CONVERGENCE OF EXCLUSION PROCESSES AND THE KPZ EQUATION TO THE KPZ FIXED POINT

JEREMY QUASTEL AND SOURAV SARKAR

ABSTRACT. We show that under the 1:2:3 scaling, critically probing large space and time, the height function of finite range asymmetric exclusion processes and the KPZ equation converge to the KPZ fixed point, constructed earlier as a limit of the totally asymmetric simple exclusion process through exact formulas. Consequently, based on recent results of [34], [11], the KPZ line ensemble converges to the Airy line ensemble.

For the KPZ equation we are able to start from a continuous function plus a finite collection of narrow wedges. For nearest neighbour exclusions, we can take (discretizations) of height functions with \( h(x) \leq C(1 + |x|) \). For non-nearest neighbour exclusions, we are restricted at the present time to a class of (random) initial data, dense in continuous functions in the topology of uniform convergence on compacts.

The method is by comparison of the transition probabilities of finite range exclusion processes and the totally asymmetric simple exclusion processes using energy estimates.

1. INTRODUCTION

The Kardar-Parisi-Zhang (KPZ) equation,

\[
\partial_t h = \lambda (\partial_x h)^2 + \nu \partial_x^2 h + \sqrt{D} \xi, \tag{1.1}
\]

where \( \xi \) is space-time white noise, the distribution valued delta correlated Gaussian field \( \langle \xi(t, x), \xi(s, y) \rangle = \delta(t - s)\delta(x - y) \), was introduced by Kardar, Parisi and Zhang in 1986 [19] as an equation for a randomly evolving height function \( h \in \mathbb{R} \) which depends on position \( x \in \mathbb{R} \) and time \( t \in \mathbb{R}_+ \). \( \lambda, \nu \) and \( D \) are physical constants. Its derivative \( u = \partial_x h \) satisfies the stochastic Burgers equation

\[
\partial_t u = \lambda \partial_x u^2 + \nu \partial_x^2 u + \sqrt{D} \partial_x \xi. \tag{1.2}
\]

A dynamic renormalization group analysis was performed on (1.2) in [13] (see also [19], [31]), predicting a dynamical scaling exponent \( z = 3/2 \) and strong coupling fixed point. In our language, this means that one expects to see a non-trivial universal fluctuation field under the 1:2:3 rescaling

\[
h \mapsto \delta h(\delta^{-3}t, \delta^{-2}x). \tag{1.3}
\]

Universal is meant in the sense that it is supposed to be the attractor as \( \delta \searrow 0 \) for a huge class of related models in one dimension; discrete growth models, directed polymer free energies, driven diffusive systems. Since the fluctuations remember the initial state, it is a scaling invariant Markov process. This process, the KPZ fixed point (see the discussion after (2.10) for the definition), was characterized in [20] through exact formulas for its transition probabilities, based on the exact solvability of a special member of the KPZ universality class, the totally asymmetric simple exclusion process (TASEP), an interacting particle system whose height function can be thought of as a single step version of (1.1), or at the stochastic Burgers level, as modelling one-way traffic flow.

Exact one point distributions have been found for the KPZ equation, and for a two way version of TASEP known as ASEP or asymmetric simple exclusion process, for a few special initial data (narrow wedge [28], [11], [26], [12], half-Brownian [29], [9], [18], Brownian [30], [17], [4], and a conjectural formula for flat [5], see also [22]). However, it is not known at the present time – nor really expected – that the solvability extends to general initial data, or to multipoint distributions (excepting perhaps the narrow wedge case). On the other hand, it was well understood at a physical level that the 1:2:3 scaling limit of the KPZ equation, as well as any exclusion process with sufficiently local jumps (finite

Date: June 3, 2024.
range asymmetric exclusion processes, or AEP), should coincide with that of TASEP. Note that such a statement contains several well known conjectures such as the convergence of multipoint distributions to those of the Airy process for narrow wedge initial data, and (following recent results [34],[11]) the convergence of the KPZ line ensemble to the Airy line ensemble. The gap was the lack of any analytic methods to prove such a convergence; all one knew how to do was find an exact formula and take its limit.

The main contribution of this article is an analytic method which compares the transition probabilities of any finite range asymmetric exclusion process to TASEP, with a close enough comparison that shows they have the same 1:2:3 scaling limit. In the weakly asymmetric limit it also allows one to prove that the KPZ equation transition probabilities, for very general initial data, converge to the KPZ fixed point transition probabilities.

The key step is an estimate of the difference of the transition probabilities of the two processes, as long as one starts with a bound on the $L^2$ norm of the Radon-Nikodym derivative with respect to a global equilibrium. The second step of the proof extends the result from such initial data, to a broader class. In the case of the KPZ equation, key tools available are a positive temperature version of the variational formulation and Brownian Gibbs property. In the case of exclusion processes, these are missing$^1$. In the nearest neighbour case (ASEP) the order of height functions is preserved. Combined with estimates using second class particles, and asymptotic fluctuations for a few special initial data due to Tracy and Widom [29],[28], we are able to extend to very general deterministic initial data. Theses are missing in the non-nearest neighbour case. Therefore the class of initial data for which we can prove the convergence for AEP at this time, while dense in a certain sense, is not as complete as one would wish. For the detailed results, see Theorem 2.2.

While we cannot prove the result for all initial data for exclusion processes at this time, to our knowledge the results for non-nearest neighbour exclusion represent the first results of KPZ universality for any data for any model which is not to some degree algebraically solvable.

Two natural questions arise; 1) Can one solve the initialization problem for AEP, i.e. show that starting from initial data having a well defined asymptotic profile, the distribution at any positive time is close enough to a finite energy measure, while not having changed its asymptotic profile? 2) How general is the method? Again, we do not have a satisfactory answer. Here we are perturbing off TASEP, and the argument leading up to (5.2) uses many special properties of that process, in particular the skew-time reversibility (see Appdx. (C)) which gives the seemingly crucial Lem. 5.1.

At the same time (and place), another proof of KPZ universality for the KPZ equation was obtained by Bálint Virág [33]. Although the proofs are very different, they both use as their launchpad the observation that, for the KPZ equation, it is only necessary to prove the convergence for a dense class of initial data. Virág uses a class of solvable initial data related to the Baik-BenArous-Peché statistics together with novel symmetries with respect to initial and final data satisfied by special solvable polymer models which can be rescaled to the KPZ equation. In this sense, the proof invites extension to further classes of partially solvable polymer models.

The dense class on which we initially prove convergence to the fixed point involves a bound on the $L^2$ norm of the Radon-Nikodym derivative with respect to equilibrium, and this key part of the argument is in Sec 3-6. The results of Sec. E prove tightness and convergence in the uniform-on-compact topology for the height functions of any model for which the convergence of probabilities for certain class of sets is known (see Prop. E.1). Once that is done, one uses known properties of the process to extend to more general initial data. This is done in Sec. 7-8. These arguments are somewhat universal in the sense that the results of Sec. 7 will hold for any model satisfying convergence for finite energy initial data and skew time reversibility. Appx. D contains the proof of the key Lem. 5.1 representing the gradient of the TASEP transition probabilities as the joint distribution of a max and argmax. Appx. E contains the proof of the main tool to extend the results from nearest neighbour to non-nearest neighbour exclusion processes, which is the strong sector bound of Lin Xu [35] and S.R.S.Varadhan [32].

$^1$An earlier version of this article on the arxiv, and the published version erroneously claim that ASEP and the KPZ equation are skew-time reversal invariant, and this is used to extend to a broader class. The present version corrects the mistake. We are very grateful to L.-C.Tsai for pointing it out to us.
condition (see Defn. [3.1]) can be thought of in this context as telling one when an operator has the 1:2:4, diffusive, Edwards-Wilkinson scaling. At our 1:2:3 scaling, such an operator can be treated as an error term. Since the method is analytic and independent of solvability of the model, it invites extensions to other particle systems, though it requires as input a comparison system for which the convergence to the KPZ fixed point is already known.

2. Models and main results

We consider finite range asymmetric exclusion processes (AEP) on \( \mathbb{Z} \) with non-zero drift. There are particles on the lattice, with at most one particle per site. The particles attempt jumps \( x \mapsto x + v \) at rate \( p(v) \) where \( \{ v \in \mathbb{Z} : p(v) > 0 \} \) is finite, and, to avoid degeneracies, is assumed that \( \{ v \in \mathbb{Z} : p^{\text{sym}}(v) = p(v) + p(-v) > 0 \} \) additively generate all of \( \mathbb{Z} \), or, in other words, the underlying single particle random walk with jump law \( p^{\text{sym}}(\cdot) \) is irreducible. In order for our results to hold it is necessary that

\[
\sum_v v p(v) \neq 0. \tag{2.1}
\]

Otherwise, the process will not move on the scales we are observing\(^2\). We can always assume it is positive, and by multiplying all the rates by a constant (i.e. changing the time scale by a constant) we may as well assume that

\[
\sum_v v p(v) = 1. \tag{2.2}
\]

The height function at any fixed time \( t \) is a simple random walk path \( h(x), x \in \mathbb{Z} \) with \( h(x+1) = h(x) \pm 1 \). If \( h(x+1) = h(x) + 1 \) we say there is a particle at \( x \) and write \( \eta(x) = 1 \) and if \( h(x+1) = h(x) - 1 \) we say there is no particle at \( x \) and write \( \eta(x) = 0 \). We can alternatively let

\[
\dot{\eta}(x) = 2\eta(x) - 1, \quad \text{and write} \quad h(x+1) = h(x) + \dot{\eta}(x). \tag{2.3}
\]

When we include the time coordinate in the notation, we write the height function as \( h(t, x) \) instead of just \( h(x) \).

As we said, a particle at \( x \) attempts jumps to site \( x + v \) at rate \( p(v) \). But the jump takes place only if the target site is unoccupied. This is called exclusion. All the particles are doing this independently of the others, and since time is continuous, there are no ties to break. Note that when the jump occurs, the height function either drops by 2 at sites \( x + 1, \ldots, x + v \), if \( v > 0 \), or increases by 2 at sites \( x + v + 1, \ldots, x \) if \( v < 0 \).

The special case where \( p(\cdot) \) is nearest neighbour is referred to as ASEP (asymmetric simple exclusion process). Simple here means nearest neighbour. The further special case when the jump law is \( p(1) = 1 \) and \( p(v) = 0 \) otherwise is called TASEP (totally asymmetric simple exclusion process); here jumps only to the right-hand neighbor are allowed.

We rescale the height function by the 1:2:3 KPZ scaling

\[
h_\varepsilon(t, x) = \varepsilon^{1/2} h(2\varepsilon^{-3/2} t, 2\varepsilon^{-1} x) + \varepsilon^{-1} t. \tag{2.4}
\]

The factors of 2 are put in to coordinate with earlier work. The height function in (2.4) lives on \( \mathbb{R}^+ \times \frac{1}{2} \varepsilon \mathbb{Z} \subset \mathbb{R}^+ \times \mathbb{R} \), where the first coordinate represents time. When there is no scope for confusion, we shall often suppress the time coordinate and the dependence on \( \varepsilon \) in the notation. We now have

\[
h(x + \frac{1}{2} \varepsilon) = h(x) + u(x) \quad \text{with} \quad u(x) \in \{-\varepsilon^{-1/2}, \varepsilon^{1/2}\}. \tag{2.5}
\]

The discrete state space \( S_\varepsilon \) = collection of such functions mapping \( \frac{1}{2} \varepsilon \mathbb{Z} \mapsto \varepsilon^{1/2} \mathbb{Z} \). These are naturally embedded in the continuous state space \( \hat{S} \) = continuous functions on \( \mathbb{R} \) by connecting \( h(x) \) and \( h(x + \frac{1}{2} \varepsilon) \) with straight lines. When we say for example, a continuous function \( h \) is initial data for AEP, we mean that at each level \( \varepsilon > 0 \), there is an initial data \( h_\varepsilon \) and \( h_\varepsilon \to h \) in the uniform-on-compact topology.

\(^2\)If we call \( m := \sum_v v p(v) \), it is fairly straightforward to check that the proof shows convergence of \( h_\varepsilon(2t, 2x) \) to \( \text{sgn}(m) \) times the KPZ fixed point running at speed \( |m| \). In particular, if \( m = 0 \) the argument shows that there is no movement on the KPZ time scale.
Call $\mu_\varepsilon(d\hat{h})$ the law of the two sided symmetric simple random walk on $\frac{1}{\varepsilon}\mathbb{Z}$, with steps of size $\varepsilon^{1/2}$ with $\hat{h}(0) = 0$. Let $\text{Leb}_\varepsilon(dh(0))$ be the discrete Lebesgue measure giving mass $\varepsilon^{1/2}$ to each $h(0) \in \varepsilon^{1/2}\mathbb{Z}$. An $h \in S_\varepsilon$ is built out of $h(0)$ and $\hat{h}(x), x \in \frac{1}{\varepsilon}\mathbb{Z}$ by

$$h(x) = h(0) + \hat{h}(x),$$

(2.6)

and the product measure

$$\nu_\varepsilon(dh) = \text{Leb}_\varepsilon(dh(0)) \times \mu_\varepsilon(d\hat{h})$$

(2.7)

is invariant for the process.

Next we turn to the KPZ equation. The solution $[3], [15], [14]$ is a Markov process with state space $\mathbb{S} = \text{continuous functions on } \mathbb{R}$. More precisely, we consider the Cole-Hopf solution, the logarithm of the solution of the stochastic heat equation (SHE),

$$h = \gamma \log Z, \quad \partial_t Z = \nu \partial_x^2 Z + \gamma^{-1}\sqrt{D}\xi Z,$$

(2.8)

where $\gamma = \nu/\lambda$. SHE is one of the few nonlinear stochastic partial differential equations for which a solvability theory is relatively straightforward; in fact, the solution can be written as an explicit chaos series $[23]$. Note that the SHE (2.8) can be solved starting from more general initial data. In particular, a narrow wedge initial data for KPZ equation means to start the SHE (2.8) with a Dirac delta measure. Multiple narrow wedges just means a sum of such, and one could also add to this a more regular function as initial data. The map $h \mapsto \delta^3 h(\delta^{-2}t, \delta^{-1}x)$ transforms the coefficients in (1.1) by $(\lambda, \nu, D) \mapsto (\delta^{2-z-b} \lambda, \delta^{2-z} \nu, \delta^{2b-z+1} D)$, so it is enough to always consider a standardized KPZ equation

$$\partial_t h = \frac{1}{4}(\partial_x h)^2 + \frac{1}{4}\partial_x^2 h + \xi$$

(2.9)

which is taken by the 1:2:3 scaling (1.3) into what we will call KPZ$_{\delta}$, into

$$\partial_t h = \frac{1}{4}(\partial_x h)^2 + \frac{1}{4}\partial_x^2 h + \delta^{1/2}\xi,$$

(2.10)

One of our main results is that as $\delta \searrow 0$ this converges to the KPZ fixed point, the 1:2:3 scaling invariant Markov process expected to govern fluctuations for all models in the KPZ class. It can be described in two ways, as a Markov process of evolving height functions $h(t, \cdot)$ through its explicit transition probabilities, or through a variational formulation.

The Markov transition probability of the KPZ fixed point starting from initial height function $h_0(x)$, $x \in \mathbb{R}$ which is UC, meaning upper semi-continuous $h_0$ with $h_0(x) \leq C(1 + |x|)$ for some $C < \infty$, is defined through its finite dimensional distributions given by

$$P_{h_0}(h(t, x_i) \leq r_i, i = 1, \ldots, m) = \det(I - K_{h_0, x_i, r_i, t})_{L^2(\mathbb{R}_+; \mathbb{R}^m)}$$

(2.11)

where the $i, j$ entry of the $m \times m$ matrix operator $K$ is given by

$$(K_{h_0, x_i, r_i, t})_{ij} = \lim_{L \to \infty} e^{-\frac{1}{3}t\partial^3-(x_i+L)\partial^2+r_i\partial}\left(P_{-L,L}^{\text{Hit}, h_0} - 1_{x_i < x_j}\right) e^{\frac{1}{3}t\partial^3+(x_j-L)\partial^2-r_j\partial}$$

(2.12)

where the integral operator kernel $P_{-L,L}^{\text{Hit}, h_0}(u_1, u_2)$ is the probability that a Brownian bridge from $u_1$ at time $-L$ to $u_2$ at time $L$ enters the hypograph of $h_0$ in between. UC is equipped with the topology of convergence of hypographs in the local Hausdorff topology. For UC functions (2.12) make sense and define Feller transition probabilities, and the process stays in this space. But in fact, one can check directly from the formula that the height function at any strictly positive time is locally Brownian, so for any strictly positive time, it lives in the more regular space of height functions which are locally Hölder of any exponent $< 1/2$. Note that if $h_0 = -\infty$ outside a box of size $\ell$, then the expression inside the limit on the right hand side of (2.12) is constant for $L > \ell$. So in that case one doesn’t have to take a limit. From the formula one can check that the process is invariant under the 1:2:3 scaling (2.4), and

$^3$it is linear in the initial data, but non-linear in the noise.

$^4$The reason for the standard in the nonlinear term is that $-x^2/t$ solves the Hopf equation $\partial_t h = \frac{1}{2}(\partial_x h)^2$ so all the parabolic shifts come without unnecessary constants. The ratio $\nu/D = 1/4$ corresponds to a diffusivity 2 for the invariant Brownian motion (it is invariant for the equation without, or with the non-linear term.) These were the choices made in the definition of the KPZ fixed point in [23]. At the level of naive scaling theory, it ensures that (2.10) converges to the KPZ fixed point without extra rescaling.
is spatially and temporally homogeneous. Brownian motion with any drift is invariant, except for a non-trivial global height shift, and the process can informally be thought of as a non-trivial evolution of Brownian motions.

An alternate description is given through the fact that the process preserves the max operation, i.e. \( h(t, x; h_0^1) \lor h(t, x; h_0^2) = h(t, x; h_0^1) \lor h(t, x; h_0^2) \) where \( h(t, x; h_0) \) denotes the solution at time \( t \), evaluated at position \( x \), starting from initial data \( h_0 \). Thus the special initial conditions \( \delta_y(x) := 0 \) for \( x = y \) and \(-\infty\) for \( x \neq y \) play the role of Dirac delta functions for linear equations. Starting from such an initial condition, and from the 1:2:3 scaling we obtain a solution we can write as

\[
h(t, x; \delta_y) := t^{1/3}A(t^{-2/3}y, t^{-2/3}x) - \frac{1}{2}(y - x)^2.
\] (2.13)

The two-parameter stochastic process \( A(x, y) \) is stationary in both variables, and is distributed as an Airy process \([20]\) in each of the two variables separately. It is called the Airy process. Unfortunately we do not know its joint distributions explicitly. In \([10]\) it is shown to be a (non-explicit) function of the Airy line ensemble, a random sequence of functions \( A_1 > A_2 > \ldots \), introduced by Prähofer and Spohn \([23]\), which is stationary, see also Corwin and Hammond \([7]\). The top line \( A_1 \) is known as the Airy\(_2\) process that, after subtracting a parabola, appear as the limiting spatial fluctuation of random growth models starting from a single point (not to be confused with the Airy\(_1\) process, elsewhere occasionally referred to as \( A_1 \), which governs those arising from a flat interface). The parabolically shifted version \( A_i \) defined as \( \widetilde{A}_i(x) = A_i(x) - x^2 \), \( i = 1, 2, \ldots \), will be called the parabolic Airy line ensemble. From the max property it is rather easy to see that the KPZ fixed point satisfies the variational principle

\[
h(t, x; h_0) = \sup_{y \in \mathbb{R}} \left\{ t^{1/3}A(t^{-2/3}y, t^{-2/3}x) - \frac{1}{2}(y - x)^2 + h_0(y) \right\}.
\] (2.14)

The equality is for fixed \( t \) as stochastic processes in \( x \). For fixed \( t \), the Airy sheet is Hölder of any exponent less than \( 1/2 \) in the space coordinates. In particular, from (eq. (104) of) Prop. 10.5 in \([10]\), it follows that for any \( x_1, x_2, y_1, y_2 \in [-b, b] \) for some \( b > 0 \), and \( d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \), we have

\[
\left| t^{1/3}A(t^{-2/3}x_1, t^{-2/3}y_1) - t^{1/3}A(t^{-2/3}x_2, t^{-2/3}y_2) \right| \leq C d^{1/2} \log^{1/2}(1 + d^{-1}),
\] (2.15)

for some random constant \( C \) satisfying \( P(C > m) \leq c_1 e^{-c_2 m^2} \) for some positive constants \( c_1, c_2 \) that depend on \( t, b \), but not on \( x_i \)’s and \( y_i \)’s. To include the \( t \) in the picture one has to go up one level to the directed landscape (see \([10]\)). From Prop. 10.5 in \([10]\) it is easy to see that, as a function of \( t, h(t, x) \) is Hölder of any exponent less than \( 1/3 \).

Note that the construction of the directed landscape in \([10]\) uses as input the model of Brownian last passage percolation (BLPP), while the transition probabilities of the KPZ fixed point in \([20]\) are obtained through limits of exact formulas for TASEP. The fact that one obtains the same object is proven by obtaining similar formulas for the KPZ fixed point in the context of BLPP \([21]\).

Let \( UC_A = \{ h : \mathbb{R} \to [-\infty, \infty), h(x) \leq A(1 + |x|) \} \) with

\[
dist(h_1, h_2) = \sum_{\ell=1}^{\infty} 2^{-\ell} [\text{HDist}_\ell(h_1(\text{hypo}(h_1), \text{hypo}(h_2)) \land 1]
\]

where the Hausdorff distance \( \text{HDist}_\ell \) is the Hausdorff distance restricting to \([-\ell, \ell]\) using the metric \( d((x_1, r_1), (x_2, r_2)) = |x_1 - x_2| + |e^{r_1} - e^{r_2}| \). \( UC = \cup_{A=1}^{\infty} UC_A \). \( UC_A \) consists of functions \( h_\varepsilon \) in \( UC_A \) satisfying \( h_\varepsilon(x + \varepsilon) = h_\varepsilon(x) \pm \varepsilon^{1/2} / \varepsilon^{1/2} Z \) for \( x \in \varepsilon Z \) and linear in between. We say such functions converge to \( h_0 \) in \( UC \) if the convergence is in some fixed \( UC_A \).

The following summarizes results of \([20]\) on the convergence of the 1:2:3 rescaled TASEP height function to the KPZ fixed point.

**Theorem 2.1** ([20]). Consider the totally asymmetric simple exclusion process and let \( h_\varepsilon(t, x) \) be the 1:2:3 rescaled height function as in (2.4). Let the initial data satisfy \( h_\varepsilon(0, x) \to h_0(x) \) where \( h_0(x) \leq C(1 + |x|) \) for some \( C < \infty \) in \( UC \). Then \( h_\varepsilon(t, x) \to h(t, x) \) in distribution, where \( h(t, x) \) is the KPZ fixed point starting from \( h_0(x) \).
In the theorem below and in the rest of the paper, convergence of measures always means convergence when tested against bounded continuous functions. We will often consider target sets $B$ of the form $\text{hyp}(g)$ for nice functions $g$ where $\text{hyp}(g)$ is the set

$$\text{hyp}(g) = \{ h : h(x) \leq g(x), x \in \mathbb{R} \}. \quad (2.16)$$

This is a very slight abuse of the standard notation where $\text{hyp}(g)$ is the hypograph of $g$; our $\text{hyp}(g)$ is the set of functions $h$ whose hypographs $\text{hypo}(h) = \{(x, y) \in \mathbb{R} \times (\mathbb{R} \cup \{-\infty\}), y \leq h(x)\}$ are contained in the hypograph of $g$.

**Theorem 2.2.** (1) Suppose that $f_{0,\varepsilon} d\nu_\varepsilon \to dp$ with $\varepsilon^{1/4}\|f_{0,\varepsilon}\|_{L^2(d\nu_\varepsilon)} \to 0$ and $\rho$-almost surely any sample path $h(\cdot)$ satisfies $h(x)$ is bounded on any compact interval and $|h(x)| \leq C(1 + |x|)$ for some $C > 0$, then started from $f_{0,\varepsilon} d\nu_\varepsilon$, the AEP height function $\tilde{h}_\varepsilon(t, x) := h_\varepsilon(2t, 2x) + \varepsilon^{-1} t/2$ converges as $\varepsilon \to 0$ in distribution to the KPZ fixed point at fixed time $t > 0$ started from $\rho$ in the uniform-on-compact topology.

(2) Let $h_{0,\varepsilon} \to h$ in UC. Then under the same scaling the ASEP height function $\tilde{h}_\varepsilon$ converges in distribution to the KPZ fixed point at fixed time $t > 0$ started from $h$ in the topology of uniform convergence on compact sets.

(3) Start KPZ, $\delta \leq 2$, with a continuous function bounded above by $A(1 + |x|)$ plus a finite collection of narrow wedges. Then, as $\delta \to 0$, the solution, after addition of the Itô factor $t/12$, converges in distribution, in the topology of uniform convergence on compact subsets, to the KPZ fixed point. The KPZ line ensemble converges to the Airy line ensemble in the same sense (see Sec. [2]).

The space-time convergence is only proved here for the KPZ equation. For the ASEP and AEP, the convergence is only proven for a fixed time, not as a process in $t$.

The key step of the proof is based on classical semigroup theory; generators and control of differences of transition probabilities through Dirichlet forms. In the next few paragraphs we attempt to provide some intuition. To describe it, and for use later, we need to write down the infinitesimal generators of our various Markov processes on the space of discrete height functions $\mathbb{S}_\varepsilon$.

The rescaled AEP, $\square$, has generator

$$L_\varepsilon f = 2\varepsilon^{-3/2} \sum_{x \in \frac{1}{2}\varepsilon\mathbb{Z}} \sum_{v \in \mathbb{Z}} p(v) \nabla_{x,v} f, \quad \nabla_{x,v} f(h) = f(h^{x,v}) - f(h). \quad (2.17)$$

Here $h^{x,v} = h$ unless there is a particle at $x$ and no particle at $x + \frac{1}{2}\varepsilon v$, in which case $h^{x,v}$ is the new height function after the particle has performed the jump.

The idea is to compare AEP to the rescaled TASEP, which has generator

$$T_\varepsilon f = 2\varepsilon^{-3/2} \sum_{x \in \frac{1}{2}\varepsilon\mathbb{Z}} \nabla_{x,1} f. \quad (2.18)$$

The operator $\nabla_{x,v}$ has a symmetrized version

$$\nabla_{x,v}^{\text{sym}} f(h) = f(h^{x,v,\text{sym}}) - f(h), \quad (2.19)$$

where $h^{x,v,\text{sym}} = h$ unless there is a particle at $x$ and no particle at $x + \frac{1}{2}\varepsilon v$ or no particle at $x$ and a particle at $x + \frac{1}{2}\varepsilon v$, in which cases $h^{x,v,\text{sym}}$ is the new height function after the particle numbers at $x$ and $x + \frac{1}{2}\varepsilon v$ have been exchanged.

The symmetric simple exclusion process (SSEP) which allows flips both ways, has generator

$$S_{\varepsilon,\delta} f := 2\delta\varepsilon^{-2} \sum_{x \in \frac{1}{2}\varepsilon\mathbb{Z}} \nabla_x^{\text{sym}} f \quad (2.20)$$

The symmetric part of the generator $L_\varepsilon$ of AEP has the same form, but with non-nearest neighbour terms $\nabla_{x,v}^{\text{sym}}$, and with $\delta = \varepsilon^{1/2}$.
The weakly asymmetric simple exclusion process (WASEP), allows flips both ways, with a slight asymmetry. Its generator is
\[ L_{\varepsilon, \delta} = T_{\varepsilon} + S_{\varepsilon, \delta}. \] (2.21)

It is known (see Prop. 6.5) that WASEP converges as \( \varepsilon \downarrow 0 \) to the KPZ equation (1.1) with \( (\lambda, \nu, D) = \left( \frac{1}{4}, \frac{1}{2}, \delta \right) \).

2.1. Heuristics and proof overview. Informally we have \( e^{T_{\varepsilon}} \rightarrow \) the semigroup of the KPZ fixed point, and \( e^{(T_{\varepsilon} + S_{\varepsilon, \delta})} \rightarrow \) the semigroup of the KPZ equation (1.1) with \( (\lambda, \nu, D) = \left( \frac{1}{4}, \frac{1}{2}, \delta \right) \). Furthermore a straightforward computation shows that \( e^{T_{\varepsilon} + S_{\varepsilon, \delta}} \rightarrow \) the semi-group of the linear stochastic differential equation (1.1) with \( (\lambda, \nu, D) = (0, \delta, \delta) \). We can think of this as saying \( S_{\varepsilon, \delta} = O(\delta) \) in some sense.

So one expects
\[ e^{t(T_{\varepsilon} + S_{\varepsilon, \delta})} - e^{tT_{\varepsilon}} \sim e^{t(T_{\varepsilon} + O(\delta))} - e^{tT_{\varepsilon}} \sim O(\delta). \] (2.22)

Taking \( \varepsilon \downarrow 0 \), this shows heuristically that the difference between the semigroups of the KPZ equation (1.1) with \( (\lambda, \nu, D) = (0, \delta, \delta) \) and the semigroup of the KPZ fixed point should be of order \( \delta \).

The heuristics for the convergence of AEP to the KPZ fixed point are similar, except that now \( \delta \) is taken to be \( \varepsilon^{1/2} \) and one needs an extra fact that the difference of the asymmetric parts of the operators scales like the symmetric part. It is true under condition (2.2) because then the difference is the generator of a mean zero asymmetric exclusion process, which is known to be diffusive [35]. Technically, the necessary estimate is provided by the strong sector condition (see (3.9) and Lem. 4.2).

It is unclear how to directly make (2.22) rigorous, or even precise. One comes up against the general problem in infinite particle systems of a lack of controllable norms. We take a different approach. A general formula for the difference of transition probabilities is presented in (3.10). It reduces the problem to the computation of the Dirichlet form of a known quantity, the transition probability of TASEP. It turns out this can be computed without exact formulas, but just using the time reversal invariance of the process. As long as one makes a small average, one can control this term well enough to obtain the convergence. But one loses a square root as compared to (2.22), so the bound one gets is not optimal. It also means one has to start with slightly randomized initial and final conditions. A second step is then to extend to more general data.

3. Difference of two Markov processes

Consider two time-homogeneous Markov processes \( h_t^{(1)}, h_t^{(2)}, t \geq 0 \) on a state space \( S \) which share a common invariant measure \( \nu \). We would like to estimate the difference between their transition probabilities
\[ p_t^{(1)}(h, B) := \text{Prob}(h_t^{(1)} \in B \mid h_0^{(1)} = h) \quad \text{and} \quad p_t^{(2)}(h, B) := \text{Prob}(h_t^{(2)} \in B \mid h_0^{(2)} = h). \] (3.1)

We could also start each in a probability measure \( f_0(h)\nu(\text{d}h) \) and ask for the difference between \( p_t^{(i)}(f_0\nu, B) := \int p_t^{(i)}(h, B)f_0(h)\nu(\text{d}h), i = 1, 2 \), i.e. the difference if we start the two processes with the same distribution \( f_0(h)\nu(\text{d}h) \). Suppose the two processes have infinitesimal generators \( L_1 \) and \( L_2 \). The transition probabilities satisfy Kolmogorov’s backward equations
\[ \partial_t p_t^{(1)}(h, B) = L_1 p_t^{(1)}(h, B) \quad \text{and} \quad \partial_t p_t^{(2)}(h, B) = L_2 p_t^{(2)}(h, B). \] (3.2)

Here \( L_i, i = 1, 2 \) are acting on the \( h \) variable. On the other hand, if we start with \( f_0(h) \), the processes at time \( t \) will have probability distributions \( f_t^{(i)}(h)\nu(\text{d}h) \) solving the forward equations
\[ \partial_t f_t^{(i)} = L_t^* f_t^{(i)}, \quad i = 1, 2. \] (3.3)

Thus
\[ \partial_s f_{t-s}^{(2)}(h)p_s^{(1)}(h, B) = -L_2 f_{t-s}^{(2)}(h)p_s^{(1)}(h, B) + f_{t-s}^{(2)}(h)L_1 p_s^{(1)}(h, B). \] (3.4)

Integrating from 0 to \( t \) and with respect to \( \nu \), we obtain
\[ p_t^{(1)}(f_0\nu, B) - p_t^{(2)}(f_0\nu, B) = \int_0^t \int f_{t-s}^{(2)}(h)(L_1 - L_2)p_s^{(1)}(h, B)\nu(h)\text{d}s. \] (3.5)
Here we have used
\[
\int f^{(2)}_t(h)p_0(h, A)\,d\nu(h) = \int_A f^{(2)}_{s}(h)\,d\nu(h) = \int \text{Prob}(h^{(2)}_s \in A \mid h^{(2)}_0 = h)f_0(h)\,d\nu(h)
\]
\[
= p^{(2)}_s(f_0, A)
\]
for measurable \( A \), and \( L_1 - L_2 \) acts on \( p^{(1)}_s \) through the variable \( h \).

One could do various things at this point, perhaps take a supremum over measurable subsets \( B \in S \) in order to compute the total variation distance. In our case, the total variation is not expected to be small, and we will be satisfied if we can prove the two probabilities are close for a wide class of fixed \( B \). To study \( \{3.5\} \), it is clear we will need information about both \( p^{(1)}_s(h, B) \), and \( f^{(2)}_s \), \( 0 \leq s \leq t \). To learn something about the latter, we might assume \( \|f_0\|_2^2 = \int f_0^2\,d\nu < \infty \) and study what happens to \( \|f^{(2)}_t\|_2^2 = \int (f^{(2)}_s)^2\,d\nu \):
\[
\|f^{(2)}_t\|_2^2 - \|f_0\|_2^2 = 2 \int_0^t \int f^{(2)}_s L_2 f^{(2)}_s\,d\nu ds = 2 \int_0^t \int f^{(2)}_s L_2 f^{(2)}_s\,d\nu ds = -2 \int_0^t \mathcal{D}_2(f^{(2)}_s)ds,
\]
(3.6)
where \( L_2^{\text{sym}} = \frac{1}{2}(L_2 + L_2^*) \) is the symmetric part of \( L_2 \). We could do the same thing with the other process to create another Dirichlet form, \( \mathcal{D}_1 \). In the applications we will be interested in, the Dirichlet forms corresponding to \( T \) and \( L \) are comparable, i.e. there is a \( 0 < c < \infty \) such that
\[
c^{-1} \mathcal{D}_1(f) \leq \mathcal{D}_2(f) \leq c \mathcal{D}_1(f).
\]
(3.7)
In \( \{3.6\} \), the non-positivity of the right hand side tells us that the integral of the Dirichlet form can be controlled by the initial \( L^2 \) norm. Together with \( \{3.7\} \), we have
\[
\int_0^t \mathcal{D}_1(f^{(2)}_s)ds \leq c \int_0^t \mathcal{D}_2(f^{(2)}_s)ds \leq C\|f_0\|_2^2,
\]
(3.8)
for some constant \( C > 0 \).

**Definition 3.1.** An operator \( A \) on \( \mathbb{R}^S \) (real valued functions on \( S \)) satisfies the strong sector condition with respect to \( \mathcal{D} \) (or \( L^{\text{sym}} \)) if there exists \( C < \infty \) such that for all \( g_1, g_2 \in \mathbb{R}^S \) with \( \mathcal{D}(g_1), \mathcal{D}(g_2) < \infty \),
\[
\left| \int g_1 A g_2\,d\nu \right| \leq C \mathcal{D}(g_1)^{1/2} \mathcal{D}(g_2)^{1/2}.
\]
(3.9)

By \( \{3.5\}, \{3.9\} \) and \( \{3.8\} \) and the Cauchy-Schwarz inequality, we have

**Lemma 3.2.** Suppose that \( L_1 - L_2 \) satisfies the strong sector condition with respect to \( \mathcal{D}_1 \). Then there exists \( 0 < C < \infty \) such that
\[
\left| p^{(1)}_t(f_0\nu, B) - p^{(2)}_t(f_0\nu, B) \right| \leq C \int_0^t \mathcal{D}_1(f^{(2)}_{t-s})^{1/2} \mathcal{D}_1(p^{(1)}_s)^{1/2} ds \leq C\|f_0\|_2 \int_0^t \mathcal{D}_1(p^{(1)}_s)ds,
\]
(3.10)

The idea now is that if process 1 is TASEP, then we know a great deal about \( p^{(1)}_s \), enough that we can essentially compute the right hand side. The method of attack is to use this bound, with a little bit of extra averaging (see the beginning of Sec. 5), to estimate the difference between the rescaled TASEP and ASEP, respectively, AEP or WASEP.

### 4. Dirichlet forms

The generators \( \{2.17\} \) and \( \{2.18\} \) are invariant with respect to the (unnormalized) measure \( \nu_\varepsilon \) defined in \( \{2.7\} \), and have Dirichlet forms
\[
\mathcal{D}_\varepsilon(f) := -\int fL_\varepsilon f\,d\nu_\varepsilon = \varepsilon^{-3/2} \sum_{x \in \frac{1}{\varepsilon} \mathbb{Z}} \sum_{v \in \mathbb{Z}} p(v) \int (\nabla^{\text{sym}}_{x,v} f)^2\,d\nu_\varepsilon
\]
(4.1)
and
\[
\mathcal{D}_\varepsilon(f):=-\int fT_\varepsilon df\,\nu_\varepsilon = \varepsilon^{-3/2} \sum_{x\in\frac{1}{2}\varepsilon\mathbb{Z}} (\nabla_{x}f)^2\,\nu_\varepsilon
\] (4.2)
respectively, where \(\nabla_{x}f\) is short for \(\nabla_{x,1}f\). \(\mathcal{D}_\varepsilon\) is in a sense the most basic Dirichlet form, summing the energies from the elementary flips. The philosophy is to compare all other Dirichlet forms to this one. So although it comes from the nearest neighbour \(p(\cdot)\) we do not indicate that or given it a special name.

**Lemma 4.1.** The Dirichlet forms (4.2) and (4.1) of TASEP and AEP are comparable (in the sense of (3.7)) with a positive constant \(c < \infty\) independent of \(\varepsilon > 0\).

**Proof.** Introduce the measure preserving transformation \(e_{x,y}h = h_{x-2\varepsilon^{-1}(y-x),\text{sym}}\), where \(h_{x,y,\text{sym}}\) is defined immediately after (2.19). We have \(e_{x,y} = e_{y,x}\) and we can write
\[
e_{x,x+\frac{1}{2}\varepsilon} = e_{x_n,x_n+\frac{1}{2}\varepsilon} \cdots e_{x_1,x_1+\frac{1}{2}\varepsilon}
\] (4.3)
where \(n = 2|v| - 2\) and, if \(v > 0\), we have \(x_1 = x, x_2 = x + \frac{1}{2}\varepsilon, \ldots, x_{v-1} = x + \frac{1}{2}\varepsilon(v - 1), x_v = x + \frac{1}{2}\varepsilon(v - 2), x_{v+1} = x + \frac{1}{2}\varepsilon(v - 3), \ldots, x_{2v-2} = x\). In terms of the gradient-of-height variables \(\eta\) (defined in (2.3)), \(e_{x,y}\) exchanges the numbers at \(x\) and \(y\), and (4.3) is just the fact that the exchange of numbers at \(x\) and \(x + \frac{1}{2}\varepsilon v\) is accomplished through nearest neighbor exchanges first along bonds from \(x, x + \frac{1}{2}\varepsilon v\) to \(x + \frac{1}{2}\varepsilon v\). Hence \(h_\varepsilon\eta(x)\) gives \(\mathcal{D}_\varepsilon(f)\) and by the measure preserving property
\[
\int (\nabla_{x}f)^2\,\nu_\varepsilon \leq n\sum_{i=1}^{n} (f(e_{x_i,x_i+\frac{1}{2}\varepsilon}h) - f(h))\,\nu_\varepsilon.
\] (4.4)
Summing over \(x\) and \(v\) gives \(\mathcal{D}_\varepsilon(f) \leq C\mathcal{D}_\varepsilon(f)\). The \(C\) depends on the jump law \(p(\cdot)\) and nothing else.

In the other direction, it is easy to construct \(f\) and \(p(\cdot)\) so that \(\mathcal{D}_\varepsilon(f) = 0\) but \(\mathcal{D}_\varepsilon(f) > 0\). On the other hand, under our assumption that \(p_{\text{sym}}(\cdot)\) is irreducible, there is a path \(x_1 = x, x_2 = x + \frac{1}{2}\varepsilon y_1, \ldots, x + \frac{1}{2}\varepsilon y_{m-1} = x_{m-1} + \frac{1}{2}\varepsilon y_m\) with \(p_{\text{sym}}(y_i) > 0\). In other words, the nearest neighbour move in the Diferchil form of the TASEP can be achieved by a finite set of allowable moves of the symmetrized AEP. As above, we can write \(\nabla_{x,1} f\) as \(f(e_{x_i,x_i+\frac{1}{2}\varepsilon y_i} \cdots e_{x_1,x_1+\frac{1}{2}\varepsilon y_1} h) - f(e_{x_i,x_i+\frac{1}{2}\varepsilon y_i} \cdots e_{x_1,x_1+\frac{1}{2}\varepsilon y_1} h)\) and, as before, we get a bound \(\mathcal{D}_\varepsilon(f) \leq C\mathcal{D}_\varepsilon(f)\) with a constant \(C\) depending only on the law \(p_{\text{sym}}(\cdot)\).

Going back to AEP we can write
\[
L_\varepsilon - T_\varepsilon = M_\varepsilon - S_\varepsilon \quad \text{where} \quad M_\varepsilon = L_\varepsilon + T_\varepsilon^* \quad S_\varepsilon = T_\varepsilon + T_\varepsilon^*,
\] (4.5)
where \(T_\varepsilon^* = e^{-3/2} \sum_{x\in\frac{1}{2}\varepsilon\mathbb{Z}} \nabla_{x,-1} f\) is just the TASEP with time reversed (or space flipped). Since we have assumed \(\sum_{v} vp(v) = 1\), \(M_\varepsilon\) is the generator of a mean zero exclusion, i.e. it is of the form (2.17) with \(\sum_{v} vp(v) = 0\). The following estimate, which is key to the non-nearest neighbour case, appeared without the heights in the thesis of Lin Xu [35] and for the environment as seen from a tagged particle by S.R.S.Varadhan (Lem. 5.2 of [32]). Our variant with the heights is basically the same as [35]. For completeness we give a proof in Appx. [4].

**Lemma 4.2.** Any mean zero exclusion generator satisfies the strong sector condition (3.9) with a constant \(C < \infty\) independent of \(\varepsilon > 0\).
From (4.5), \( L_ε - T_ε \) satisfies the strong sector condition because we have matched drifts. In the nearest neighbour case, i.e. ASE, \( M_ε = 0 \) and the strong sector condition is obvious.

**Example 4.3.** Suppose we add to our AEP generator \( L_ε \) a generator \( L_ε^{\text{pert}} \) defined through a Dirichlet form, i.e. \(-\int f L_ε^{\text{pert}} df = \mathcal{D}_ε^{\text{pert}}(f)\) which is comparable to \( \mathcal{D}_ε(f) \). One checks easily that the argument goes through. We have in mind Dirichlet forms of the following type,

\[
\varepsilon^{-3/2} \int \left( \sum_{x \in \varepsilon \mathbb{Z}} \alpha_x (\nabla_x \text{sym} f)^2 + \sum_{x,x' \in \varepsilon \mathbb{Z}} \beta_{x,x'} (\nabla_{x,x'} \text{sym} f)^2 \right) d\nu_ε \tag{4.6}
\]

with functions \( \alpha_x, \beta_{x,x'} \) with \( \alpha_x \) not depending on the particle numbers at \( x - \varepsilon, x \) and \( \beta_{x,x'} \) not depending on the particle numbers at \( x - \varepsilon, x', x', x'' \) but on variables in a local (order \( \varepsilon \)) neighbourhood of both, and \( \nabla_{x,x'} f(h) = \frac{1}{\varepsilon^2} f(h^{\text{sym}}_{x,x'}) - f(h) \) where the height function \( h^{\text{sym}}_{x,x'} \) is flipped both at \( x \) and \( x' \), and the sum is over \(|x - x'| \leq C \varepsilon\). As long as \( \alpha_x, \beta_{x,x'} \) are uniformly bounded, the Dirichlet form is comparable to our standard one, and the key part of our argument continues to hold (so that the statement of Thm. 2.2(1) holds for such processes.) Of course, there are many other examples of this type. The relevance of this example is that such models (i.e. with generator \( L_ε^{\text{pert}} \)) are expected under appropriate scaling to have a hydrodynamic limit of Cahn-Hilliard type [2], [27]. Therefore they can be thought of as a very simple toy versions of the Kuramoto-Sivashinsky equation (though with a stabilizing instead of a destabilizing second order term) which has also been conjectured to lie in the KPZ universality class [25].

5. **Dirichlet form of the TASEP transition probabilities**

Let \( p_t^{\text{TASEP,} \varepsilon}(h, B) \) denote the probability for the rescaled TASEP height function at time \( t \) to be in set \( B \), given that initially it was \( h \), and \( p_t^{\text{AEP,} \varepsilon}(h, B) \) the analogue for AEP. We will show in Appx. D that \(|\nabla^\text{sym} p_t^{\text{TASEP,} \varepsilon}(h, \text{hyp}(g))|\) is roughly the joint probability

\[
R(x, h(0); \hat{h}) := \mathbb{P}^{\text{TASEP}}\left( \arg \max_y \{ h_t(y; -g) + h(y) \} = x, h_t(x; -g) + h(x) \in \{ \varepsilon^{1/2}, 2\varepsilon^{1/2} \} \right). \tag{5.1}
\]

Here \( h_t(\cdot; -g) \) denotes the TASEP height function at time \( t \) started from \(-g\), and \( h(x) = h(0) + \hat{h}(x) \). Except for the slightly annoying issue of possible non-uniqueness of the argmax we would have that for each \( \hat{h} \), \( R(x, h(0); \hat{h}) \) is a probability measure on \( \frac{1}{2} \varepsilon \mathbb{Z} \times \varepsilon^{1/2} \mathbb{Z} \). More precisely we will prove the following in Appx. D.

**Lemma 5.1.** For any \( t > 0 \) and any \( g \in S_ε \) such that \( g(x) \geq C'(1 + |x|) \) for some \( C' > 0 \), there exists a universal constant \( C < \infty \) such that

\[
\int \left( \sum_{x \in \frac{1}{2} \varepsilon \mathbb{Z}} \sum_{h(0) \in \varepsilon^{1/2} \mathbb{Z}} |\nabla^\text{sym} p_t^{\text{TASEP,} \varepsilon}(h, \text{hyp}(g))| \right)^2 \mu_ε(d\hat{h}) \leq C, \tag{5.2}
\]

where \( \mu_ε \) is as defined in (2.6).

From this one might expect \( R(x, h(0); \hat{h}) \sim \varepsilon^{3/2} \). However, the Dirichlet form has the sums over \( \frac{1}{2} \varepsilon \mathbb{Z} \) and \( \varepsilon^{1/2} \mathbb{Z} \) outside the square, and a computation gives \( \mathcal{D}_ε(p_t^{\text{TASEP,} \varepsilon}) = O(1) \), from which (5.10) tells us nothing. The reason is that for fixed \( \hat{h} \), the argmax distribution is highly intermittent in \( x \). We can get around this obstacle by looking at slightly averaged transition probabilities.

We can have our shift operator \( \tau_y \) use \( y \in \mathbb{R} \) and act on functions on \( \frac{1}{2} \varepsilon \mathbb{Z} \) by \( \tau_y h(\cdot) = h(\cdot - [y]_ε) \) and on the sets \( B \) by \( \tau_y B = \{ h(\cdot) : h(\cdot - [y]_ε) \in B \} \) where \([y]_ε \) denotes the nearest point in \( \frac{1}{2} \varepsilon \mathbb{Z} \). There is another shift, \( \sigma_r \) which does the analogue to the heights themselves, \( \sigma_r h(x) = h(x) + [r]_{\varepsilon^{1/2}} \) and \( \sigma_r B = \{ h(\cdot) : h(\cdot) - [r]_{\varepsilon^{1/2}} \in B \} \), where \([r]_{\varepsilon^{1/2}} \) denotes the nearest point in \( \varepsilon^{1/2} \mathbb{Z} \).
We will perform a little Gaussian average in $r$ and $y$. Let $y_1, y_2$ and $r$ be independent, Gaussian, mean 0 and variance $a^2$ for some $a > 0$ and let

$$p^\text{TASEP,$\epsilon$}(f_0$,$B) := \mathbb{E}[p^\text{TASEP,$\epsilon$}(f_0, \sigma_r \tau_{y_1+y_2} B)]$$

(5.3)

where $\mathbb{E}$ is the expectation over $y_1, y_2, r$. The expectation $\mathbb{E}$ has nothing to do with any physical probability space and is just a convenient way to write the small convolution, which smooths out the transition probability, reducing the intermittency. In the proof below, we use several times the symmetry of the convolution. But otherwise, it is not very important that it is Gaussian. We could have combined $y_1$ and $y_2$ as $y' = y_1 + y_2$ but writing it this way will clarify the manipulations which follow. Let

$$\hat{p}^\text{TASEP,$\epsilon$}(h, B) := \mathbb{E}[p^\text{TASEP,$\epsilon$}(\tau_{y_1} h, \sigma_r \tau_{y_2} B)].$$

(5.4)

Lemma 5.2. With the above notation,

$$\hat{p}^\text{TASEP,$\epsilon$}(f_0, B) - \hat{p}^\text{AEP,$\epsilon$}(f_0, B) = \int_0^t \int_{f_{t-s,\epsilon}} f(t, \epsilon) - L(\epsilon) \hat{p}^\text{TASEP,$\epsilon$}(h, B) d\nu \, ds,$$

(5.5)

Proof. After the Gaussian smoothing, (5.5) becomes

$$\hat{p}^\text{TASEP,$\epsilon$}(f_0, B) - \hat{p}^\text{AEP,$\epsilon$}(f_0, B) = \mathbb{E} \left[ \int_0^t \int_{f_{t-s,\epsilon}} f(t, \epsilon) - L(\epsilon) \hat{p}^\text{TASEP,$\epsilon$}(h, \sigma_r \tau_{y_1+y_2} B) d\nu \right].$$

(5.6)

Now

$$\int_{f_{t-s,\epsilon}} f(t, \epsilon) - L(\epsilon) \hat{p}^\text{TASEP,$\epsilon$}(h, \sigma_r \tau_{y_1+y_2} B) d\nu = \int (T_\epsilon - L(\epsilon)) \int_{f_{t-s,\epsilon}} f(t, \epsilon) \hat{p}^\text{TASEP,$\epsilon$}(h, \sigma_r \tau_{y_1+y_2} B) d\nu$$

and from the spatial homogeneity of TASEP,

$$\hat{p}^\text{TASEP,$\epsilon$}(h, \tau_{y_1+y_2} B) = \hat{p}^\text{TASEP,$\epsilon$}(\tau_{y_1} h, \tau_{y_2} B).$$

(5.7)

This gives (5.5).\hfill \Box

Lemma 5.3. There exists a constant $C > 0$ such that for any fixed $t > 0$, with the definitions above, and the Dirichlet form in (4.2),

$$\mathcal{D}_\epsilon(\hat{p}^\text{TASEP,$\epsilon$}(\cdot, \text{hyp}(g))) \leq Ce^{1/2}a^{-2}.$$ (5.9)

Proof. Let $B$ be of the form (2.16). If we write out the left hand side of (5.9) we get

$$e^{-3/2} \sum_{x \in \mathbb{Z}^d} \left( \mathbb{E}\left[ \nabla^\text{sym}_x p^\text{TASEP,$\epsilon$}(\tau_{y_1} h, \sigma_r \tau_{y_2} B) \right] \right)^2 d\nu,$$

(5.10)

where $\nabla^\text{sym}_x f(h) = f(h^{x,\text{sym}}) - f(h)$ from (2.19) with $v = 1$. Now,

$$\tau_{y_1} h^{x,\text{sym}} = \tau_{y_1}(h^{x+y_1,\text{sym}}),$$

(5.11)

the height shift $\sigma_r$ commutes with $h \mapsto h^{x,\text{sym}}$, and the TASEP transition probabilities are invariant under spatial and height shifts, i.e. $p^\text{TASEP,$\epsilon$}(h, \tau_y B) = p^\text{TASEP,$\epsilon$}(h, B)$ and $p^\text{TASEP,$\epsilon$}(h, \tau_y B) = p^\text{TASEP,$\epsilon$}(h, B)$. Note that the notations $h^{x+y_1,\text{sym}}$ above and $\nabla^\text{sym}_{x+y_1}$ below actually stand for $h^{x+y_1,\text{sym}}$ and $\nabla^\text{sym}_{x+y_1}$ respectively. So

$$\nabla^\text{sym}_x (p^\text{TASEP,$\epsilon$}(\tau_{y_1} h, \sigma_r \tau_{y_2} B)) = p^\text{TASEP,$\epsilon$}(\tau_{y_1} h^{x,\text{sym}}, \sigma_r \tau_{y_2} B) - p^\text{TASEP,$\epsilon$}(\tau_{y_1} h, \sigma_r \tau_{y_2} B)$$

(5.12)

$$= p^\text{TASEP,$\epsilon$}(\tau_{y_1} (h^{x+y_1,\text{sym}}), \sigma_r \tau_{y_2} B) - p^\text{TASEP,$\epsilon$}(\tau_{y_1} h, \sigma_r \tau_{y_2} B)$$

$$= p^\text{TASEP,$\epsilon$}(\sigma_r (h^{x+y_1,\text{sym}}), \tau_{y_2+y_1} B) - p^\text{TASEP,$\epsilon$}(\sigma_r h, \tau_{y_2+y_1} B).$$

(5.13)

So

$$\nabla^\text{sym}_x (p^\text{TASEP,$\epsilon$}(\tau_{y_1} h, \sigma_r \tau_{y_2} B)) = (\nabla^\text{sym}_{x+y_1} p^\text{TASEP,$\epsilon$})(\sigma_r h, \tau_{y_2+y_1} B).$$

(5.14)

We can rewrite the expectation over $y_1, y_2$ as the expectation in $y' = y_2 + y_1$ of the expectation of $y_1$ given $y'$. Now the distribution of $y_1$ given $y'$ is Gaussian, mean $y'/2$ and variance $a^2/2$. Furthermore,
by Jensen’s inequality, we can take the expectation over \( y' \) outside of the square. Putting this together, (5.10) is bounded above by

\[
\varepsilon^{-5/2} \mathbb{E} \left[ \int_{\mathbb{S}_\varepsilon} \varepsilon^{3/2} \sum_{x \in \mathbb{Z}, h(0) \in \varepsilon^{1/2} \mathbb{Z}} \left( \varepsilon^{3/2} \sum_{y_1 \in \mathbb{Z}, r \in \varepsilon^{1/2} \mathbb{Z}} \varphi(y_1, r) \nabla_{x+y_1}^{\text{sym}} p_{t}^{\text{TASEP}, \varepsilon} (\sigma_{-r}, \tau y, B) \right)^2 \, d\mu_\varepsilon \right],
\]

(5.15)

where the outside expectation is over \( y' \) and \( \varphi(y_1, r) \) is the orthogonal projection of the Gaussian density \( \frac{(y_1 - y'/2)^2}{\sqrt{\pi a^2}} \) onto \( L^2 \) of the discrete measure giving mass \( \varepsilon^{3/2} \) to each point of \( \frac{1}{2} \mathbb{Z} \times \varepsilon^{1/2} \mathbb{Z} \).

With \( \psi \) chosen as \( \psi(x, r) = \nabla_{x}^{\text{sym}} p_{t}^{\text{TASEP}, \varepsilon} (\sigma_{-r}, \tau y, B) \) and since \( \varphi(y_1, r) = \varphi(y_1, -r) \), it is easy to see that the term being squared in (5.15) is precisely \( \varphi \ast \psi \). By Young’s inequality

\[
\| \varphi \ast \psi \|_2 \leq \| \varphi \|_2 \| \psi \|_1
\]

applied on that discrete space, so we can bound

\[
\varepsilon^{3/2} \sum_{x \in \frac{1}{2} \mathbb{Z}, h(0) \in \varepsilon^{1/2} \mathbb{Z}} \left( \varepsilon^{3/2} \sum_{y_1 \in \mathbb{Z}, r \in \varepsilon^{1/2} \mathbb{Z}} \varphi(y_1, r) \nabla_{x+y_1}^{\text{sym}} p_{t}^{\text{TASEP}, \varepsilon} (\sigma_{-r}, \tau y, B) \right)^2
\leq C_1 a^{-2} \left( \varepsilon^{3/2} \sum_{x \in \frac{1}{2} \mathbb{Z}, h(0) \in \varepsilon^{1/2} \mathbb{Z}} \left| \nabla_{x}^{\text{sym}} p_{t}^{\text{TASEP}, \varepsilon} (h, \tau y, B) \right| \right)^2,
\]

where the factor \( C_1 a^{-2} \) is the square of the \( L^2 \) norm of \( \varphi \), and we used that the \( L^2 \) norm of the projection is bounded by the full \( L^2 \) norm. Together with (5.2) this gives (5.9).

6. CONVERGENCE STARTING FROM FINITE ENERGY INITIAL DATA

Let \( p_{t}^{\text{TASEP}, \varepsilon} (h, B) \), \( p_{t}^{\text{AEP}, \varepsilon} (h, B) \) and \( p_{t}^{\text{WASEP}, \varepsilon} (h, B) \) denote the transition probabilities for TASEP, AEP and WASEP with generators (2.18), (2.17) and (2.21) respectively.

**Lemma 6.1.** Let \( B \) be of the form (2.16). Let \( p_{t}^{\text{TASEP}, \varepsilon} (f_0, \nu, \varepsilon, B), p_{t}^{\text{AEP}, \varepsilon} (f_0, \nu, \varepsilon, B) \) and \( p_{t}^{\text{WASEP}, \varepsilon} (f_0, \nu, \varepsilon, B) \) denote the slightly smoothed out transition probabilities as in (5.3), by Gaussians with mean 0 and standard deviation \( a \). Then there is a \( C < \infty \) depending only on the jump law \( p(\cdot) \), such that

\[
\left| p_{t}^{\text{AEP}, \varepsilon} (f_0, \nu, \varepsilon, B) - p_{t}^{\text{TASEP}, \varepsilon} (f_0, \nu, \varepsilon, B) \right| \leq C a^{-1} \| f_0, \nu, \varepsilon \|_2 \varepsilon^{1/4}
\]

(6.1)

and

\[
\left| p_{t}^{\text{WASEP}, \varepsilon} (f_0, \nu, \varepsilon, B) - p_{t}^{\text{TASEP}, \varepsilon} (f_0, \nu, \varepsilon, B) \right| \leq C a^{-1} \| f_0, \nu, \varepsilon \|_2 \delta^{1/2}.
\]

(6.2)

**Proof.** Lem. (5.2) applies in both cases. By the strong sector condition, Lem (4.2) and the equivalence of Dirichlet forms, Lem (4.1) we can replicate (5.10) and use Lem. (5.3) to obtain the desired bounds.

From here, we have the following proposition.

**Proposition 6.2.** Suppose that \( f_0, \nu, \varepsilon \) d\nu \rightarrow d\rho \) with \( \varepsilon^{1/4} \| f_0, \nu, \varepsilon \|_2 \rightarrow 0 \). Then,

\[
\lim_{\varepsilon \rightarrow 0} p_{t}^{\text{AEP}, \varepsilon} (f_0, \nu, \varepsilon, \text{hyp} (g)) = p_{t}^{\text{FP}} (\rho, \text{hyp} (g)).
\]

(6.3)

uniformly for \( g \) uniformly continuous on \( \mathbb{R} \) with \( g(x) \geq C (1 + |x|) \).

**Proof.** If \( B \) is a set of the type (2.16) where \( g \) is uniformly continuous on all of \( \mathbb{R} \), the difference of \( \tau \varepsilon B \) and \( B \) can be made arbitrarily small by choosing \( z \) sufficiently small. In other words, given \( \gamma > 0 \), for \( z \) sufficiently small, \( \text{hyp} (g - \gamma) \subset \tau z B \subset \text{hyp} (g + \gamma) \). Thus, for any fixed constant \( \eta > 0 \), we fix \( a \) to be sufficiently small so that it follows from (6.1) by taking \( \varepsilon \rightarrow 0 \) that

\[
\liminf_{\varepsilon \rightarrow 0} p_{t}^{\text{AEP}, \varepsilon} (f_0, \nu, \varepsilon, \text{hyp} (g)) \geq \liminf_{\varepsilon \rightarrow 0} p_{t}^{\text{TASEP}, \varepsilon} (f_0, \nu, \varepsilon, \text{hyp} (g - 2 \gamma)) - \eta.
\]

(6.4)
As this holds for all $\eta > 0$, we have
\begin{equation}
\liminf_{\varepsilon \to 0} p_t^{AE, \varepsilon} (f_0, \varepsilon \nu, \text{hyp}(g)) \geq \liminf_{\varepsilon \to 0} p_t^{TASEP, \varepsilon} (f_0, \varepsilon \nu, \text{hyp}(g - 2\gamma))
\end{equation}
and similarly
\begin{equation}
\limsup_{\varepsilon \to 0} p_t^{AE, \varepsilon} (f_0, \varepsilon \nu, \text{hyp}(g)) \leq \limsup_{\varepsilon \to 0} p_t^{TASEP, \varepsilon} (f_0, \varepsilon \nu, \text{hyp}(g + 2\gamma)).
\end{equation}
Since we know from [20] that $p_t^{TASEP, \varepsilon} (f_0, \varepsilon \nu, \text{hyp}(g \pm 2\gamma)) \to p_t^{FP} (\rho, \text{hyp}(g \pm 2\gamma))$, the KPZ fixed point transition probabilities, and (6.5) and (6.6) hold for all $\gamma > 0$, we thus conclude the required proposition.

**Remark 6.3.** The initial data is generic in the following sense. $f_0, \varepsilon \nu \to d\rho$ with $\|f_0, \varepsilon \nu\|_2 \to 0$, could, for example, follow closely some nice deterministic function inside a box $[-L, L]$ and then be Brownian motion outside, with $f_0, \varepsilon \nu$ looking like an appropriate discretization. Note however that if the function $g$ is not chosen to go to infinity quickly enough outside the box, the right hand side of (6.3) may vanish, and the Proposition, while true, provides little information. This justifies the growth condition in $g$.

**Remark 6.4.** It is easy to see that everything above goes through if we replace the target set $B$ by sets $\text{epi} (g)$ where $\text{epi} (g) = \{ h : h(x) \geq g(x), x \in \mathbb{R} \}$ for any uniformly continuous $g$ such that $g(x) \leq C(1 - |x|)$ for some $C > 0$. The only change happens in (5.1), where the arg max gets replaced by arg min. Hence (5.2) still holds and we get the same statement as (6.3) with $\text{hyp}(g)$ replaced by $\text{epi}(g)$ for $g(x) \leq C(1 - |x|)$ for some $C > 0$.

Next we draw conclusions from (6.2). Under TASEP $dh_{\varepsilon} = -4\varepsilon^{-1} 1_{\lambda} dt + dM$ where $M$ is a martingale, while under WASEP, $dh_{\varepsilon} = [4\delta \varepsilon^{-3/2} (1_{\nu} - 1_{\lambda}) - 4\varepsilon^{-1} 1_{\lambda}] dt + d\tilde{M}$ where $\tilde{M}$ is another martingale. The 4 in TASEP is because the height function jumps by $2\varepsilon^{1/2}$ at rate $2\varepsilon^{-3/2}$. On the lattice $\frac{1}{2}\varepsilon \mathbb{Z}$,
\begin{equation}
41_{\lambda} = -\frac{\varepsilon}{4} \nabla - h \nabla^+ h + \frac{\varepsilon^3}{4} \nabla^+ - \nabla^+ h + 1,
\quad 41_{\nu} = -\frac{\varepsilon}{4} \nabla - h \nabla^+ h + \frac{\varepsilon^3}{4} \nabla^+ - \nabla^+ h + 1,
\end{equation}
where $\nabla^+ h(x) = (\varepsilon / 2)^{-1} (h(x) - h(x + \varepsilon / 2))$, $\nabla^+ h(x) = (\varepsilon / 2)^{-1} (h(x + \varepsilon / 2) - h(x))$, and the martingales are approximating white noises, one can see that these are discretizations of the KPZ equation. At the same time it is clear that for TASEP the factor in front of the second order discrete derivative is too small ($\varepsilon^{-3/2}$ instead of $\varepsilon^{-2}$) and thus there is no way to scale TASEP to the KPZ equation. From the discrete equation for WASEP one can read off the following result. Note that the Itô factor $t/12$ comes from the difference of the formal expression for the KPZ equation, and the Cole-Hopf solution.

**Proposition 6.5 (Bertini, Giacomin [3]).** Let $h_{\varepsilon} (t, x)$ be the height function at time $t$ of WASEP, $\varepsilon, \delta$. Suppose that $h_{\varepsilon} (0, x) \to h_0 (x)$ where $h_0 (x)$ is a continuous function with $h(x) \leq C(1 + |x|)$ for some $C < \infty$. Then $h_{\varepsilon} (t, x) + \varepsilon^{-1} t + t/12$ converges to the Cole-Hopf solution $h(t, x)$ of the KPZ equation (1.1) with $(\lambda, \nu, D) = (1/4, \delta/4, \delta)$ starting from $h_0 (x)$.

The formal argument above makes it clear why the global height shift $\varepsilon^{-1} t$ is the same in all our models. In (6.2), we first take the limit as $\varepsilon \downarrow 0$. The conclusion is that as $\varepsilon \downarrow 0$, if $f_0, \varepsilon \nu \to d\rho$ the left hand side of (6.2) converges to $|p_t^{KPZ, \delta} (\rho, B) - p_t^{FP} (\rho, B)|$ and therefore,
\begin{equation}
|p_t^{KPZ, \delta} (\rho, B) - p_t^{FP} (\rho, B)| \leq C \delta^{1/2} a^{-1} \limsup_{\varepsilon \to 0} \|f_0, \varepsilon \nu\|_2.
\end{equation}
Finally, we take the limit as $\delta \downarrow 0$. Again, as in the proof of Proposition 6.2, if $g$ is uniformly continuous the averaging makes an arbitrarily small error. So we have

**Proposition 6.6.** Suppose that $f_0, \varepsilon \nu \to d\rho$ with $\|f_0, \varepsilon \nu\|_2 \leq C_1$ for some universal constant $C_1$ (not depending on $\varepsilon$), and $g$ is uniformly continuous on all of $\mathbb{R}$ with $g(x) \geq C(1 + |x|)$ for some $C > 0$. Then,
\begin{equation}
\lim_{\delta \to 0} p_t^{KPZ, \delta} (d\rho, \text{hyp}(g)) = p_t^{FP} (d\rho, \text{hyp}(g)).
\end{equation}
The rest of the work of the paper is to use properties of ASEP and the KPZ equation to extend the result to distributional convergence in the uniform-on-compact topology, and broaden the class of initial data. In the original version (and the published version) this was based on an erroneous claim that these models were skew-time reversal invariant. Note that this was not used except after this point. In the following, the error is corrected.

7. Deterministic initial conditions for ASEP

We now begin the proof of Thm 1. First we obtain the result for narrow wedges, then a lower and upper bound for general data. In the following $h(t, x; h_0)$ denotes the KPZ fixed point starting at $h_0$ and $h_{\varepsilon}(t, x; h, 0)$ the (sped-up) ASEP evolution starting from $h_{\varepsilon, 0}$.

7.1. Finite energy approximations. For any $k \in \mathbb{N}, \gamma > 0, y_1 < y_2 < \ldots < y_k, b_1, b_2, \ldots, b_k \in \mathbb{R}$, let $g(x)$ be the continuous piecewise linear function such that $g(x) = b_1$ for $x \leq y_1$, $g(x) = b_k$ for $x \geq y_k$ and $g(x), y_i \leq x \leq y_{i+1}$ is the line segment joining $b_i$ and $b_{i+1}$ for $i = 1, \ldots, k - 1$. The randomized version is $g(x) + \Theta(x)$ where $\Theta(x), x \in \mathbb{R}$ is the following random process: $\Theta(y_i)$ are uniformly distributed on $[0, \gamma], i = 1, \ldots, k$; $\Theta(y_i - x), x > 0$, and $\Theta(y_k + x), x > 0$, are Brownian motions, and $\Theta(x), y_i < x < y_{i+1}$ are Brownian bridges, all of them independent of each other. So $g(x) + \Theta(x)$ is a continuous randomized version of $g(x)$, deviating from $g(x)$ by at most $\gamma$ at the points $y_i$, by Brownian bridges in between the $y_i$ and by Brownian motions at either end. On $\mathbb{S}_\varepsilon$, we use the natural analogue with nearest neighbour random walks (see Appx. A) for the precise definition. The Radon-Nikodym derivative with respect to a two sided, symmetric random walk is called $g(x)$.

Lemma 7.1. Let $a = \max |b_{i+1} - b_i|/|y_{i+1} - y_i|$ be the maximal slope and $B = b_k - b_1$ be the length of the approximation interval. There exists $C, c < \infty$ such that

$$\|f_{0, \varepsilon}\|L^2(d\nu) \leq Ce^{ca^2B}. \quad (7.1)$$

7.2. Narrow wedges. For $y \in \mathbb{R}$ and $b \in \mathbb{R}$ the narrow wedge of height $b$ at $y$ for the KPZ fixed point is

$$\mathcal{D}_y^b(x) = b \text{ if } x = y, \quad \mathcal{D}_y^b(x) = -\infty \text{ otherwise.}$$

For $y_1 < y_2 < \ldots < y_k, b_1, b_2, \ldots, b_k \in \mathbb{R}$ the multiple narrow wedge $\mathcal{D}_{\gamma, s}^b$ is $\mathcal{D}_{\gamma, s}^b(x) = \max_i \mathcal{D}^{y_i, b_i}(x)$.

These have deterministic approximations the narrow wedge of height $b$ at $x$ is $\mathcal{D}_{s,L}^b(x)$ is $\max\{b - s|x - y|, -L\}$ and the multiple narrow wedge is $\mathcal{D}_{s,L}^b(x) = \max_i \mathcal{D}_{s,L}^{y_i, b_i}(x)$.

The Brownian multiple narrow wedge $\mathcal{D}_{\gamma, s}^b$ is the randomized version of this piecewise linear function as in Sec. 7.1. On $\mathbb{S}_\varepsilon$, we use a straightforward analogue with nearest neighbour random walks and call it $\mathcal{D}_{\gamma, s}^b$.

Prop. 6.2 together with Lem. 7.1 shows convergence to the KPZ fixed point starting from $\mathcal{D}_{\gamma, s}^b$, if we start ASEP with $\mathcal{D}_{\gamma, s}^b$; if

$$\lim_{\varepsilon \to 0} P(h_{\varepsilon}(t, \cdot; \mathcal{D}_{\gamma, s}^b, \varepsilon) \leq g) = P(h(t, \cdot; \mathcal{D}_{\gamma, s}^b) \leq g) \quad (7.2)$$

We now work to extend this convergence to initial data true multiple narrow wedges, which have slopes $\pm \varepsilon^{-1/2}$ instead of the fixed $s$ as $\varepsilon \to 0$, i.e. $\mathcal{D}_{\gamma, s = \gamma \varepsilon^{-1/2}}^b$. From the continuity of the KPZ fixed point we have that $\lim_{s, L \to \infty} P(h(t, \cdot; \mathcal{D}_{\gamma, s}^b) \leq g) = P(h(t, \cdot; \mathcal{D}_{\gamma, s}^b) \leq g)$ and hence

$$\lim_{s, L \to \infty} \lim_{\varepsilon \to 0} P(h_{\varepsilon}(t, \cdot; \mathcal{D}_{\gamma, s}^b, \varepsilon) \leq g) = P(h(t, \cdot; \mathcal{D}_{\gamma, s}^b) \leq g) \quad (7.3)$$
The $S_\epsilon$ randomized multiple narrowish wedge dominates the multiple narrow wedge,
\begin{equation}
\partial_{\gamma,s,L,\epsilon}^{\overrightarrow{y,b}} \geq \partial_{\gamma,s,L,\epsilon}^{\overrightarrow{y,b}} \gamma,s=\epsilon^{-1/2},L=-\infty,\epsilon,\gamma \notag \end{equation}
and hence at all later times the discrepancy process is non-negative,
\begin{equation}
\Delta_{s,L,\epsilon}(t,x) := h_\epsilon(t,x;\partial_{\gamma,s,L,\epsilon}^{\overrightarrow{y,b}}) - h_\epsilon(t,x;\partial_{\gamma,s,\epsilon^{-1/2},L=-\infty,\epsilon}^{\overrightarrow{y,b}}) \geq 0. \tag{7.5}
\end{equation}

Here we have used the basic coupling of different initial data and the same randomness in the $\gamma$.

Consider first a single narrow wedge at $y$ and its approximation. From [28] we have one point convergence starting from the narrow wedge, i.e. for fixed $x$,
\begin{equation}
\lim_{\gamma \to 0} \lim_{\epsilon \to 0} P(h_\epsilon(t,x;\partial_{\gamma,s,L,\epsilon}^{\overrightarrow{y,b}}) \leq g(x)) = P(h(t,x;\partial_{y,b}) \leq g(x)). \tag{7.6}
\end{equation}

From the one point single wedge case of (7.3) we conclude that \( \lim_{s,L \to \infty} \lim_{\epsilon \to 0} P(\Delta_{s,L,\epsilon}(t,x) > 0) = 0 \) and therefore by the union bound that
\begin{equation}
\lim_{s,L \to \infty} \lim_{\epsilon \to 0} P(\Delta_{s,L,\epsilon}(t,x_i) > 0, i = 1, \ldots, m) = 0. \tag{7.7}
\end{equation}
This means we have convergence in finite dimensional distributions starting from the single narrow wedge
\begin{equation}
\lim_{\gamma \to 0} \lim_{\epsilon \to 0} P(h_\epsilon(t,x_i;\partial_{\gamma,s,L,\epsilon}^{\overrightarrow{y,b}}) \leq g(x_i), i = 1, \ldots, m) = P(h(t,x_i;\partial_{y,b}) \leq g(x_i), i = 1, \ldots, m). \tag{7.8}
\end{equation}

Theorem 1.1 of Corwin-Dimitrov [6] gives us in addition the tightness of the rescaled ASEP height function at fixed time $t > 0$ in the uniform-on-compact topology, upgrading (7.8) to uniform convergence on compact sets.

We next construct a half-randomized wedge at 0, $\phi_{\gamma,s,L,\epsilon}^{0,0}$, which has slope $-\epsilon^{-1/2}$, i.e. is a true narrow wedge, to the right of 0, but coincides with the randomized narrowish wedge on the left of 0. Throughout the evolution, it is trapped above the corresponding true narrow wedge and below the corresponding randomized narrowish wedge,
\begin{equation}
h_\epsilon(t,x;\phi_{\gamma,s,L,\epsilon}^{0,0}) \geq h_\epsilon(t,x;\phi_{\gamma,s,L,\epsilon}^{0,0}) \geq h_\epsilon(t,x;\phi_{\epsilon^{-1/2,\infty}}^{0,g}), \tag{7.9}
\end{equation}
so it also converges to the fixed point starting with a continuum narrow wedge. In particular, the discrepancy between the solutions starting from the half-randomized wedge and the narrow wedge converges to zero uniformly. This discrepancy can be built in the following way. Consider the particles in the randomized narrow wedge as first class particles. Then add second class particles on the left of 0 to lower the height function there to the narrow wedge. The change in the discrepancy process at $x$ up to time $t$ is equal to the flux of these second class particles across $x$ in the positive direction during that time period. The presence of first class particles produces a drift of second class particles to the left. Therefore if we increase the number of first class particles to the right of 0 the flux is decreased (see Lem. [3.1]). From this we conclude that if we compare starting with randomized narrowish wedge to the left of 0 and narrow wedge to the left of 0, both with any configuration of first class particles to the right of 0, the resulting discrepancy process is asymptotically trivial. By symmetry the same thing holds if do this on the right side, and by translation invariance we could replace 0 by any point $y$ and height $b$.

Since $\gamma > 0$ is arbitrary, Prop. [5.2] implied that we have convergence (in the sense of finite dimensional distributions) to the KPZ fixed point starting from any continuous function on a finite interval $[a,b]$ as long as we pasted half-randomized wedges at the ends, i.e. on the right with slope $-s$ down to $-L$, then a free Brownian motion, and the flipped picture on the left. The preceding argument means we now have proven the same convergence if we put (deterministic) slope $-\epsilon^{-1/2}$ on the right of $b$ and $\epsilon^{-1/2}$ on the left of $a$ (i.e. the minimal boundary conditions). In particular,

**Proposition 7.2.** Let $h_0$ be continuous on $[a,b]$ and $-\infty$ on $[a,b]^c$, and $h_{\epsilon,0} \in S_\epsilon$ converging to $h_0$ uniformly on $[a,b]$. Then
\begin{equation}
\limsup_{\epsilon \to 0} P(h_\epsilon(t,x_i;h_{\epsilon,0}) \leq r_i, i = 1, \ldots, m) \leq P(h(t,x_i;h_0) \leq r_i, i = 1, \ldots, m). \tag{7.10}
\end{equation}
We next argue that we can start ASEP with true multiple narrow wedges $\delta \bar{g}, \tilde{b} = \epsilon^{-1/2}, L = -\infty$. With this initial data, run this process up to small time $\delta$. We know from the result for the single narrow wedge with a non-zero $\gamma$ that if we choose $\gamma \gg \delta^{1/2}$, as $\epsilon \to 0$ with probability approaching 1, $h_\epsilon(t = \delta, x) \geq g(x) := \max_{1 \leq j \leq k} \{ b_j - \delta^{-1}[x - y_j]^2 \} - \gamma$ on compact sets of $x$. Choose a large compact set $[-L, L]$ and replace $g$ outside of $[-L, L]$ by slopes $-\epsilon^{-1/2}$ on the right and $\epsilon^{-1/2}$ on the left to obtain $g_{L, \epsilon}$, a lower bound for $h_\epsilon(t = \delta, \cdot)$ everywhere. But we know that as $\epsilon \to 0$ then $L \to \infty$ and then $\delta \to 0$, the rescaled ASEP with initial data $g_{L, \epsilon}$ converges to the KPZ fixed point starting from $\delta \bar{g}, \tilde{b}$. This provides a lower bound limit of the multiple narrow wedge. The corresponding upper bound is provided by randomized narrowish wedges. Hence starting with multiple narrow wedges we have convergence in finite dimensional distribution to the KPZ fixed point starting with the corresponding continuum multi-narrow wedge.

**Proposition 7.3.**

$$\lim_{\epsilon \to 0} P(h_\epsilon(t, x_i; \delta \bar{g}, \tilde{b})_{s=\epsilon^{-1/2}, L = -\infty}) \leq r_i, \ i = 1, \ldots, m) = P(h(t, x_i; \delta \bar{g}, \tilde{b}) \leq r_i, \ i = 1, \ldots, m).$$

From this we obtain immediately from the variational formula the general lower bound. Keep in mind the lower bound is saying that the discrete approximations are statistically dominating the fixed point in the limit.

**Proposition 7.4.** Let $h_{\epsilon, 0} \to h_0 \in UC$. Then

$$\limsup_{\epsilon \to 0} P(h_\epsilon(t, x_i; h_{\epsilon, 0}) \leq r_i, \ i = 1, \ldots, m) \leq P(h(t, x_i; h_0) \leq r_i, \ i = 1, \ldots, m), \quad (7.11)$$

i.e. in the limit the ASEP height functions are distributionally lower bounded by the fixed point.

**Proof.** Let $\delta > 0$, $t > 0$, and $x_1, \ldots, x_m$ be fixed. Since $h_0 \in UC$ and using the variational formula for the KPZ fixed point, we can find $k$, $y_\epsilon^1, \ldots, y_\epsilon^k$ and $\alpha_\epsilon^1, \ldots, \alpha_\epsilon^k$ such that $h_{\epsilon, 0}(x) = \sum_{i=1}^k a_i - \epsilon^{-1/2} |x - y_j|$ satisfies $h_{\epsilon, 0} \leq h_{\epsilon, 0}$ everywhere and $\liminf_{\epsilon \to 0} h(t, x_i; h_{\epsilon, 0}) \geq h(t, x_i; h_0) - \delta$, for $i = 1, \ldots, m$. Since we have convergence of ASEP starting with such multiple narrow wedges, by taking $\delta$ small we obtain 7.11 from the ordering property of the height functions of ASEP.

### 7.3. **Upper bound.**

Consider ASEP with first and second particles with the following initial condition: No particles at sites $x \geq 0$. At $x < 0$, we put a particle with probability $\frac{1}{2}$ and then independently let it be first class with probability $1 - \rho \epsilon^{1/2}$ and second class with probability $\rho \epsilon^{1/2}$. The corresponding rescaled initial height functions are $h_\epsilon(x) = \epsilon^{1/2} h_\epsilon(\epsilon^{-1} x)$ for all particles and $h_\epsilon$ for first class particles, living on $\epsilon \mathbb{Z}$, or $\mathbb{R}$, by linear interpolation. Both height functions are equal to $-\epsilon^{-1/2} x$ for $x \geq 0$. For $x < 0$, $h_\epsilon$ is a symmetric random walk taking jumps of size $\epsilon^{1/2}$ at times $\epsilon \mathbb{Z}$ with $h_\epsilon(0) = 0$. $h_\epsilon$ is similar but with a drift $-\rho$. Since they both vanish at the origin, we have coupled them so that $h_\epsilon(x) \leq h_\epsilon(x)$ and this remains so throughout the evolution.

**Lemma 7.5.** In the above setup, for any $0 < \rho < \infty$, if $x_\epsilon \to \infty$, for each $t > 0$, $\bar{h}_\epsilon(t, x_\epsilon) - h_\epsilon(t, x_\epsilon) \to 0$ in distribution.

**Proof.** The setup is ASEP with step Bernoulli initial conditions in the anti-shock (rarefaction) regime and an exact determinantal formula was obtained by Tracy and Widom in [29] and asymptotics worked out in Theorems 2 and 3 of that article. Their asymptotics give that in our setting both $h_\epsilon(t, x_\epsilon) + \frac{1}{2} x_\epsilon^2$ and $h_\epsilon(t, x_\epsilon) + \frac{1}{2} x_\epsilon^2$ converge in distribution to $t^{1/2}$ times the Tracy-Widom GUE distribution (as long as $x_{\epsilon} \to \infty$). Since they are coupled, ordered and have the same limiting distribution the difference converges to 0 in distribution.

Let $h_0$ be a continuous function with $h_0(x) \leq C(1 + |x|)$ and suppose we start our ASEP height functions with a measure $f_{0,\epsilon} \mu_{\epsilon}$ satisfying the energy bound $\| f_{0,\epsilon} \|_{L^2(\mu_{\epsilon})} = \alpha(\epsilon^{-1/4})$ and such that $h_\epsilon(0, x) \to h_0(x)$ locally uniformly. Our basic result for randomized initial data says that $h_\epsilon(t, x) \to h(t, x)$, the KPZ fixed point at time $t > 0$ (in distribution, in the sense of uniform convergence on compact sets). To satisfy the energy bound, by Lem. 7.1 one could choose the
height functions driftless Brownian motions outside a box \([-x_\varepsilon, x_\varepsilon]\) with \(x_\varepsilon \sim c|\log \varepsilon|\) as long as \(c\) is sufficiently small.

We now show that we can instead take the Brownian motions to have drift \(\rho > 0\) on \((x_\varepsilon, \infty)\) and \(-\rho\) on \((-\infty, -x_\varepsilon)\). To this end, let \(\tilde{f}_{0,\varepsilon}\) and \(\nu_\varepsilon\) be the marginals of \(f_{0,\varepsilon}, \nu_\varepsilon\) on the box \([-x_\varepsilon, x_\varepsilon]\), and define a new measure \(\rho_\varepsilon\) to correspond to those marginal distribution on the box and the drifted Brownian motions outside that box.

**Lemma 7.6.** In the setup above with drifts \(\pm \rho\), as long as the energy inside the box, \(\| \tilde{f}_{0,\varepsilon} \|_{L^2(\nu_\varepsilon)} = o(\varepsilon^{-1/4})\), we still have that \(h_\varepsilon(t, x)\) converges to the KPZ fixed point in \(\rho_\varepsilon\) probability.

**Proof.** To compare the driftless Brownian motion on \((-\infty, -x_\varepsilon)\) to the one with drift \(-\rho\), we couple them so that the latter has first class particles and the former is built from it by adding second class particles. There are no second class particles on \([-x_\varepsilon, \infty)\) and the height functions coincide there. The difference in height functions just corresponds to a height function built out of the second class particles, and by Lem. 7.5 and Lem. B.1, if we look at \(\bar{h}_\varepsilon(t, x_i) - \tilde{h}_\varepsilon(t, x_i), i = 1, \ldots, m\), with \(x_i \in \mathbb{R}\) fixed (i.e. independent of \(\varepsilon\), the random vector goes to 0 in distribution. Now we make the same comparison for our height functions on \((+x_\varepsilon, \infty)\) to obtain the result.

**Proposition 7.7.** Let \(h_{\varepsilon,0} \to h_0\) in UC. Then

\[
\liminf_{\varepsilon \to 0} P(h_\varepsilon(t, x_i; h_{\varepsilon,0}) \leq r_i, i = 1, \ldots, m) \geq P(h(t, x_i; h_0) \leq r_i, i = 1, \ldots, m),
\]

(7.12)

i.e. in the limit the ASEP height functions are distributionally upper bounded by the fixed point.

**Proof.** Let \(\delta > 0\). Since \(h_0\) is upper semi-continuous and by continuity in UC of the KPZ fixed point, there is a continuous function \(h_0 \geq h_0\) with \(h(t, x_i; h_0) \leq h(t, x_i; h_0) + \delta\). Hence we can assume without loss of generality that \(h_0\) is continuous. Let \(C < \infty\) so that the convergence takes place in \(UC\), in particular, we have a bound \(h_{\varepsilon,0}(x) \leq C(1 + |x|)\).

Let \(x_\varepsilon' < x_\varepsilon\) both going to infinity, but no faster than \(c \log \varepsilon\) so as to keep the energy bound needed in Prop. 6.2 by Lem. 7.1. We can build a measure \(\rho_\varepsilon\) on height functions such that: 1. The marginal \(\tilde{f}_{0,\varepsilon}d\nu_\varepsilon\) on \([-x_\varepsilon, x_\varepsilon]\) satisfies \(\| \tilde{f}_{0,\varepsilon} \|_{L^2(\nu_\varepsilon)} = o(\varepsilon^{-1/4})\), 2. With \(\rho_\varepsilon\) probability converging to 1, the height functions \(\tilde{h}_\varepsilon(t, x)\) under \(\rho_\varepsilon\) satisfy \(\tilde{h}_\varepsilon(x) - \tilde{h}_{\varepsilon,0}(x) < \delta\) for all \(x \in [-x_\varepsilon', x_\varepsilon']\), \(\tilde{h}_\varepsilon(x) > h_{\varepsilon,0}(x)\) for all \(x \in [-x_\varepsilon, x_\varepsilon]\), and \(\tilde{h}_\varepsilon(\pm x_\varepsilon) > (C + 1)(1 + |x_\varepsilon|)\). 3. \(\tilde{h}_\varepsilon(x)\) is Brownian motion with drift \(\rho x\) for \(x > x_\varepsilon\) starting at \(\tilde{h}_\varepsilon(x_\varepsilon)\), and analogously on \((-\infty, -x_\varepsilon)\), with a picture reflected across 0 and an independent Brownian motion with drift \(-\rho\). Here we choose some \(\rho > C\) (for example, \(\rho\) could be \(C + 1\)).

Since the probability that a Brownian motion with drift \(\rho > C\) starting at a point \(x_\varepsilon \to \infty\) ever hits the line \(Cx\) goes to 0 as \(\varepsilon \to 0\), we have that \(\tilde{h}_\varepsilon(x) > h_{\varepsilon,0}(x)\) for all \(x \in \mathbb{R}\) with probability going to 1 as \(\varepsilon \to 0\). Furthermore \(\tilde{h}_\varepsilon(t, x)\) converges to the KPZ fixed point in \(\rho_\varepsilon\) probability by the previous lemmas. But since initially \(\tilde{h}_\varepsilon\) is not more than \(\delta\) larger than \(h_{\varepsilon,0}\) on an interval \([-x_\varepsilon', x_\varepsilon']\) \(\to \mathbb{R}\) we have shown that \(\liminf_{\varepsilon \to 0} P(h_\varepsilon(t, x_i; h_{\varepsilon,0}) \leq r_i, i = 1, \ldots, m) \geq P(h(t, x_i; h_0) \leq r_i - \delta, i = 1, \ldots, m)\). Since this is true for any \(\delta > 0\) the upper bound follows.

### 8. Bootstrapping KPZ using uniform Hölder continuity

Though the results in Sec. E and Sec. 7 would also work for the KPZ equation, the KPZ equation enjoys the additional property that it satisfies a positive temperature version of the variational formula (8.2). This allows us to give an easier proof of the extension from finite energy to general initial data.

The KPZ,\(\delta\) proto-Airy sheet \(A_{\delta,t}(x, y)\) is defined as follows. First of all we consider the white noise in KPZ,\(\delta\) (2.10) to be fixed and solve with any admissible data using that same white noise, to produce a stochastic flow on the space of admissible data. Note that the solution of KPZ,\(\delta\) with nice initial data \(h(0, x) = h_0(x)\) is simply defined as \(h(t, x) = 4\delta \log Z(t, x)\) where \(Z(t, x)\) solved the stochastic heat equation

\[
\partial_t Z = \delta \partial_x^2 Z + \frac{1}{4} \delta^{1/2} \xi Z,
\]

(8.1)
with initial data \( Z(0, x) = e^{(4\delta)^{-1} h_0(x)} \). The "initial data" narrow wedge at \( y \) for KPZ, \( \delta \), means \( h(t, x) = 4\delta \log Z(t, x) \) where \( Z(0, x) = \delta_y(x) \). We call the process with the parabola removed \( h(t, x) + \frac{1}{4}(x - y)^2 := A_{\delta,t}(x, y) \). By linearity in the initial data of the stochastic heat equation (8.1) we have for any nice initial data \( h_0(x) \),

\[
  h_\delta(t, y; h_0) = 4\delta \log \int \exp \left\{ (4\delta)^{-1} (A_{\delta,t}(x, y) - \frac{1}{4}(x - y)^2 + h_0(x)) \right\} dx.
\]

where \( h_\delta(t, y; h_0) \) is the solution of KPZ, \( \delta \) starting from \( h_0 \).

Let \( || \cdot ||_{\alpha, B} \) denote the Hölder norm \( || f ||_{\alpha, B} = \sup_{x_1, x_2} |f(x_2) - f(x_1)|/|x_2 - x_1|^\alpha \) where the supremum is over \([-B, B]\) or \([-B, B]^2\), depending on whether the function has one or two real variables.

**Lemma 8.1.** (Corwin-Hammond[8])

1. For any \( 0 < \alpha < 1/2 \) and \( B < \infty \), \( \lim_{K \to \infty} \mathbb{P}(||A_{\delta,t}(x, y)||_{\alpha, B} \geq K) = 0 \) uniformly in \( \delta > 0 \).

2. For any \( b > 0 \), \( \lim_{A \to \infty} \mathbb{P}(A_{\delta,t}(x, y) \geq A + b(|x|^2 + |y|^2)) = 0 \) uniformly in \( \delta > 0 \).

**Proof.** Part (1) follows from the absolute continuity of the narrow wedge KPZ equation with respect to Brownian motion on any compact set and the tightness of the Radon-Nikodym derivative for all \( \delta > 0 \) (see Theorem 1.2 of [8]) and the symmetry in the two coordinates of \( A_{\delta,t}(x, y) \).

Part (2) follows from Lemma 4.1 of [8] and the symmetry in the two coordinates of \( A_{\delta,t}(x, y) \). \( \square \)

Suppose our initial data \( h_0 \) satisfies \( h_0(x) \leq A + C|x|^2 \) for all \( x \in \mathbb{R} \), for some \( b < (2t)^{-1} \). Let \( \sigma > 0 \). Then, with probability greater than \( 1 - \sigma \), we can choose \( B \) large enough that the integral in (8.2) over \(|x| > B\) is bounded by \( \sigma \), uniformly in \( \delta > 0 \) for \( y \) in a compact \([M, M]\). It is also immediate from (8.2) that for any \( 0 < \alpha < 1/2 \) and \( B < \infty \), \( \lim_{K \to \infty} \mathbb{P}(||h(t, \cdot)||_{\alpha, B} \geq K) = 0 \) uniformly in \( \delta > 0 \). This gives the tightness of \( h(t, \cdot) \).

Next, we want to show that \( h_\delta(t, \cdot; h_0) \) converges to the KPZ fixed point \( h(t, \cdot; h_0) \) in distribution in the topology of uniform-on-compact. Suppose our initial data \( h_0 \) satisfies \( h_0(x) \leq A + C|x| \) for some \( A, C \) and for all \( x \in \mathbb{R} \). Then, as in the last paragraph, we can truncate the initial condition in a compact interval \([-B, B]\) so that the integral in (8.2) over \(|x| > B\) is bounded by \( \sigma \), uniformly in \( \delta > 0 \) for all \( y \in [-M, M] \) with probability greater than \( 1 - \sigma \). Furthermore, by the same argument as at the beginning of the proof of Prop. [7.7] we can assume that \( h_0 \) is continuous.

Define a new initial condition \( \tilde{h}_0 \) from \( h_0 \) such that \( \tilde{h}_0(x) = \tilde{h}_0(x) \) for \( x \in [-B, B] \) and \( \tilde{h}_0(B + x) = h_0(B) + B(x) \) for \( x > 0 \) and \( \tilde{h}_0(-B - x) = h_0(B) + B(-x) \) for \( x > 0 \), where \( B \) is a two-sided Brownian motion passing through \( 0 \). Then by the last paragraph, for all \( y \in [-M, M] \) with probability greater than \( 1 - \sigma \),

\[
  |h_\delta(t, y; h_0) - h_\delta(t, y; \tilde{h}_0)| \leq 2\sigma
\]

and the same bound holds for the KPZ fixed point. The proof of the statement for the KPZ fixed point follows from the fact that the supremum in the variation formula for the fixed point is attained in a compact interval of \( y \) with high probability. This is proved in Lemma [E.3]

Now consider two randomized versions \( \tilde{h}_0, \tilde{h}_0 \) of \( h_0 \) (see beginning of Sec. [7.1]), such that for \( x \in [-B, B] \), with probability \( 1 - \sigma \), \( \tilde{h}_0(x) \geq h_0(x) \geq \tilde{h}_0(x) \) and \( \tilde{h}_0(x) = \tilde{h}_0(x) = \tilde{h}_0(x) \) for all \(|x| > B\). Running the coupled dynamics, we have with probability \( 1 - \sigma \) for all \( y, h_\delta(t, y; h_0) \geq h_\delta(t, y; \tilde{h}_0) \geq h_\delta(t, y; \tilde{h}_0) \) that is,

\[
  p_{t}^{\text{KPZ,} \delta} (\tilde{h}_0, \text{hyp} g) - \sigma \leq p_{t}^{\text{KPZ,} \delta} (h_0, \text{hyp} g) \leq p_{t}^{\text{KPZ,} \delta} (\tilde{h}_0, \text{hyp} g) + \sigma.
\]

For \( g \) as in Prop. [6.6] we have as \( \delta \to 0 \),

\[
  p_{t}^{\text{FP}} (h_0, \text{hyp} g) - \sigma \leq \liminf p_{t}^{\text{KPZ,} \delta} (h_0, \text{hyp} g) \leq \limsup p_{t}^{\text{KPZ,} \delta} (h_0, \text{hyp} g) \leq p_{t}^{\text{FP}} (h_0, \text{hyp} g),
\]

which tells us that \( \lim_{\delta \to 0} p_{t}^{\text{KPZ,} \delta} (h_0, \text{hyp} g) = p_{t}^{\text{FP}} (h_0, \text{hyp} g) \), and therefore by the approximation in the first paragraphs of the proof, that the same is true if we start with \( h_0 \). Together with the
tightness, this proves the convergence in distribution of \( h_\delta(t,\cdot;h_0) \) to the fixed point \( h(t,\cdot;h_0) \) in the uniform-on-compact topology.

Next we work towards multi-narrow wedge initial data, i.e. \( Z(0,x) = \delta^{\bar{x},\bar{a}}(x) \). An approximation of it in the \( \alpha \)-Hölder class satisfying \( h_\delta(x) \leq A - |x| \) is \( \delta^{\bar{x},\bar{a}}(\delta/2) \log(\delta/2)^{-1}s \) where \( \delta^{\bar{x},\bar{a}}(x) = \min\{\delta^{\bar{x},\bar{a}}(x), A - |x|\} \). Let \( \sigma > 0 \). On a set with probability greater than \( 1 - \sigma \), the \( \alpha \)-Hölder norm in \( y \) of \( A_{\delta,t}(x,y) \), on set \( [x - s^{-1}(a + L), x + s^{-1}(a + L)] \times [-B,B] \) is less than \( K \) and \( A_{\delta,t}(x,y) \leq A + b(|x|^2 + |y|^2) \) for some \( b < t^{-1} \). Here \( \bar{x}, \bar{a} \) are the smallest and largest \( x_i \) and \( a_i, \bar{a} \) are their corresponding \( a_i \). The second interval is just the smallest interval containing the wedge parts \( \delta^{\bar{x},\bar{a}}(\delta/2) \log(\delta/2)^{-1}s \). Since these wedge parts can vary at most \( r_{s,\delta,L} := \max_i s^{-1}(a_i + \delta \log \delta^{-1}s + L) \), we have, on this set,

\[
4\delta \log \left( \sum \mu_i e^{(4\delta)^{-1}(a_i + A_{\delta,t}(x_i,y) - \frac{1}{t}(x_i-y)^2 - Kr_{s,\delta,L})} + R_L(y) \right) \leq 4\delta \log \left( \sum \mu_i e^{(4\delta)^{-1}(a_i + A_{\delta,t}(x_i,y) - \frac{1}{t}(x_i-y)^2 + Kr_{s,\delta,L})} + R_L(y) \right),
\]

where \( \mu_i = \frac{1}{2}(4\delta)^{-1}s \int_{s^{-1}(a_i+L)}^{s^{-1}(a_i+L)} e^{-(4\delta)^{-1}s|x|}\,dx \) and \( R_L(y) := \int \exp \left\{ (4\delta)^{-1} \left( A_{\delta,t}(x,y) - \frac{1}{t}(x-y)^2 + (-L) \vee (A - |x|) \right) \right\} \,dx \).

Since \( A_{\delta,t}(x,y) \leq A + b(|x|^2 + |y|^2) \) for \( R_L \to 0 \) uniformly on compact sets of \( y \) as \( L \to \infty \). Letting \( \delta \to \infty \) followed by \( L, s \to \infty \) we have

\[
\limsup_{\delta \to 0} \mathbb{P}(\max_{i,j} |a_i + b_j + A_{\delta,t}(x_i,y_j) - \frac{1}{t}(x_i-y_j)^2| \leq 0) - \mu \log(\delta^{-1} \log(\delta^{-1})) \leq 0.
\]

Since \( \sigma > 0 \) is arbitrary, we have the convergence of the finite dimensional distributions of the KPZ,\( \delta \) starting with \( k \) narrow wedges to the KPZ fixed point. Together with the tightness of the KPZ,\( \delta \) height function, we get the distributional convergence in the uniform-on-compact topology. A similar argument proves the same result starting with \( k \) narrow wedges plus a continuous function bounded above by some \( A(1 + |x|) \). This proves the first statement in Thm. 2.2 (3).

9. Convergence of the KPZ line ensemble to the Airy line ensemble

Recall the KPZ ensemble from [8]. The Hamiltonian \( H_t : \mathbb{R} \to [0,\infty) \) is \( H_t(x) = e^{t^{1/3}x} \). Let \( \mathcal{L} = (\mathcal{L}_i, I_2, \ldots) \) be an \( \mathbb{N} \times \mathbb{R} \) indexed line ensemble. Fix \( k_1 \leq k_2 \) with \( k_1, k_2 \in \mathbb{Z} \), an interval \( (a,b) \subset \mathbb{R} \) and two vectors \( \bar{x}, \bar{y} \in \mathbb{R}^{k_2-k_1+1} \). Given two measurable functions \( f,g : (a,b) \to \mathbb{R} \cup \{\pm \infty\} \), the law \( \mathbb{P}_{H_t}^{k_1,k_2,(a,b),\bar{x},\bar{y},f,g} \) on \( \mathcal{L}_{k_1}, \ldots, \mathcal{L}_{k_2} : (a,b) \to \mathbb{R} \) has the following Radon-Nikodym derivative with respect to \( \mathbb{P}_{\text{free}}^{k_1,k_2,(a,b),\bar{x},\bar{y}} \), the law of \( k_2 - k_1 + 1 \) independent Brownian bridges taking values \( \bar{x} \) at time \( a \) and \( \bar{y} \) at time \( b \):

\[
\mathbb{P}_{H_t}^{k_1,k_2,(a,b),\bar{x},\bar{y},f,g} \left( \mathcal{L}_{k_1}, \ldots, \mathcal{L}_{k_2} \right) = \exp \left\{ -\sum_{i=k_1}^{k_2} \int_a^b H_t((\mathcal{L}_{i+1}(u) - \mathcal{L}_i(u)) \,du \right\}
\]

with \( \mathcal{L}_{k_1-1} = f \) and \( \mathcal{L}_{k_2+1} = g \), and \( Z_{H_t}^{k_1,k_2,(a,b),\bar{x},\bar{y},f,g} \) is the normalizing constant. We say that the line ensemble \( \mathcal{L} \) has the \( H_t \)-Brownian Gibbs property if for all \( K = \{k_1, \ldots, k_2\} \subset \mathbb{N} \) and \( (a,b) \subset \mathbb{R} \), the conditional distribution of \( \mathcal{L}|_{K \times (a,b)} \) given \( \mathcal{L}|_{(\mathbb{N} \setminus \mathbb{N}) \setminus K \times (a,b)} \) is \( \mathbb{P}_{H_t}^{k_1,k_2,(a,b),\bar{x},\bar{y},f,g} \). Here \( f = \mathcal{L}_{k_1-1} \) and \( g = \mathcal{L}_{k_2+1} \), with the convention that if \( k_1 = 1 \) then \( f \equiv +\infty \).

Theorem 9.1 ([8], Thm. 2.15). For all \( t \geq 1 \), there exists an \( \mathbb{N} \times \mathbb{R} \) indexed line ensemble \( \mathcal{H}^t = (\mathcal{H}^t_1, \mathcal{H}^t_2, \ldots) \) such that

(1) The lowest indexed curve (top line) \( \mathcal{H}^t_1 : \mathbb{R} \to \mathbb{R} \) is equal in distribution to the scaled time \( t \) Cole-Hopf solution to the narrow wedge initial data KPZ equation.
(2) The ensemble \( \mathcal{H}^t \) has the \( H_t \)-Brownian Gibbs property.
We call any such line ensemble a (scaled) KPZ line ensemble.

Building on our work in the last section, we are now ready to prove the convergence of the entire (scaled) KPZ line ensemble to the Airy line ensemble, thereby proving Conjecture 2.17 of [3]. It follows from our main result combined with the following two recent results:

Theorem 9.2 ([4]). For \( t > 0 \) the scaled KPZ line ensemble is tight and any subsequential limit is a non-intersecting line ensemble with the Brownian Gibbs property.

Theorem 9.3 ([1]). A Brownian Gibbsian line ensemble is completely characterized by the finite-dimensional distributions of its top curve.

Corollary 9.4. The \( \mathbb{N} \times \mathbb{R} \) indexed line ensemble defined by the map \( (n, x) \mapsto 2^{1/3}(\eta_n^1(x) + x^2/2) \) converges in distribution as a line ensemble to the Airy line ensemble \( \{A_n(x) : n \in \mathbb{N}, x \in \mathbb{R} \} \) as \( t \to \infty \), in the uniform-on-compact topology.

Proof. By Thm. 9.2 and Prohorov’s theorem, it is enough to show that any subsequential limit of the scaled KPZ line ensemble, which is the the scaled time \( t \) Hopf Cole solution to the narrow wedge initial data KPZ equation, converge to those of the top line of the Airy line ensemble. Hence the finite dimensional distributions of the top line of any subsequential limit of the scaled KPZ line ensemble match with those of the top line of the Airy line ensemble. Since any subsequential limit also has the Brownian Gibbs property, it follows from Thm. 9.3 that it is the Airy line ensemble (after the parabolic shift). \( \square \)

APPENDIX A. ENERGY OF RANDOMIZED PROFILES

In this appendix we provide a detailed proof of (7.1).

Fix any \( 0 < \gamma, s, L < \infty \). Approximate \( \nu \) by the corresponding random walk-approximation in the \( S_\gamma \)-lattice and by replacing \( y_i \) by \( [y_i]_\gamma \), \( b_i \) by \( [b_i]_{\varepsilon 1/2} \), \( \gamma \) by \( [\gamma]_{\varepsilon 1/2} \), \( L \) by \( [L]_{\varepsilon 1/2} \) (where \( [x]_\gamma \) denotes the nearest point in \( \mathbb{Z} \)).

First we show that if \( \eta_\varepsilon^0 \) denotes the random walk approximation of a standard Brownian motion with drift \( \alpha \) on \( [0, d] \), then the Radon Nikodym derivative \( f_\varepsilon \) of \( \eta_\varepsilon^0 \) with respect to \( \eta_\varepsilon^0 \) satisfies \( \limsup \int f_\varepsilon^2 d\eta_\varepsilon^0 < \infty \). (By an abuse of notation, \( \eta_\varepsilon^0 \) denotes both a measure and its realization as a random walk.) Recall that \( \eta_\varepsilon^0 \) is constructed as follows. Let \( B \) be a standard Brownian motion. For any \( x \in \{0, 1, \ldots, \lfloor d \varepsilon^{-1} \rfloor \} \), if \( B((x + 1)\varepsilon) - B(x\varepsilon) + a \varepsilon > 0 \) then \( \eta_\varepsilon^0((x + 1)\varepsilon) - \eta_\varepsilon^0(x\varepsilon) = \varepsilon^{1/2} \), else \( \eta_\varepsilon^0((x + 1)\varepsilon) - \eta_\varepsilon^0(x\varepsilon) = -\varepsilon^{1/2} \). Then for any sequence \( \xi \) of length \( n = \lfloor d \varepsilon^{-1} \rfloor \) with entries in \( \pm \varepsilon^{1/2} \), \( f_\varepsilon(\xi) = 2^n p_{\varepsilon}^{S_\varepsilon} (1 - p_n)^{n - S_\varepsilon} \), where \( p_n = P(B(\varepsilon) + a \varepsilon > 0) \) and \( S_\varepsilon \) denotes the number of \( \varepsilon^{1/2} \)'s in \( \xi \). Thus (here \( E \) denotes expectation with respect to \( \eta_\varepsilon^0 \)), \( E f_\varepsilon^2 = 2^{2n} (1 - p_n)^{2n} E \left( \frac{p_n}{1 - p_n} \right)^{2S_\varepsilon} \leq (1 + (2p_n - 1)^2)^n \leq e^{n(2p_n - 1)^2} \). Since \( \sqrt{n}(p_n - 1/2) = \sqrt{n}(P(B(1) > -a \sqrt{\varepsilon}) - 1/2) \to a \sqrt{d} \phi(0) \), where \( \phi \) is the density of a standard Gaussian variable,

\[
\limsup \int f_\varepsilon^2 d\eta_\varepsilon^0 \leq e^{4a^2 d \phi^2(0)} . \tag{A.1}
\]

Now we prove (7.1). First, for simplicity, assume that \( k = 1 \). Due to translation invariance of \( \nu_\varepsilon \), we can assume, without loss of generality, that \( y_1 = 0 \). Then for any curve \( \rho \) in \( S_\varepsilon \), by conditioning on \( \rho(0) \in \{[b_1]_{\varepsilon 1/2}, \ldots, [b_1]_{\varepsilon 1/2} + [\gamma]_{\varepsilon 1/2} \} \) and using (A.1) on \( [-s^{-1}(\rho(0) + L), s^{-1}(\rho(0) + L)] \),

\[
\limsup \nu_\varepsilon = \limsup \int f_{0, \varepsilon}(\rho)^2 d\nu_\varepsilon \leq 2\gamma^{-1} \exp \left\{ 4s(b_1 + L + \gamma)\phi^2(0) \right\} . \tag{A.2}
\]

Next we prove the proposition for \( k = 2 \); it is easy to see that the general case is conceptually similar to this with heavier notations. Again we can assume \( y_1 = 0 \). Let \( \mu_\varepsilon, \eta_\varepsilon^0 \) denote the restrictions of \( f_{0, \varepsilon} \nu_\varepsilon \) and \( \nu_\varepsilon \) on \([0, y_2]\) with \( \rho(0) \equiv 0 \) (that is, \( \eta_\varepsilon^0 \) is the random walk measure starting from 0). Because of

\^5Here we are taking \( t \to \infty \). To harmonize with the previous section take \( t = \delta^{-3} \).
and conditioning on $\rho(0)$, it is enough to prove that the Radon-Nikodym derivative $g_\varepsilon$ of $\mu_\varepsilon$ with respect to $\eta_\varepsilon^\omega$ satisfies $\limsup_\varepsilon \int g_\varepsilon^2 d\eta_\varepsilon^0 < \infty$. Let $\psi(x) = \hat{\phi}_{s,L}(x)$ with $k = 2$ and $y_1 = 0, b_1 = 0$. Let $\tilde{\eta}_\varepsilon^\omega$ denote the random walk approximation of $B(x) + \psi(x)$ on $[0, y_2]$, where $B$ is a standard Brownian motion. Then for any curve $\rho$ in $S_\varepsilon$ with $\rho(0) = 0$ and $\rho(y_2) \in \{[b_2]_{\varepsilon^{1/2}}, \ldots, [b_2]_{\varepsilon^{1/2}} + [\gamma]_{\varepsilon^{1/2}}\}$, 

$$
g_\varepsilon(\rho) = \frac{\mu_\varepsilon(\rho)}{\eta_\varepsilon^0(\rho)} \leq \frac{\tilde{\eta}_\varepsilon^\omega(\rho) \mathbb{P}(U = \rho(y_2) - [b_2]_{\varepsilon^{1/2}})}{\mathbb{P}(\tilde{\eta}_\varepsilon^0(\rho_2) = \rho(y_2) - [b_2]_{\varepsilon^{1/2}}) \eta_\varepsilon^0(\rho)} \leq C_{\gamma,y_2} \eta_\varepsilon^0(\rho),$$

(A.3) where $U$ denotes a discrete uniform random variable on $\{0, 1, 2, \ldots, [\gamma]_{\varepsilon^{1/2}}\}$ and $C_{\gamma,y_2}$ is some constant depending only on $y_2, \gamma$. Following the same argument as in (A.1), the Radon-Nikodym derivative of $\eta_\varepsilon^\omega$ with respect to $\eta_\varepsilon^0$ is bounded in $L^2(\eta_\varepsilon^0)$. Hence, from (A.3), it follows that $\limsup_\varepsilon \int g_\varepsilon^2 d\eta_\varepsilon^0 < \infty$.

**APPENDIX B. COUPLING**

Consider ASEP where some of the particles are first class (1) and some are second class (2). We will write $1_x$ for the indicator function that there is a 1 at site $x$, and $2_x$ for the indicator function that there is a 2 there. There is at most one particle at any site. All particles try to jump at rate $p$ to the right and $q < p$ to the left. The exclusion rule is that 1’s are blocked by 1’s, and 2’s are blocked by 2’s and 1’s, but 1’s perform their jump even if a 2 is at the target, with the 2 exchanging position with the 1 in order to make this possible; this being the sense in which they are “second class”. The magic property of ASEP is that both the first class particle processes and the first + second class particle processes are running as ASEP.

The following lemma makes precise the intuitive fact that the presence of first class particles shifts the locations of second class particles to the left (since $p > q$).

Start two versions of the system above, $(1, 2)$ and $(\hat{1}, \hat{2})$ where there are more $\hat{1}$'s than 1's in the sense that $\hat{1}_x \geq 1_x$ for all $x$. On the other hand, the 2's are in initially in pairs at the same sites $2^i(0) = \hat{2}^i(0)$. In other words, initially the two sets of second class particles are the same. We write $2^i(t), \hat{2}^i(t)$ for their positions at time $t$.

**Lemma B.1.** In the above setup there exists a coupling of the two systems $(\hat{1}, \hat{2})$ and $(1, 2)$ so that $\hat{2}^i(t) \leq 2^i(t)$ for all $i$ and $t \geq 0$.

**Proof.** Letting all the particles jump together as much as possible we see that the condition that there are more $\hat{1}$'s than 1's is preserved in time. Then it suffices to show that if at some time $t$ we have $\hat{2}^i(t) \leq 2^i(t)$ for all $i$, and for some $i_0$ we have $\hat{2}^{i_0}(t) = 1^{i_0}(t) = x$, then the rate of jumping to $x + 1$ is greater for $\hat{2}^{i_0}$ than for $2^{i_0}$ and the rate of jumping to $x - 1$ is greater for $\hat{2}^{i_0}$ than for $2^{i_0}$.

In general at $x + 1$ we have the following nine possibilities:

$$
\begin{align*}
0 & 1 \hat{1} 2 \hat{1} 2 0 \hat{2} \hat{0} 1 \\
0 & 1 \hat{2} 0 0 \hat{1} 1 \hat{2} 2
\end{align*}
$$

Clearly in the first three cases, the rate of jumping to $x + 1$ is the same for $\hat{2}^{i_0}$ and $2^{i_0}$. In the fourth case, $\hat{1}^0_0$ the rate of jumping of $\hat{2}^{i_0}$ is $q$ and that of $2^{i_0}$ is $p$. Since $p > q$ we are good. In the fifth case, $\hat{2}^0_0$ the corresponding rates are $0 \leq p$. The sixth and seventh cases $\hat{0}^1_1$ and $\hat{2}^1_1$ are disallowed by the condition $1_x \geq 1_x$. The eighth and ninth cases $\hat{0}^1_2$ and $\hat{2}^1_2$ are disallowed by the condition $\hat{2}^{i_0+1}(t) \leq 2^{i_0+1}(t)$.

At $x - 1$ we have in general the same nine cases, with the sixth and seventh disallowed. As before, in the first three cases the particles just jump, or not, together. In the fourth case, $\hat{1}^0_0$ the jump rates are $p > q$, as desired. But now it is the fifth case, $\hat{2}^0_0$ that is disallowed by the condition $\hat{2}^{i_0-1}(t) \leq 2^{i_0-1}(t)$.

The eighth case $\hat{0}^1_2$ gives rates $q \geq 0$, and the ninth case $\hat{2}^1_2$ gives rates $p > 0$, as desired. □
APPENDIX C. SKEW-TIME REVERSAL INVARIANCE

Skew time reversal invariance is the statement that the probability of the event
\[ \sup_{x \in \mathbb{R}} h(t, x; f) - g(x) \leq 0 \]  \hspace{1cm} (C.1)
is symmetric under \((f, g) \mapsto (-g, -f)\), where \(h(t, x; f)\) is the height function at \((t, x)\) given that initially \(h(0, \cdot) = f(\cdot)\). It holds for TASEP and the KPZ fixed point due to their variational description. To see that it is false for ASEP consider
\[ f \]such that \(f(\cdot)\) is a positive temperature version of skew-time reversibility: From the fact that exponentiating conservation of the model, the last statement is the probability that \(h\) from the right to the left so that \(P_\delta(h(t) \leq g) = e^{-\int \delta} = \) the probability that the first particle has not jumped left up to time \(t\). But \(P_{-g}(h(t) \leq -f) \) is the probability that starting with \(-|x|\) we are less than or equal to \(-|x|\) at time \(t\), which is 1.

For KPZ\(_\delta\),
\[ \partial_t h = \frac{1}{4} (\partial_x h)^2 + \delta \partial_x^2 h + (2\delta)^{1/2} \xi, \]  \hspace{1cm} (C.2)
what is true is a positive temperature version of skew-time reversibility: From the fact that exponentiating \(h\) gives the multiplicative stochastic heat equation which is linear in the initial data, and whose fundamental solution is statistically invariant under start and end point, the probability of the event
\[ \sup_{\delta} h(t, x; f) - g(x) \leq 0 \]  \hspace{1cm} (C.3)
is symmetric under \((f, g) \mapsto (-g, -f)\). Choosing \(f\) to be a single narrow wedge at 0 and \(-g\) to be two narrow wedges at, say, ±\(a\), one can easily check that the positive temperature version of skew-time reversibility is inconsistent with the zero temperature version.

APPENDIX D. PROOF OF LEM. 5.1

Lemma D.1. Let \(X(h_t, \hat{h})\) denote the number of points of maxima of \(h_t(\cdot; -g) + \hat{h}(\cdot)\). Then
\[ \sum_{x \in \mathbb{Z}} \sum_{h(0) \in \mathbb{Z}^{1/2}} |\nabla_x^{\text{sym}} p_t \text{TASEP}, \varepsilon(h, \leq g)| \leq \sum_{k=1}^{\infty} k^{\text{TASEP}}(X(h_t, \hat{h}) = k) = E^{\text{TASEP}}(X(h_t, \hat{h})). \]  \hspace{1cm} (D.1)

Proof. For simplicity of notation, we denote \(h_x,1, \text{sym} \) by \(h_x\). First of all if \(h_x = h\), then clearly \(\nabla_x^{\text{sym}} p_t \text{TASEP}, \varepsilon = 0\). On the other hand, if \(h\) has a local maximum at \(x\), then, by skew-time reversibility of TASEP height function (see Appdx C), if \(h_t(y, -g) = h_x(y, -g)\) denotes the TASEP height function at scale \(\varepsilon\) at time \(t\) and position \(y\) started from \(-g\), \(\nabla_x^{\text{sym}} p_t \text{TASEP}, \varepsilon(h, \leq g)|\) is given by
\[ |\nabla_x^{\text{sym}} p_t \text{TASEP}, \varepsilon(h, \leq g)| = |\text{TASEP}, \varepsilon(-g, \leq -h, x) - \text{TASEP}, \varepsilon(-g, \leq -h, y)|, \]
which is the probability that \(h_t(y, -g) \leq -h^2(y)\) for all \(y \in \mathbb{Z}\), but for some \(y\), \(h_t(y, -g) > -h(y)\). Of course, necessarily \(y = x\), and one can see that this is the probability that \(h_t(y, -g) + h(y) \leq 0\) for all \(y\) and \(h_t(x, -g) + h(x) \leq \varepsilon^{1/2} \) or \(2\varepsilon^{1/2}\). Thus
\[ |\nabla_x^{\text{sym}} p_t \text{TASEP}, \varepsilon(h, \leq g)| \leq \text{TASEP}(\arg \max \{h_t(y, -g) + h(y)\} = x, h_t(x, -g) + h(x) \in \{\varepsilon^{1/2}, 2\varepsilon^{1/2}\}) \].

Here \(\text{TASEP}\) denotes the probability with respect to the TASEP dynamics for a fixed \(h\). The right-hand-side above is equal to \(\text{TASEP}(\arg \max \{h_t(y, -g) + \hat{h}(y)\} = x, h_t(x, -g) + \hat{h}(x) \in \{\varepsilon^{1/2} - h(0), 2\varepsilon^{1/2} - h(0)\})\), so this probability summed over all \(h(0) \in \varepsilon^{1/2} \mathbb{Z}\) for each fixed \(h\) equals \(\text{TASEP}(\arg \max \{h_t(y, -g) + \hat{h}(y)\} = x)\). Note that only one of the two possibilities \(h_t(x, -g) + h(x) = \varepsilon^{1/2}\), \(h_t(x, -g) + h(x) = 2\varepsilon^{1/2}\) happens; it is because there is a parity conservation in the model.

On the other hand, if \(h\) has a local minimum at \(x\), \(|\nabla_x^{\text{sym}} p_t \text{TASEP}, \varepsilon(h, \leq g)|\) is given by the probability that \(h_t(y, -g) \leq -h(y)\) for all \(y \in \mathbb{Z}\), but for some \(y\), \(h_t(y, -g) > -h^2(y)\). Clearly again \(y = x\) and \(h_t(x, -g) = -h^2(x) + \varepsilon^{1/2} \) or \(2\varepsilon^{1/2}\). Since \(-h(x) = -h^2(x) + 2\varepsilon^{1/2}\), again by the parity conservation of the model, the last statement is the probability that \(h_t(y, -g) + h(y) \leq 0\) for all \(y\) and \(h_t(x, -g) + h(x) = 0\); or \(h_t(y, -g) + h(y) \leq -\varepsilon^{1/2}\) for all \(y\) and \(h_t(x, -g) + h(x) = -\varepsilon^{1/2}\). That
is, \( \nabla_x p_t \) is given by \( \mathbb{P}^{\text{TASEP}}(\arg\max \{ h_t(y; -g) + h(y) \} = x, h_t(x; -g) + h(x) \in \{-\varepsilon^{1/2}, 0\}) \). As in the last paragraph, summed over all \( h(0) \in \varepsilon^{1/2}\mathbb{Z} \) for each fixed \( \hat{h} \), we get \( \mathbb{P}^{\text{TASEP}}(\arg\max \{ h_t(y; -g) + h(y) \} = x) \).

Considering the possibility of multiple points of maxima, we have (D.1) when we sum over all \( x \in \varepsilon\mathbb{Z} \).

Thus the square of the left hand side of (D.1) is bounded by \( \mathbb{E}^{\text{TASEP}}[(X(h_t, \hat{h}))^2] \). Till now \( \hat{h} \) was fixed; now we take an expectation over \( \mu_{\varepsilon} \), so that the left-hand side of (5.2) is bounded by \( \mathbb{E}[(X(h_t, \hat{h}))^2] = \sum_{k=0}^{\infty} k^2 \mathbb{P}(X(h_t, \hat{h}) = k) \), where \( \mathbb{P} = \mathbb{P}^{\text{TASEP}} \times \mu_{\varepsilon} \) denotes the product measure of the TASEP dynamics and the random walk measure \( \mu_{\varepsilon} \) for \( \hat{h} \), and \( \mathbb{E} \) denotes the corresponding expectation. Conditioning on \( h_t(\cdot; -g) \), to get (5.2) it is enough to show

**Lemma D.2.** For all \( \varepsilon > 0 \) and \( k \geq 1 \) and any fixed function \( f(\cdot) \in \mathbb{S}_\varepsilon \) such that \( f(x) + \hat{h}(x) \to -\infty \) as \( |x| \to \infty \), there are absolute constants \( C, c > 0 \) (not depending on \( \varepsilon, f, k \)) such that

\[
\mu_{\varepsilon}(X(f, \hat{h}) = k) \leq Ce^{-ck^{1/4}}. \tag{D.2}
\]

Note that as \( -g(x) \leq -C(1 + |x|) \), one has that for any fixed \( \varepsilon > 0 \) almost surely the TASEP height function at time \( t, h_t(x; -g) + \hat{h}(x) \to -\infty \) as \( |x| \to \infty \). In fact, one can get the same linear bounds with a slightly worse constant and a height shift. To see this, note that we can bound the initial data above by both \( -C(1 + x) \) or \( -C(1 - x) \). By adjusting the drift, we can put a random walk height function above each of these. Because the TASEP height function preserves order, at a later time the height function will be dominated by the evolution of either one of these random walk initial data. Besides the height shift, these two are equilibria, and so we conclude that our height function is again bounded by the two random walks with drift at time \( t \). These, in turn can be bounded above by \( C'(1 - x) \) and \( C'(1 + x) \).

The proof of Lem. [D.2] is adapted from [16] who consider \( f \equiv 0 \) on a finite interval. Intuitively, \( f \equiv 0 \) is the worst case, and it is plausible one could avoid what follows by making this intuition rigorous.

**Proof of Lem. [D.2]** Fix \( k > 1 \) and any function \( f \in \mathbb{S}_\varepsilon \) such that \( f(x) + \hat{h}(x) \to -\infty \) as \( |x| \to \infty \). Because of the decay condition on \( f + \hat{h} \), all maxima are attained in a compact interval almost surely, that is, there exists \( L_k > 0 \) such that all points of maxima for \( f + \hat{h} \) are in \([-L_k, L_k]\) with probability at least \( 1 - e^{-k} \). Let us call this event \( A_k \). Henceforth, we restrict ourselves to points only in the compact set \([-L_k, L_k]\). First observe that \( h + f \) can attain its maximum at consecutive points. Let \( \mathcal{E}_k \) denote the event that there is a string of \( \sqrt{k} \) consecutive points at which \( h + f \) attains the maximum. Then we show that \( \mu_{\varepsilon}(\mathcal{E}_k) \leq Ce^{-c\sqrt{k}} \) for some absolute constants \( C, c > 0 \).

To this end, on the event \( \mathcal{E}_k \), let \( \{ x^*, x^* + \varepsilon, \ldots, x^* + be \} \) be the last string of at least \( \sqrt{k} \) points at which \( f + \hat{h} \) attains the maximum. We consider the last \( \sqrt{k}/3 \) points of this string, that is the points \( \{ x^* + \varepsilon(b - \sqrt{k}/3 + 1), \ldots, x^* + \varepsilon \} \) (to avoid cumbersome notations, we assume \( \sqrt{k}/3 \) is an integer). With \( u \) as defined in (2.2), and \( m \) the maximum value of \( f + \hat{h} \), we have that \( m := h(x) + f(x) = \hat{h}(x + \varepsilon) + f(x + \varepsilon) \), that is \( u(x) = f(x) - f(x + \varepsilon) \) for all \( x \in \mathbb{I} := \{ x^* + \varepsilon(b - \sqrt{k}/3), \ldots, x^* + \varepsilon(b - 2) \} \). That is, given the function \( f \), the values of \( \mathbb{I} \) for each \( x \in \mathbb{I} \) are fixed. Now for each such configuration of \( u \) in \( \mathcal{E}_k \), we construct \( 2\sqrt{k}/3 - 2 \) new and distinct configurations by assigning different values of \( u(x) \) for all \( x \in \mathbb{I} \) and setting \( u(x^*) + \varepsilon(b - \sqrt{k}/3), u(x^* + \varepsilon(b - 1)) \) such that for each of these new configurations the values of \( \hat{h} + f \) at \( x^* + \varepsilon(b - \sqrt{k}/3), x^* + \varepsilon \) are different from those of the original configuration. If we define this map as \( \phi \), then for each configuration \( u \in \mathcal{E}_k \), \( \phi(u) \) is a set of \( 2\sqrt{k}/3 - 2 \) elements and the sets are disjoint for different configurations \( u \). To see the disjointness, observe that it is easy to recognize the map \( \phi^{-1} \). For any \( u_{**} \in \phi(u) \) and the walk \( h_{**} \) defined from the sequence \( u_{**} \) as \( h_{**}(x + \varepsilon) = h_{**}(x) + u_{**}(x) \), either \( h_{**} + f \) or \( x = x^* + \varepsilon \) is in which case the number of points of maxima of \( h_{**} + f \) is at most \( \sqrt{k}/3 \), and we can recover \( u \) by identifying the last string of length at least \( 2\sqrt{k}/3 \) of consecutive points before a point of
maximum where \( \hat{h}_s + f \) is flat. On the other hand, if \((\hat{h}_s + f)(x) \leq (\hat{h} + f)(x)\) for all \( x - \varepsilon \in \mathcal{I} \) and \((\hat{h}_s + f)(x^* + b\varepsilon) < (\hat{h} + f)(x^* + b\varepsilon)\), the number of points of maxima of \( \hat{h}_s + f \) is at least \( 2\sqrt{k}/3 \) and we can recover \( u \) by identifying the last string of points of maxima of length at least \( 2\sqrt{k}/3 \). This gives that

\[
\mu_\varepsilon(\mathcal{E}_k \cap \mathcal{A}_k) \leq 2^{-\sqrt{k}/3 + 2}.
\]

Now on the event \( \mathcal{E}^k_1 \cap \{X(f, \hat{h}) = k\} \), there are at least \( \sqrt{k} \) disjoint blocks of points of maxima for \( \hat{h} + f \), each of size less than \( \sqrt{k} \). That is, let \( \{x_1, x_1 + \varepsilon, \ldots, x_i + (b_i - 1)\varepsilon\} \) for \( i = 1, 2, \ldots, \ell \) be the points of maxima for \( \hat{h} + f \), where \( b_1 + b_2 + \ldots + b_{\ell} = k \) and \( x_i + b_i \varepsilon < x_{i+1}, b_i < \sqrt{k} \) for all \( i \) and hence \( \ell \geq \sqrt{k} \). First assume that, in addition to the above, there are at least \( k^{1/3} \) blocks such that in each block, there is an \( x \) such that \( u(x - \varepsilon) = -\varepsilon^{1/2} \) (clearly then \( f(x - \varepsilon) = \varepsilon^{1/2} \)). Let us call this event \( \mathcal{E}^k_1 \) and enumerate these blocks as before, with \( \ell \geq k^{1/3} \) this time. Now for any \( i = 1, 2, \ldots, k^{1/4} \), and any configuration \( u \in \mathcal{E}^k_L \), we define a new configuration \( \Phi_i(u) = u_* \) as follows. Let \( y_i \) be the first \( x \) such that \( x + \varepsilon \) is in the \( (\ell - i) \)-th block and \( u(y_i) = -\varepsilon^{1/2} \). Define \( u_{\varepsilon}(y_i) = \varepsilon^{1/2} \) and \( u_{\varepsilon}(x) = u(x) \) for all other \( x \); and define the walk \( \hat{h}_s \) from the sequence \( u_* \) as \( \hat{h}_s(x + \varepsilon) = \hat{h}_s(x) + u_*(x) \). Then \((\hat{h}_s + f)(x) = m + 2\varepsilon^{1/2} \) for \( x \in \{x_j, x_j + \varepsilon, \ldots, x_j + (b_j - 1)\varepsilon\} \) for all \( \ell - i + 1 \leq j \leq \ell \) and for all points in the \( (\ell - i) \)-th block after \( y_i \). Let \( \Phi_{j_1, j_2, \ldots, j_m} := \Phi_{j_1} \circ \cdots \circ \Phi_{j_m} \) and denote the set

\[
\Phi(u) = \{\Phi_S(u) : S \subseteq \{1, 2, \ldots, k^{1/4}\}\}
\]

for each \( u \in \mathcal{E}^k_1 \). Thus \( \Phi(u) \) has \( 2^{k/4} \) elements and for different configurations \( u \) the sets \( \Phi(u) \) are disjoint. Observe that the disjointness follows because we can identify \( u \) from \( u_* \) by first finding the maximum value \( m_* \) attained for the first time at \( x_* \), say, and then the highest value less than \( m_* \) that is attained in at least \( k^{1/3}/2 \) blocks of points less than \( x_* \). Hence \( \mu_\varepsilon(\mathcal{E}^k_1) \leq 2^{-k^{1/4}} \). Similarly if there are at least \( k^{1/3} \) blocks such that in each block, there is an \( x \) such that \( u(x) = \varepsilon^{1/2} \) (clearly then \( f(x) = -\varepsilon^{1/2} \) then), and we call this event \( \mathcal{E}^k_2 \), then reversing the sequence \( u \) and switching the \( \varepsilon^{1/2} \) to \( -\varepsilon^{1/2} \) and vice versa, we see as before that \( \mu_\varepsilon(\mathcal{E}^k_2) \leq 2^{-k^{1/4}} \). On the event that there does not exist such \( k^{1/3} \) blocks with \( x \) such that \( u(x) = \varepsilon^{1/2} \) or \( k^{1/3} \) blocks with \( x \) such that \( u(x) = -\varepsilon^{1/2} \), it means that there are at most \( 2k^{1/3} \) blocks of size at least \( 2 \), that is, there are at least \( k/2 \) blocks of points of maxima that are all of size \( 1 \) (since \( k - 2k^{1/3} \geq k/2 \) for \( k \) large). We call this event \( \mathcal{E}^k_3 \). Note that on this event, there exists a sequence of at least \( \sqrt{k} \) consecutive blocks of maxima, all of size \( 1 \). If for a sequence \( u \) of finite length, there are \( \sqrt{k} \) points of maxima and each maximum is attained at a block of size \( 1 \), then by defining the map \( \psi_1 \) as \( \psi_1(u) = u_* \) where \( u_*(x_i) = \varepsilon^{1/2} \) and \( u_*(x) = u(x) \) for all other \( x \), where \( x_i \) is the \( i \)-th point of maximum counted from the right and \( \psi(u) = \{\psi_S(u) : S \subseteq \{1, 2, \ldots, k^{1/3}\}\} \), we see that \( \psi(u) \) has \( 2^{k/3} \) elements and for different configurations \( u \) the sets \( \psi(u) \) are disjoint. This implies that \( \mu_\varepsilon(\mathcal{E}^k_3) \leq 2^{-k^{1/3}} \). Thus

\[
\mu_\varepsilon\left(\{X(f, \hat{h}) = k\} \cap \mathcal{E}^k \cap \mathcal{A}_k\right) \leq 2^{-k^{1/4} + 2}.
\]

Putting all this together we have

\[
\mu_\varepsilon(X(f, \hat{h}) = k) \leq \mu_\varepsilon\left(\{X(f, \hat{h}) = k\} \cap \mathcal{E}^k \cap \mathcal{A}_k\right) + \mu_\varepsilon(\mathcal{E}^k \cap \mathcal{A}_k) + \mu_\varepsilon(\mathcal{A}_k) \leq Ce^{-ck^{1/4}}
\]

for some absolute constants \( C, c > 0 \).

**Appendix E. Tightness and convergence in the uniform-on-compact topology**

**Proposition E.1.** Let \( t > 0 \) be fixed and for each \( \varepsilon > 0 \), let \( h_\varepsilon \) be a random element of \( \mathbb{S}_c \). Assume that for all \( g_1, g_2 \) (uniformly) continuous with \( g_1(x) = C_1 + C_2|x - c_3| \) and \( g_2(x) = C_1 - C_2|x - c_3| \) for all sufficiently large \( |x| \), for some \( C_1, C_2 > 0 \) and \( c_3 \in \mathbb{R} \), we have

\[
\lim_{\varepsilon \to 0} P(h_\varepsilon(x) \leq g_1(x), x \in \mathbb{R}) = p_{t}^{\text{FP}}(h, \text{hyp}(g_1)), \quad \lim_{\varepsilon \to 0} P(h_\varepsilon(x) > g_2(x), x \in \mathbb{R}) = p_{t}^{\text{FP}}(h, \text{epi}(g_2)),
\]

for some function \( h \) which is bounded on any compact interval and \( |h(x)| \leq C(1 + |x|) \) for some \( C > 0 \). Here \( P \) is the probability with respect to the randomness in \( h_\varepsilon \). Then (the distributions of)
\{h_\varepsilon\}_{\varepsilon>0} are tight in the uniform-on-compact topology. In particular, we have the following modulus of continuity uniform in \varepsilon. For any \(b > 0\) small enough,

$$
\limsup_{\varepsilon \to 0} \mathbb{P}(\omega_\varepsilon(b) > mb^{1/2} \log^{7/6}(1 + b^{-1})) \leq ce^{-d\lambda^3/2}, \tag{E.2}
$$

where, for \(b \in (0, 1)\), \(\omega_\varepsilon(b) := \sup_{|x-y| \leq b, x, y \in [-L, L]} |h_\varepsilon(x) - h_\varepsilon(y)|\) for any fixed compact set \([-L, L]\).

**Remark E.2.** Since \(\text{hyp}(g)\) sets for uniformly continuous \(g\) growing to \(\infty\) at some rate form a separating class for the KPZ fixed point started from any \(h\) with \(h(x) \leq C(1 + |x|)\), by Prop. 6.2, Rem. 6.4 and Prop. E.1 we get the uniform-on-compact part of Thm. 2.2.1.

Note that no attempt is being made to optimise the exponent of the logarithms in Prop. E.1 and the 7/6 just comes from the proof.

Let \(s\) denote the following function on \(\mathbb{R}\)

$$
s(x) = \min\{|x|^{1/2} \log^{1/2}(1 + |x|^{-1}), 1\} + |x|. \tag{E.3}
$$

We make use of the following modulus of continuity estimate for the KPZ fixed point in the proof of Proposition E.1.

**Lemma E.3.** Fix \(t > 0\) and \(L < \infty\). For each \(k \in \mathbb{N}\), let \(h^{FP}_t\) denote the KPZ fixed point at time \(t\) started from the initial condition \(h\) with \(|h(x)| \leq C(1 + |x|)\) and

$$
\mathcal{E}_k := \{|h^{FP}_t(x) - h^{FP}_t(y)| \leq ks(x-y), |h^{FP}_t(x)| \leq k, \text{ for all } x \in [-L, L], y \in \mathbb{R}\}. \tag{E.4}
$$

Then there are \(c < \infty\) and \(d > 0\) (depending on \(L, t\) but not \(k\)) such that

$$
P(\mathcal{E}_k) \geq 1 - ce^{-dk^3/2}. \tag{E.5}
$$

Here we use \(P\) to denote probabilities with respect to the KPZ fixed point, while we reserve \(\mathbb{P}\) for the approximating processes.

**Proof.** [10] Cor. 10.7. shows that for the Airy sheet there exists \(d > 0\) and \(c < \infty\) so that

$$
P \left( \sup_{x,y \in \mathbb{R}} \frac{|\mathcal{A}(x,y)|}{\log^2(|x| + |y| + 2)} > m \right) \leq ce^{-d\lambda^3/2}. \tag{E.6}
$$

Therefore, from the variational formula (2.14) and \(h(y) \leq C(1 + |y|)\), there exists perhaps new \(c < \infty\) and \(d > 0\) so that

$$
P \left( \sup_{x \in [-L-1, L+1]} |h^{FP}_t(x)| > m \right) \leq ce^{-d\lambda^3/2}. \tag{E.7}
$$

Next we show that for any such \(x \in [-L - 1, L + 1]\) and any initial condition \(h(y) \leq C(1 + |y|)\), with high probability, the supremum in the variation formula (2.14) for \(h^{FP}_t(x)\) is attained in a compact interval of \(y\). To this end, notice that from \(h(y) \leq C(1 + |y|)\) and (E.6) and the negative parabola in the variational problem, for fixed \(x \in [-L-1, L+1]\),

$$
\sup_{|y| \geq \lambda} \left\{ t^{1/3} \mathcal{A}(t^{-2/3}y, t^{-2/3}x) - \frac{1}{7} (y-x)^2 + h(y) \right\} \geq h^{FP}_t(x)
$$

only if \(h^{FP}_t(x) \leq -\alpha \lambda^2\) for some \(x \in [-L-1, L+1]\), or, for some \(y\) and \(\alpha > 0\), \(t^{1/3} \mathcal{A}(t^{-2/3}y, t^{-2/3}x) \geq \alpha \lambda^2\). From (E.6) and (E.7) either happens with probability less than some \(ce^{-d\lambda^3}\). Thus, if \(S_\lambda\) denotes the event that the absolute value of the argmax in \(y\) of \(t^{1/3} \mathcal{A}(t^{-2/3}y, t^{-2/3}x) - \frac{1}{7} (y-x)^2 + h(y)\) is at most \(\lambda\) for all \(x \in [-L - 1, L + 1]\), then

$$
P(S^c_\lambda) \leq ce^{-d\lambda^3}.$$
Therefore for any \( x \in [-L, L] \) and \( y \in [-L - 1, L + 1] \), on the event \( S_\lambda \), with \( A_t(x, y) := t^{1/3} \mathcal{A}(t^{-2/3}y, t^{-2/3}x) - \frac{1}{2}(y - x)^2 \), we have

\[
|h_{t}^{FP}(x) - h_{t}^{FP}(y)| \leq \sup_{z \in [-\lambda, \lambda]} |A_t(z, x) - A_t(z, y)|
\]

\[
\leq \sup_{z \in [-\lambda, \lambda]} |A_t(z, x) - A_t(z, y)|
\]

\[
\leq \sup_{z \in [-\lambda, \lambda]} |t^{1/3} \mathcal{A}(t^{-2/3}z, t^{-2/3}x) - t^{1/3} \mathcal{A}(t^{-2/3}z, t^{-2/3}y)| + \frac{2(L + \lambda + 1)}{t}|x - y|
\]

\[
\leq Ks(x - y)
\]

with \( P(K > m) \leq ce^{-dm^{3/2}} \), where the last line follows from (2.15) and \( c, d \) are constants depending on \( L, t \).

On the other hand, from the negative parabola in the variational formula (2.14) and (E.6), it follows that for \( h(y) \leq C(1 + |y|) \) for all \( y \),

\[
h_t^{FP}(x) \leq K_1(1 + |x|)
\]

for some random constant \( K_1 \) satisfying \( P(K_1 > m) \leq ce^{-dm^{3/2}} \). Similarly, for \( h(y) \geq -C(1 + |y|) \) for all \( y \), it follows from the variational formula (2.14) and (E.6) that

\[
h_t^{FP}(x) \geq t^{1/3} \mathcal{A}(t^{-2/3}x, t^{-2/3}x) + h(x) \geq -K_1(1 + |x|)
\]

for all \( x \in \mathbb{R} \).

Thus for all \( x \in [-L, L] \) and \( |y| > L + 1 \), using (E.7)

\[
|h_t^{FP}(x) - h_t^{FP}(y)| \leq |h_t^{FP}(x)| + |h_t^{FP}(y)| \leq K_1 + K_1(1 + |y|)
\]

\[
\leq K_1|x - y| + K_1(2 + L) \leq Ks(x - y),
\]

\[
P(K > m) \leq ce^{-dm^{3/2}}
\]

for constants \( c, d \) that depend on \( L \). Together, we have the lemma. \( \Box \)

Now, we are ready to prove Prop. E.1

**Proof of Prop. E.1** Without loss of generality we can assume that the compact interval is \([-L, L] \); so it is enough to show that \( \{ h_{e \mid [-L, L]} \} \) is tight in the uniform topology.

Fix \( k > 0 \) large and \( a, b > 0 \) small (to avoid cumbersome notations assume \( k, a^{-1}, b^{-1} \in \mathbb{N} \)). Consider the rectangle \([-L, L] \times [-k, k] \) and split it into rectangles of vertical and horizontal dimensions \( a \) and \( b \). That is, let \( x_i = -L + ib \) for \( i \in I := \{0, 1, 2, \ldots, 2Lb^{-1}\} \) and \( f_j = -k + ja \) for \( j \in J := \{0, 1, 2, 2ka^{-1}\} \). By our assumption in (E.1), for any \( i, j \),

\[
\mathbb{P}(h_{\varepsilon}(x) > f_{j+1} + ks(x - x_i) \text{ for some } x \in \mathbb{R}) \Rightarrow P(h_t^{FP}(x) > f_{j+1} + ks(x - x_i) \text{ for some } x \in \mathbb{R}),
\]

and

\[
\mathbb{P}(h_{\varepsilon}(x) \leq f_j - ks(x - x_i) \text{ for some } x \in \mathbb{R}) \Rightarrow P(h_t^{FP}(x) \leq f_j - ks(x - x_i) \text{ for some } x \in \mathbb{R})
\]

as \( \varepsilon \to 0 \).

On the event \( E_k \) defined in (E.4), we have

\[
\{ h_t^{FP}(x) > f_{j+1} + ks(x - x_i) \text{ for some } x \in \mathbb{R} \} \subseteq \{ h_t^{FP}(x_i) > f_{j+1} \},
\]

and

\[
\{ h_t^{FP}(x) \leq f_j - ks(x - x_i) \text{ for some } x \in \mathbb{R} \} \subseteq \{ h_t^{FP}(x_i) \leq f_j \}.
\]

Now, let us define the “tube”-sets

\[
T_{i,j,k} := \{(x, y) \in \mathbb{R}^2 : f_j - ks(x - x_i) < y \leq f_{j+1} + ks(x - x_i)\}.
\]

Then, using (E.9), (E.10), (E.11), (E.12) and a union bound,

\[
\limsup_{\varepsilon \to 0} \mathbb{P}((x, h_{\varepsilon}(x)) \in T_{i,j,k}^c) \text{ for some } x \in \mathbb{R} \leq P(h_t^{FP}(x_i) \in (f_j, f_{j+1}]) + 2P(E_k^c).
\]

(E.13)
To put it another way, abusing notation to let \( h_\varepsilon \) denote the graph \( (x, h_\varepsilon(x)) \), \( x \in \mathbb{R} \),
\[
\liminf_{\varepsilon \to 0} \mathbb{P}(h_\varepsilon \in T_{i,j,k}) \geq P(h_\varepsilon^{\text{FP}}(x_i) \in (f_j, f_{j+1}) - 2P(E_\varepsilon^c). \quad (E.14)
\]
For fixed \( i \), the sets \( T_{i,j,k} \) are disjoint for all \( j = 0, 1, 2, \ldots, 2ka^{-1} \). Summing the above probabilities, we have again by \( (E.4) \),
\[
\liminf_{\varepsilon \to 0} \mathbb{P}(h_\varepsilon \in \cup_{j \in J} T_{i,j,k}) \geq P(h_\varepsilon^{\text{FP}}(x_i) \in [-k, k]) - (4ka^{-1} + 1)P(E_\varepsilon^c) \geq 1 - (4ka^{-1} + 2)P(E_\varepsilon^c). \quad (E.15)
\]
If we let \( \mathcal{G}^{\varepsilon}_{a,b,k} := \{ \forall i \in I, h_\varepsilon \in \cup_{j \in J} T_{i,j,k} \} \), this implies by another union bound
\[
\liminf_{\varepsilon \to 0} \mathbb{P}(\mathcal{G}^{\varepsilon}_{a,b,k}) \geq 1 - 10kLa^{-1}b^{-1}P(E_\varepsilon^c). \quad (E.16)
\]
For any \( x, y \in [-L, L] \) with \( |x - y| \leq b \), there exists \( i \in I \) such that \( |x - x_i|, |y - x_i| \leq b \). On the event \( \mathcal{G}^{\varepsilon}_{a,b,k} \), if \( j \in J \) is such that \( h_\varepsilon(x_i) \in (f_j, f_{j+1}] \), we have \( h_\varepsilon \in T_{i,j,k} \). Hence for \( b \) sufficiently small,
\[
|h_\varepsilon(x) - h_\varepsilon(x_i)| \leq a + ks(x - x_i) \leq a + 2kb^{1/2} \log^1/(1 + b^{-1}),
\]
and the same holds for \( |h_\varepsilon(y) - h_\varepsilon(x_i)| \). Thus \( \omega_\varepsilon(b) \leq 2a + 4kb^{1/2} \log^1/(1 + b^{-1}) \). Choosing \( a = b^{1/2} \), we have for \( b \) sufficiently small, by \( (E.16) \),
\[
\limsup_{\varepsilon \to 0} \mathbb{P}(\omega_\varepsilon(b) \geq 5kb^{1/2} \log^1/(1 + b^{-1})) \leq 10kLa^{-3/2}P(E_\varepsilon^c). \quad (E.17)
\]
Choosing \( k = C\varepsilon \log^2/3(1 + b^{-1}) \), and using Lem. \( \text{[E.3]} \), we obtain \( \text{[E.2]} \).

From the Arzela-Ascoli theorem, tightness in the uniform-on-compact topology follows from the modulus of continuity estimate \( \text{[E.2]} \), as long as we also have a bound on \( \limsup_{\varepsilon \to 0} \mathbb{P}(|h_\varepsilon(0)| > k) \). The latter is obtained by an easier version of the same argument, letting one of the \( x_i \) equal 0. \( \square \)

### APPENDIX F. STRONG SECTOR CONDITION (PROOF OF LEM. \( \text{[E.2]} \))

We recall the proof in \( \text{[E.5]} \) as recounted in \( \text{[E.3]} \), and describe the minor modifications necessary to include the height \( h(0) \). We start with a definition. An **irreducible cycle** \( C \) is a sequence of integers \( y_0, \ldots, y_k \) with \( y_0 = y_k = 0 \) and \( y_i \neq y_j \) for any other \( i \neq j \). \( \pi_C(a_j) = \frac{1}{2} \) for \( a_j = y_j - y_{j-1} \) and zero otherwise. Then it is shown in \( \text{[E.5]} \) that any \( p(\cdot) \) with mean 0 and finite support can be written as a finite sum \( p(x) = \sum_i w_i \pi_C(x) \) where \( w_i > 0 \), \( \sum_i w_i = 1 \) and \( C_i \) are irreducible cycles. From this one obtains the representation
\[
Mf = \sum_{x \in \mathbb{Z}} \sum_{C_{i+x}} M_{C_{i+x}} M_{C_{i+x}} = \frac{1}{k} \sum_{i=0}^{k-1} \eta(x + y_i)(1 - \eta(x + y_{i+1})) \nabla_{x+y_i,a_{i+1}} f \quad (F.1)
\]

We have a Markov generator \( A = M_{C_{i+x}} \) acting on functions on particle configurations \( \eta \in \{0, 1\}^\mathbb{Z} \) on \( \mathcal{G} := \{y_0 + x, \ldots, y_k - x\} \). We can furthermore restrict to the subset \( \mathcal{G}_k \) of particle configurations with a fixed number \( \ell \) of particles and our measure just becomes the uniform measure. Let \( A = \frac{1}{2}(A + A^*) \) denote the symmetrization of \( A \). It has the same range as \( A \), namely mean zero functions (since the uniform measure is uniquely invariant for both). Since the configuration space is finite, there is a finite \( B \) such that
\[
\sum_{\eta \in \mathcal{G}_k} Ag(\eta)(-A)^{-1}Ag(\eta) \leq B^2 \sum_{\eta \in \mathcal{G}_k} g(\eta)(-A)(\eta). \quad (F.2)
\]
Equivalently, by Cauchy-Schwarz inequality, for any \( \alpha > 0 \),
\[
\sum_{\eta \in \mathcal{G}_k} f(\eta)Ag(\eta) \leq \frac{\alpha}{2} \sum_{\eta_1 \in \mathcal{G}_k} f(\eta)(-A)f(\eta) + \frac{B}{2\alpha} \sum_{\eta \in \mathcal{G}_k} g(\eta)(-A)g(\eta). \quad (F.3)
\]
Now let \( f \) and \( g \) be functions of \( h(0) \) and \( \eta \in \{0, 1\}^\mathbb{Z} \). We can think of general function \( f \) and \( g \) of \( h(0) \in \mathbb{Z} \) and \( \eta \in \{0, 1\}^\mathbb{Z} \) of being first of all a function of \( \eta \) \( \mathcal{G}_k \), then of \( \ell \), then of all the other variables \( h(0) \) and \( \eta(x), x \in \mathbb{Z} \setminus \mathcal{G} \). The inequality \( \text{[F.3]} \) clearly holds, with these general functions, and the dependence on the extra variables \( \ell \) and \( \eta(x), x \in \mathbb{Z} \setminus \mathcal{G} \) there but not written. Nothing in \( \text{[F.3]} \) affects the variables \( \eta(x), x \in \mathbb{Z} \setminus \mathcal{G} \). In terms of \( h(0) \), there are two cases. If all elements of \( \mathcal{G} \) are
either in \{0, 1, \ldots\} or in \{\ldots, -2, -1\} then no move of \( A \) affects \( h(0) \). Averaging the inequality over the measure \( \nu \) gives in this case
\[
\int fM_{C+x} g d\nu \leq \frac{B\alpha}{2} \int f(-M_{C+x} f) d\nu + \frac{B}{2\alpha} \int g(-M_{C+x} g) d\nu. \tag{F.4}
\]
In the second case we have elements of \( \mathcal{C} \) in both \{0, 1, \ldots\} and \{\ldots, -2, -1\}, so some moves affect \( h(0) \). We claim the same inequality holds. For suppose \( y_i + x \leq -1 \) and \( y_{i+1} + x \geq 0 \). The corresponding term in each term of our inequality on \( \mathcal{Y}_\ell \) reads either
\[
\sum_{\eta \in \mathcal{Y}_\ell} f(h(0), \eta) \eta_{y_{i}+x}(1-\eta_{y_{i+1}+x})(g(h(0) - 2, \eta)_{y_{i}+x,y_{i+1}+x}) - g(h(0), \eta)) \tag{F.5}
\]
on the left hand side, or the same thing with two \( f \)'s or two \( g \)'s on the right hand side. We can take the sum over \( h(0) \in \mathbb{Z} \) and pass it through the finite sum over \( \mathcal{Y}_\ell \), then take the expectation with respect to the marginal distribution of \( \ell \) and the \( \eta(x), x \in \mathbb{Z} \setminus \mathcal{C} \) under \( \nu \). It is clearly the same if we had instead \( y_i + x \geq 0 \) and \( y_{i+1} + x \leq -1 \), with the \(-2\) replaced by a \(+2\). Doing this to all the terms shows that (F.4) holds in the second case as well.

Now we can sum (F.4) over \( x \in \mathbb{Z} \) and \( C_i \) with weights \( w_i \) to obtain
\[
\int fM g d\nu \leq \frac{B\alpha}{2} \int f(-M f) d\nu + \frac{B}{2\alpha} \int g(-M g) d\nu. \tag{F.6}
\]
Optimizing over \( \alpha > 0 \) gives (3.9).

Acknowledgements. JQ would like to thank Bálint Virág for enlightening discussions. Both authors were supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Gideon Amir, Ivan Corwin, and Jeremy Quastel, Probability distribution of the free energy of the continuum directed random polymer in 1 + 1 dimensions, Comm. Pure Appl. Math. 64 (2011), no. 4, 466–537. MR 2796514

[2] L. Bertini, C. Landim, and S. Olla, Derivation of Cahn-Hilliard equations from Ginzburg-Landau models, J. Statist. Phys. 88 (1997), no. 1-2, 365–381. MR 1468389

[3] Lorenzo Bertini and Giambattista Giacomin, Stochastic Burgers and KPZ equations from particle systems, Comm. Math. Phys. 183 (1997), no. 3, 571–607. MR 1462228

[4] Alexei Borodin, Ivan Corwin, Patrik Ferrari, and Bálint Vető, Height fluctuations for the stationary KPZ equation, Math. Phys. Anal. Geom. 18 (2015), no. 1, Art. 20, 95. MR 3366125

[5] Pasquale Calabrese and Pierre Le Doussal, Interaction quench in a Lieb-Liniger model and the KPZ equation with flat initial conditions, J. Stat. Mech. Theory Exp. (2014), no. 5, P05004, 19. MR 3224219

[6] Ivan Corwin and Evgeni Dimitrov, Transversal fluctuations of the ASEP, stochastic six vertex model, and Hall-Littlewood Gibbsian line ensembles, Comm. Math. Phys. 363 (2018), no. 2, 435–501. MR 3851820

[7] Ivan Corwin and Alan Hammond, Brownian Gibbs property for Airy line ensembles, Inventiones mathematicae 195 (2014), no. 2, 441–508.

[8] ______., KPZ line ensemble, Probability Theory and Related Fields 166 (2016), no. 1-2, 67–185.

[9] Ivan Corwin and Jeremy Quastel, Crossover distributions at the edge of the rarefaction fan, Ann. Probab. 41 (2013), no. 3A, 1243–1314. MR 3098678

[10] Duncan Dauvergne, Janosch Ortmann, and Bálint Virág, The directed landscape, arXiv:1812.00309 (2018).

[11] Evgeni Dimitrov and Konstantin Matetski, Characterization of Brownian Gibbsian line ensembles, arXiv:2002.00684 (2020).

[12] Victor Dotsenko, Distribution function of the endpoint fluctuations of one-dimensional directed polymers in a random potential, J. Stat. Mech. Theory Exp. (2013), no. 2, P02012, 20. MR 3041935

[13] D. Forster, David R. Nelson, and Michael J. Stephen, Large-distance and long-time properties of a randomly stirred fluid, Phys. Rev. A (3) 16 (1977), no. 2, 732–749. MR 459274

[14] Massimiliano Gubinelli and Nicolas Perkowski, KPZ reloaded, Comm. Math. Phys. 349 (2017), no. 1, 165–269. MR 3592748

[15] M. Hairer, Solving the KPZ equation, XVIIth International Congress on Mathematical Physics, World Sci. Publ., Hackensack, NJ, 2014, p. 419. MR 3204494

[16] Joseph Helfer and Daniel T. Wise, A note on maxima in random walks, Electron. J. Combin. 23 (2016), no. 1, Paper 1.17, 10. MR 3464722
[17] T. Imamura and T. Sasamoto, *Fluctuations of the one-dimensional polynuclear growth model with external sources*, Nuclear Phys. B **699** (2004), no. 3, 503–544. MR 2098552

[18] Takashi Imamura and Tomohiro Sasamoto, *Replica approach to the KPZ equation with the half Brownian motion initial condition*, J. Phys. A **44** (2011), no. 38, 385001, 29. MR 2835150

[19] Mehran Kardar, Giorgio Parisi, and Yi-Cheng Zhang, *Dynamic scaling of growing interfaces*, Phys. Rev. Lett. **56** (1986), 889–892.

[20] Konstantin Matetski, Jeremy Quastel, and Daniel Remenik, *The KPZ fixed point*, arXiv:1701.00018 (2016).

[21] Mihai Nica, Jeremy Quastel, and Daniel Remenik, *One-sided reflected Brownian motions and the KPZ fixed point*, Forum Math. Sigma **8** (2020), Paper No. e63, 16. MR 4190063

[22] Janosch Ortmann, Jeremy Quastel, and Daniel Remenik, *A Pfaffian representation for flat ASEP*, Comm. Pure Appl. Math. **70** (2017), no. 1, 3–89. MR 3581823

[23] Michael Prähofer and Herbert Spohn, *Scale invariance of the PNG droplet and the Airy process*, Journal of Statistical Physics **108** (2002), no. 5-6, 1071–1106.

[24] Jeremy Quastel, *Introduction to KPZ*, Current developments in mathematics, 2011, Int. Press, Somerville, MA, 2012, pp. 125–194. MR 3098078

[25] Dipankar Roy and Rahul Pandit, *One-dimensional Kardar-Parisi-Zhang and Kuramoto-Sivashinsky universality class: limit distributions*, Phys. Rev. E **101** (2020), no. 3, 030103(R), 6. MR 4083969

[26] Tomohiro Sasamoto and Herbert Spohn, *Exact height distributions for the KPZ equation with narrow wedge initial condition*, Nuclear Phys. B **834** (2010), no. 3, 523–542. MR 2628936

[27] Anamaria Savu, *Hydrodynamic scaling limit of continuum solid-on-solid model*, J. Appl. Math. (2006), Art. ID 69101, 37. MR 2231983

[28] Craig A. Tracy and Harold Widom, *Asymptotics in ASEP with step initial condition*, Comm. Math. Phys. **290** (2009), no. 1, 129–154. MR 2520510

[29] ———, *On ASEP with step Bernoulli initial condition*, J. Stat. Phys. **137** (2009), no. 5-6, 825–838. MR 2570751

[30] ———, *Formulas for ASEP with two-sided Bernoulli initial condition*, J. Stat. Phys. **140** (2010), no. 4, 619–634. MR 2670733

[31] H. van Beijeren, R. Kutner, and H. Spohn, *Excess noise for driven diffusive systems*, Phys. Rev. Lett. **54** (1985), no. 18, 2026–2029. MR 789756

[32] S. R. S. Varadhan, *Self-diffusion of a tagged particle in equilibrium for asymmetric mean zero random walk with simple exclusion*, Ann. Inst. H. Poincaré Probab. Statist. **31** (1995), no. 1, 273–285. MR 1340041

[33] Bálint Virág, *The heat and the landscape I*, arXiv: 2008.07241 (2020).

[34] Xuan Wu, *In preparation.*

[35] Lin Xu, *Diffusive scaling limit for mean zero asymmetric simple exclusion processes*, ProQuest LLC, Ann Arbor, MI, 1993, Thesis (Ph.D.)–New York University. MR 2690363

(J. Quastel) Department of Mathematics, University of Toronto, 40 St. George Street, Toronto, Ontario, Canada M5S 2E4

Email address: quastel@math.toronto.edu

(S. Sarkar) University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WB.

Email address: ss2871@cam.ac.uk