# Hairer-Quastel universality in non-stationarity via energy solution theory[*]

Kevin Yang[†]

## Abstract

The paper addresses *probabilistic* aspects of the KPZ equation and stochastic Burgers equation by providing a solution theory that builds on the energy solution theory [15, 19, 20, 23]. The perspective we adopt is to study the stochastic Burgers equation by writing its solution as a probabilistic solution [22] plus a term that can be studied with deterministic PDE considerations. One motivation is universality of KPZ and stochastic Burgers equations for a certain class of stochastic PDE growth models, first studied in [28]. For this, we prove universality for SPDEs with general nonlinearities, thereby extending [28, 29], and for many non-stationary initial data, thereby extending [21].

Our perspective lets us also prove explicit rates of convergence to white noise invariant measure of stochastic Burgers for *non-stationary* initial data, in particular extending the spectral gap result of [23] beyond stationary initial data, though for non-stationary data our convergence will be measured in Wasserstein distance and relative entropy, not via the spectral gap as in [23]. Actually, we extend the spectral gap in [23] to a log-Sobolev inequality. Our methods can also analyze fractional stochastic Burgers equations [23]; we discuss this briefly.

Lastly, we note our perspective on the KPZ and stochastic Burgers equations gives a first *intrinsic* notion of solutions for general continuous initial data, in contrast to Hölder regular data needed for regularity structures [26], paracontrolled distributions [14], and Hölder-regular Brownian bridge data for energy solutions [19, 20].

---

[†]University of California, Berkeley, United States of America.
 E-mail: kevin_yang95@berkeley.edu

## 1 Introduction

The *Kardar-Parisi-Zhang* (KPZ) equation is a conjecturally canonical model for interface fluctuations derived in [30] via non-rigorous renormalization group heuristics. Such "interface fluctuations" include flame propagation, infected individuals in an epidemic, bacterial colony and tumor growth, and crack formations; see [8] and [30]. However, rigorous proof of this so-called "weak KPZ universality" problem has posed a challenging mathematical problem, primarily because the KPZ equation is a *singular stochastic PDE*; there are ultraviolet "UV" divergences whose rigorous analysis is highly nontrivial and was ultimately resolved by Hairer [25] using a first version of what would be later developed into a theory of regularity structures [26]. We explain this singular nature of KPZ and related SPDEs in the context of weak universality. First, however, let us introduce the KPZ equation below, which we pose on the space-time set $\mathbb{R}_{\geqslant 0} \times \mathbb{T}$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the unit one-dimensional torus, and introduce a Gaussian space-time white noise on $\mathbb{T}$, defined to be a Gaussian random field with covariance kernel formally given by $\mathbf{E}\xi_{\mathfrak{t},\mathrm{x}}\xi_{\mathfrak{s},\mathrm{y}} = 2\delta_{\mathfrak{t}=\mathfrak{s}}\delta_{\mathrm{x}=\mathrm{y}}$:

$$\partial_{\mathfrak{t}}\mathbf{h} = \Delta\mathbf{h} + \beta|\nabla\mathbf{h}|^2 + \xi. \tag{1.1}$$

The constant $\beta$ is taken to be a non-negative number; its sign is not too important as we can always take $-\mathbf{h}$ instead of $\mathbf{h}$, and the noise is statistically invariant under sign change. The $\beta$ constant will eventually be a homogenized coefficient in our main universality result. We will be mostly concerned with generalizations of (1.1) in which we replace $\beta|\nabla\mathbf{h}|^2$ by a general function of $\nabla\mathbf{h}$. In particular, let us define $\mathbf{g}^\varepsilon$ as the solution to the following SPDE, in which $\mathbf{F}$ is a general function and $\varepsilon$ is a small positive parameter we eventually take to zero in order to recover the KPZ equation:

$$\partial_{\mathfrak{t}}\mathbf{g}^\varepsilon = \Delta\mathbf{g}^\varepsilon + \varepsilon^{-1}\mathbf{F}(\varepsilon^{1/2}\nabla\mathbf{g}^\varepsilon) + \xi. \tag{1.2}$$

We eventually consider solutions to a slightly and technically modified version of (1.2), but for transparency of illustration we consider formally (1.2). To illustrate weak universality and singular behavior, one can Taylor expand the $\mathbf{F}$ nonlinearity on the RHS of (1.2) about its value at 0. The first two terms, which are constant and linear in $\nabla\mathbf{g}^\varepsilon$, can be removed when we insert them back into the SPDE (1.2) by global time-shift and linear change-of-reference along a constant speed characteristic. We are left with the quadratic term in $\nabla\mathbf{g}^\varepsilon$ plus terms that are formally vanishing in the small $\varepsilon$ limit, so we are led to believe $\mathbf{g}^\varepsilon$ converges to (1.1) for $\beta = 2^{-1}\mathbf{F}''(0)$. The convergence to KPZ is true, but the argument is *wrong* as the effective coefficient is global in $\mathbf{F}''$. We make precise the correct effective coefficient later in this introduction, but we remark some version of both claims in the previous sentence are proved in [21, 28, 29].

In [28, 29], the method of proving universality is based on the theory of regularity structures in [26]. In [21], the method of proof is based on *energy solution theory* [15, 20]. On the one hand, because regularity structures is based on pathwise and analytic considerations, the notion of convergence in [28, 29] is much stronger than that in [21]. Moreover, the nature of the smoothing of (1.1) is much more general than in [21], as [21] strongly depends on both the Markovian structure of the smoothed SPDEs (1.2) and the explicit nature of invariant measures. On the other hand, the paper [21] treats a much larger set of nonlinearities than [28, 29]. For example, [21] can treat (1.2) for $\mathbf{F}(\mathrm{x}) = |\mathrm{x}|$, but [28, 29] are far from treating this non-smooth nonlinearity.

Currently, however, energy solution theory [21] is exclusively applicable to Brownian initial data. One point of this paper is to extend [21] to arbitrary continuous initial conditions for (1.1) and extend the energy solution theory from stationary Brownian initial data to general continuous initial data; this would actually provide a first intrinsic

solution theory for singular SPDEs with arbitrary continuous initial data instead of quantitatively Hölder initial data.

Before we proceed with a precise discussion of universality, however, we must introduce the following *stochastic Burgers equation* (SBE), whose solution $\mathbf{u}$ is related to $\mathbf{h}$ from (1.1) by taking a weak derivative $\mathbf{u} = \nabla\mathbf{h}$:

$$\partial_t \mathbf{u} = \Delta \mathbf{u} + \beta\nabla(\mathbf{u}^2) + \nabla\xi. \tag{1.3}$$

Technically, results of [21] are stated at the level of (1.3), not (1.1), but as shown in [20] the difference between (1.1) and (1.3), which is a constant eliminated via differentiation, poses no problem. In particular, our aforementioned extension of energy solutions [15, 20] will study (1.3) instead of (1.1), but it can treat (1.1) without much difficulty.

Having introduced the stochastic Burgers equation (1.3), let us now introduce exactly what we mean by an "energy solution". The following is taken verbatim from the beginning of Section 2 of [22] instead of [20]. We explicitly mention the differences between notions of energy solution in [22] and [20] below as well.

- Suppose $\mathbf{u} \in \mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$, where $\mathscr{D}(\mathbb{T})$ is the topological dual of $\mathscr{C}^\infty(\mathbb{T})$ (the latter is equipped with the usual Frechet topology). Suppose there exists $\rho \in \mathscr{C}_c^\infty(\mathbb{R})$ such that $\int_{\mathbb{R}} \rho(\mathrm{x})\mathrm{dx} = 1$ and, for $\rho^N(\cdot) = N\rho(N\cdot)$, the limit

$$\int_0^t \nabla(\mathbf{u}_{\mathfrak{s}})^2 \mathrm{d}\mathfrak{s}(\Phi) := \lim_{N\to\infty} \int_0^t (\mathbf{u}_{\mathfrak{s}} * \rho^N)^2(-\nabla\Phi)\mathrm{d}\mathfrak{s} \tag{1.4}$$

  exists for all $\Phi \in \mathscr{C}^\infty(\mathbb{T})$ locally uniformly in $t$ in probability. Above, the convolution is convolution on the torus.
- Suppose in addition that for any $\Phi \in \mathscr{C}^\infty(\mathbb{T})$, the process

$$\mathbf{B}(t;\Phi) := \mathbf{u}_t(\Phi) - \mathbf{u}_0(\Phi) - \int_0^t \mathbf{u}_{\mathfrak{s}}(\Delta\Phi)\mathrm{d}\mathfrak{s} - \int_0^t \nabla(\mathbf{u}_{\mathfrak{s}})^2\mathrm{d}\mathfrak{s}(\Phi) \tag{1.5}$$

  is a martingale in $t \geqslant 0$ with quadratic variation $2t\int_{\mathbb{T}} |\nabla\Phi(\mathrm{x})|^2\mathrm{dx}$. Lastly, suppose the following energy estimate holds:

$$\mathbf{E}|\int_{\mathfrak{s}}^t (\mathbf{u}_r * \rho^N)^2(-\nabla\Phi) - (\mathbf{u}_r * \rho^M)^2(-\nabla\Phi)\mathrm{d}r|^2 \lesssim \frac{|t - \mathfrak{s}|}{M \wedge N}\int_{\mathbb{T}} |\nabla\Phi(\mathrm{x})|^2\mathrm{dx}. \tag{1.6}$$

- If the two bullet points above hold, we say $\mathbf{u}$ is an energy solution to (1.3).

In [20], a backwards representation of the $\mathbf{u}$ process is also required. However, this backwards representation is used only for some technical estimates. As noted in [22] (see Theorem 2.8, for example), this backwards representation can be dropped in cases that we eventually specialize to without losing this technical benefit.

In turns out that our approach to universality (Theorem 1.5) will be intimately connected to establishing explicit rates of convergence to the invariant measure for (1.3) that were previously missing from the literature for *general* continuous initial conditions; the paper [23] studied spectral gaps that only provide information, a priori, for initial data very close to the white noise invariant measure, which turn out to be basically obtainable via appropriate discretization of (1.3) by Ornstein-Uhlenbeck processes plus an asymmetric interaction. Actually, we will enhance the spectral gap in [23] to a much more powerful log-Sobolev inequality using basically the same discretization method via Fourier smoothing.

We conclude this introduction by noting our work successfully treats *general continuous* data, going well beyond the usual class of Hölder regular initial data required

to analyze SPDEs like (1.1) and (1.3) via regularity structures [26] or paracontrolled distributions [14]. In particular, this highlights the utility of our perspective for (1.1) and (1.3), which ultimately agree with the same solutions obtained via other methods [14, 19, 26]. It also even provides a first *intrinsic* solution theory to SPDEs like (1.1) and (1.3) for general continuous initial data; although a to-be-introduced Cole-Hopf solution theory also treats such initial data, this requires [4] appealing to an auxiliary SPDE.

## 1.1 Universality

We start with the following construction or ansatz that we clarify afterwards.

**Definition 1.1.** *Let us first define a "UV projection" map $\Pi_N : \mathscr{L}^2(\mathbb{T}) \to \mathscr{L}^2(\mathbb{T})$, in which $N$ is a positive integer, via the basis-prescription $\Pi_N(\exp(2\pi i k \mathrm{x})) = \mathbf{1}(|k| \leqslant N) \exp(2\pi i k \mathrm{x})$ for all $k \in \mathbb{Z}$. We now define $\mathbf{h}^{N,1}$ as the solution to the following SPDE on $\mathbb{R}_{\geqslant 0} \times \mathbb{T}$ with regularized noise:*

$$\partial_t \mathbf{h}^{N,1} \;=\; \Delta \mathbf{h}^{N,1} + \varepsilon_N^{-1} \Pi_N \mathbf{F}(\varepsilon_N^{1/2} \nabla \mathbf{h}^{N,1}) + \Pi_N \xi; \tag{1.7}$$

*In (1.7), $\varepsilon_N = N^{-1}\pi$ is a normalization factor used in [21]. The function $\mathbf{F}$ is general; we specify conditions on it later in Assumption 1.3.*

*We additionally define $\mathbf{h}^{N,2}$ to be the solution to the following noise-free PDE on $\mathbb{R}_{\geqslant 0} \times \mathbb{T}$ that is "controlled" by $\mathbf{h}^{N,1}$, in the sense that the following PDE has randomness coming only from $\mathbf{h}^{N,1}$:*

$$\partial_t \mathbf{h}^{N,2} \;=\; \Delta \mathbf{h}^{N,2} + \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1} + \mathbf{h}^{N,2})) - \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla \mathbf{h}^{N,1}). \tag{1.8}$$

*Lastly, we define $\mathbf{h}^N = \mathbf{h}^{N,1} + \mathbf{h}^{N,2}$, we define $\mathbf{u}^{N,1} = \nabla \mathbf{h}^{N,1}$, and we define $\mathbf{u}^N = \nabla \mathbf{h}^N = \nabla \mathbf{h}^{N,1} + \nabla \mathbf{h}^{N,2}$. Let us emphasize that $\mathbf{u}^{N,1}$ satisfies the regularized SPDE given below on $\mathbb{R}_{\geqslant 0} \times \mathbb{T}$:*

$$\partial_t \mathbf{u}^{N,1} \;=\; \Delta \mathbf{u}^{N,1} + \varepsilon_N^{-1} \Pi_N \nabla \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1}) + \Pi_N \nabla \xi. \tag{1.9}$$

**Remark 1.2.** Because (1.8) is a nonlinear PDE, let us must mention exactly what we mean by "solution". Precisely, we require $\mathbf{h}^{N,2}(\mathfrak{t}, \cdot) \in \mathscr{C}^1(\mathbb{T})$, where $\mathscr{C}^1(\mathbb{T})$ is the space of continuously differentiable functions on $\mathbb{T}$. Moreover, we require $\|\mathbf{h}^{N,2}(\mathfrak{t}, \cdot)\|_{\mathscr{C}^1(\mathbb{T})}$ is continuous in $\mathfrak{t}$. We also require $\mathbf{h}^{N,2}$ to solve

$$\begin{aligned} \mathbf{h}^{N,2}(\mathfrak{t}, \mathrm{x}) \;=\;& (\mathrm{e}^{\mathfrak{t}\Delta} \mathbf{h}^{N,2}(0, \cdot))(\mathrm{x}) \\ &+ \int_0^{\mathfrak{t}} \left( \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1}(\mathfrak{s}, \cdot) + \mathbf{h}^{N,2}(\mathfrak{s}, \cdot))) - \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla \mathbf{h}^{N,1}(\mathfrak{s}, \cdot)) \right) \right) \mathrm{d}\mathfrak{s}. \end{aligned}$$

Existence and uniqueness of such mild solutions to (1.8) will be guaranteed under our assumptions on $\mathbf{F}$, as long as $\mathbf{h}^{N,2}(0, \cdot) \in \mathscr{C}^1(\mathbb{T})$; see Lemma A.3.

Let us now explain Definition 1.1, beginning with the $\mathbf{h}^{N,1}$ equation. Let us specialize to (1.7) for $\mathbf{F}(\mathrm{x}) = \mathrm{x}^2$. In this case, in [19, 20] it is shown that (1.7) with this quadratic nonlinearity $\mathbf{F}$ is an appropriate regularization of the continuum KPZ equation (1.1) that agrees with the Cole-Hopf solution of [4] and regularity structures solution of [25], *if* the initial data to (1.7) is a Brownian bridge on $\mathbb{T}$ (that is independent of the space-time white noise $\xi$) or a random initial data whose law as a probability measure on $\mathscr{C}(\mathbb{T})$ has relative entropy that is uniformly bounded with respect to Brownian bridge measure. Actually, as noted in [13] the choice of $\Pi_N$ regularization is not so important, and we can actually use a much smoother cutoff (for example convolution by a time-1 heat kernel with Fourier cutoff at level $N$) for which the results of [19, 20] still hold.

We proceed to explain $\mathbf{h}^{N,2}$. Because the solution to (1.9) is supported on non-zero Fourier modes $|k| \leqslant N$, we can replace $\mathbf{u}^{N,1}$ in (1.9) by $\Pi_N^{-1} \mathbf{u}^{N,1}$. We can also replace

$\Pi_N \nabla \xi$ by $\Pi_N^2 \nabla \xi$. Then, applying $\Pi_N^{-1}$ to (1.9) formally gives the stochastic Burgers equation with general $\mathbf{F}$ and smoothed noise (this is formal because $\Pi_N$ is not invertible). Moreover, in [21], convergence of $\mathbf{u}^{N,1}$ means convergence of fixed Fourier modes, and applying $\Pi_N^{-1}$ to $\mathbf{u}^{N,1}$ does not affect this convergence. We point out that this formal relation between (1.9) and stochastic Burgers with general $\mathbf{F}$ can be made somewhat rigorous in the large-$N$ limit. In particular, in the quadratic case $\mathbf{F}(\mathbf{x}) = \mathbf{x}^2$, the $\Pi_N$-Fourier smoothing operator in front of $\mathbf{F}$ in the $\mathbf{h}^{N,1}$ equation (1.7) does not change the large-$N$ limit of the solution $\mathbf{h}^{N,1}$, as in either case one gets the same solution to the KPZ equation; see [20, 25]. Thus, let us pretend (1.7) does not have $\Pi_N$ in front of $\mathbf{F}(\mathbf{x}) = \mathbf{x}^2$. Then if $\mathbf{h}^{N,2}$ solves (1.8), it is easy to see that $\mathbf{h}^N = \mathbf{h}^{N,1} + \mathbf{h}^{N,2}$ solves the same equation (1.7) as $\mathbf{h}^{N,1}$ without $\Pi_N$ in front of $\mathbf{F}$, though with different initial data. So for the specific regularization $\Pi_N$ in Definiiton 1.1, we formally write the solution to a regularized stochastic Burgers equation with general nonlinearity into the two pieces (1.7) and (1.8). We finish this explanation of Definition 1.1 by noting the decomposition $\mathbf{h}^N = \mathbf{h}^{N,1} + \mathbf{h}^{N,2}$ is by no means canonical, namely if we think of $\mathbf{h}^N$ as a solution to a regularized SPDE. In particular, we will specify an initial data for $\mathbf{h}^N$, and we have a choice as to how to decompose this initial data into initial data for $\mathbf{h}^{N,1}$ and $\mathbf{h}^{N,2}$ equations. We clarify this in Theorem 1.5.

Let us now introduce the necessary assumptions for the nonlinearity $\mathbf{F}$ in Definition 1.1 and the initial data of consideration for $\mathbf{h}^N$, which we re-emphasize is, technically, a pair of initial data for $\mathbf{h}^{N,1}$ and $\mathbf{h}^{N,2}$, respectively. We will provide explanation afterwards.

**Assumption 1.3.** The nonlinearity $\mathbf{F} : \mathbb{R} \to \mathbb{R}$ in *Definition 1.1* is uniformly Lipschitz. Moreover, if $\mathbf{P}^{\mathrm{G}}$ is standard Gaussian measure $\mathcal{N}(0,1)$ on $\mathbb{R}$, then $\mathbf{F}, \mathbf{F}' \in \mathscr{L}^2(\mathbb{R}, \mathbf{P}^{\mathrm{G}})$.

**Assumption 1.4.** Assume $\mathbf{h}^N(0, \cdot) = \Pi_N \mathbf{h}(0, \cdot)$, where $\mathbf{h}(0, \cdot)$ is a continuous function on $\mathbb{T}$ independent of $\xi$ such that $\mathbf{h}(0,0) = 0$, where $0 \in \mathbb{T}$ is interpreted by embedding $\mathbb{T} = \mathbb{R}/\mathbb{Z} \simeq [0,1)$, and $\mathbf{h}^N(0, \cdot) = \Pi_N \mathbf{h}(0, \cdot) \to \mathbf{h}(0, \cdot)$ uniformly on $\mathbb{T}$ with probability 1.

Let us first explain Assumption 1.3. It is not difficult to see [19] that the Fourier coordinates of $\mathbf{u}^{N,1}$ in (1.9) solves a finite-dimensional SDE. The Lipschitz assumption guarantees this SDE has global solutions in time. In particular, we could assume a.e. differentiability of $\mathbf{F}$ instead and prove results until blow-up times, but we choose to take the Lipschitz assumption for convenience and clarity of presentation. The *uniformly* Lipschitz assumption is also used for deriving global-in-time solutions to (1.8) for each fixed $N$. Similarly, we could impose locally Lipschitz instead of uniformly Lipschitz and work until a random blow-up time (for classical solutions) for (1.8). We choose uniformly Lipschitz again for convenience and clarity. There are other assumptions besides uniformly Lipschitz, however, that would ensure (classical) global solutions for both (1.9) and (1.8). For example, establishing global solutions for (1.9) requires only locally Lipschitz $\mathbf{F}$, since the invariant measure for (1.9) extends local solutions to global ones; see Lemma A.1. Global solutions for (1.8), which is a viscous Hamilton-Jacobi equation with random Hamiltonian, can be obtained for continuously differentiable $\mathbf{F}$ satisfying a polynomial bound for any arbitrary but fixed degree; see Theorem 3.1 in [3]. (We note [3] is not published, and the proof of Theorem 3.1 therein is quite complicated. Also, in [29], the specific choice $\mathbf{F}(\mathbf{x}) = |\mathbf{x}|$ was mentioned as an interesting problem, and this choice satisfies Assumption 1.3. For these reasons, we give details just for uniformly Lipschitz $\mathbf{F}$.) It also seems possible to remove the polynomial bound on $\mathbf{F}$ and upgrade its regularity to $\mathscr{C}^{1,\alpha}$. In any case, all these choices for conditions on $\mathbf{F}$ would generalize what is done in [29].

Moreover, the condition $\mathbf{F}, \mathbf{F}' \in \mathscr{L}^2(\mathbb{R}, \mathbf{P}^{\mathrm{G}})$ comes from the assumption on nonlinearities in [21], which serves as a stationary model of what is considered herein. Technically, it means that the Hermite polynomial expansions, or equivalently eigenfunction expan-

sion in $\mathscr{L}^2(\mathbb{R}, \mathbf{P}^{\mathrm{G}})$ with respect to the Ornstein-Uhlenbeck operator of Malliavin calculus, converges in $\ell^2(\mathbb{Z})$ with both the standard constant weight and with the weight $|k|$. This is similar to how a function and its derivative belong to $\mathscr{L}^2(\mathbb{T})$ given its Fourier series converges in $\ell^2(\mathbb{Z})$ after we multiply its $k$-th Fourier coordinate by $1 + |k|$. Let us make this precise. By Gaussian integration-by-parts, we get $2^{-1}\mathbf{E}^{\mathrm{G}}\mathbf{F}'' = 2^{-1}\mathbf{E}^{\mathrm{G}}\mathbf{F}\mathsf{H}_2$, where $\mathsf{H}_2$ is the second Hermite polynomial with respect to the Gaussian measure $\mathbf{P}^{\mathrm{G}}$ as in [21]. In particular, we note $\mathbf{E}^{\mathrm{G}}\mathbf{F}'' = \mathbf{E}^{\mathrm{G}}\mathbf{F}\mathsf{H}_2 < \infty$. Indeed, the Hermite polynomial $\mathsf{H}_2$ is a unit-length Hermite basis vector in $\mathscr{L}^2(\mathbb{R}, \mathbf{P}^{\mathrm{G}})$, so $\mathbf{E}^{\mathrm{G}}\mathbf{F}\mathsf{H}_2 \leqslant (\mathbf{E}^{\mathrm{G}}\mathbf{F}^2)^{1/2}$, and this last second moment is finite by Assumption 1.3.

Let us now explain Assumption 1.4. We first note that it only requires almost sure continuity of the initial data function $\mathbf{h}(0, \cdot)$, instead of quantitative Hölder regularity as in [21, 28, 29], making Assumption 1.4 quite general. Observe that $\mathbf{h}(0, \cdot)$ is allowed to be random; moreover, if $\mathbf{h}(0, \cdot)$ is a Brownian bridge on $\mathbb{T}$, defined by independent Gaussian random variables of variance $1 + |k|^{-2}$ for $k \neq 0$ Fourier coordinate and assumed to be independent of $\xi$, then the convergence $\Pi_N \mathbf{h}(0, \cdot) \to \mathbf{h}(0, \cdot)$ is classical. We clarify the $k = 0$ Fourier coefficient for Brownian bridge just has to be chosen independent of the $k \neq 0$ coefficients and of $\xi$. We emphasize, again, that we can replace $\Pi_N$ with a smoother Fourier cutoff if desired; this would make Assumption 1.4 and the required uniform convergence less strict; let us remark that some convergence is needed in Assumption 1.4 [28, 29], as otherwise large-$N$ limits of $\mathbf{h}^N$ are out of the question. Lastly, when we refer to Assumption 1.4 in Theorem 1.5, we will specify ways of decomposing initial data $\mathbf{h}^N$ into initial data for $\mathbf{h}^{N,1}$ and $\mathbf{h}^{N,2}$ equations of Definition 1.1.

We now present the first main theorem of this article. To this end, we must first specify a theory of solutions to (1.1) and (1.3) without regularization; note Definition 1.1 addresses only Fourier regularized (S)PDEs. We will employ the following *Cole-Hopf solution theory*, which agrees with regularity structures [26], paracontrolled distributions [14], and energy solution theory [19, 20], at least for sufficiently Hölder continuous initial data (and for energy solutions, only for initial data with uniformly bounded "relative entropy" with respect to Brownian bridge initial data; see the definition of relative entropy in Section C, though we will re-clarify this when more relevant). The Cole-Hopf solution to (1.1) is defined (and explained) as follows.

- The Cole-Hopf solution is defined as $\mathbf{h} = \beta^{-1}\log\mathbf{z}$, where $\partial_t\mathbf{z} = \Delta\mathbf{z} + \beta\mathbf{z}\xi$ is a linear SPDE on $\mathbb{R}_{\geqslant 0} \times \mathbb{T}$ called the *stochastic heat equation* (SHE).

- In [32], an almost sure comparison principle was shown for SHE. Thus, the logarithm is well-defined as long as we start $\mathbf{z}$ with positive (continuous) initial data.

- We define the Cole-Hopf solution to the stochastic Burgers equation (1.3) to be $\mathbf{u} = \nabla\mathbf{h} = \beta^{-1}\log\mathbf{z}$, where the derivative is interpreted in the weak sense.

We clarify that the Cole-Hopf solution agrees with the solution of KPZ via regularity structures only if we include a divergent counter-term in (1.1). In particular, if we were to claim the convergence of $\mathbf{h}^{N,1}$ in Definition 1.1 to the Cole-Hopf solution of (1.1), we would need to introduce a term of the form $-C_N$ to (1.7), where $C_N$ is constant and $|C_N| \to \infty$ as $N \to \infty$. However, our results are for the stochastic Burgers equation, for which this constant $C_N$ plays no role because we take its spatial derivative. Lastly, we note that the space $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$ of continuous $\mathscr{D}(\mathbb{T})$-valued paths is equipped with the locally uniform (in $\mathbb{R}$) topology. Here, $\mathscr{D}(\mathbb{T})$ is the space of generalized functions on $\mathbb{T}$, or equivalently the topological dual of $\mathscr{C}^\infty(\mathbb{T})$.

**Theorem 1.5.** *Consider initial data* $\mathbf{h}(0, \cdot)$ *as in* Assumption 1.4, *and assume* $\mathbf{E}^{\mathrm{G}}\mathbf{F}'' \neq 0$. *There exists a one-parameter family of decompositions* $\mathbf{h}^N = \mathbf{h}^{N,1,\varepsilon} + \mathbf{h}^{N,2,\varepsilon}$, *as in* Definition 1.1, *parameterized by* $\varepsilon > 0$ *such that:*

- *For any $\varepsilon > 0$, the function $\mathbf{h}^{N,1,\varepsilon}$ solves (1.7) with initial data given by $\Pi_N$ acting on Brownian bridge on $\mathbb{T}$ independent of $\xi$ and conditioned to be within $\varepsilon$ of $\mathbf{h}(0, \cdot)$. Moreover, the gradient $\mathbf{u}^{N,1,\varepsilon} = \nabla \mathbf{h}^{N,1,\varepsilon}$ converges as $N \to \infty$ to the Cole-Hopf solution of (1.3) with $\beta = 2^{-1}\mathbf{E}^{\mathrm{G}}\mathbf{F}''$ and with initial data given by the weak derivative of Brownian bridge independent of $\xi$ and conditioned to be within $\varepsilon$ of $\mathbf{h}(0, \cdot)$. Here, $\mathbf{E}^{\mathrm{G}}$ is expectation with respect to Gaussian measure $\mathbf{P}^{\mathrm{G}}$ in Assumption 1.3.*
- *For any $\varepsilon > 0$, $\mathbf{h}^{N,2,\varepsilon}$ solves (1.8) with initial data given by the difference $\mathbf{h}^N(0, \cdot) - \mathbf{h}^{N,1,\varepsilon}(0, \cdot)$. Moreover, we have $|\mathbf{h}^{N,2,\varepsilon}(\mathbf{t}, \mathbf{x})| \leqslant \varepsilon + \mathrm{o}_N$ for all $\mathbf{t} \geqslant 0$ and $\mathbf{x} \in \mathbb{T}$ with probability 1, where $\mathrm{o}_N \to 0$ almost surely and uniformly in $\mathbf{t} \geqslant 0$ and $\mathbf{x} \in \mathbb{T}$.*

*We deduce that the law of $\mathbf{u}^N = \nabla \mathbf{h}^N = \nabla \mathbf{h}^{N,1,\varepsilon} + \nabla \mathbf{h}^{N,2,\varepsilon}$, as a probability measure on $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$, converges to the Cole-Hopf solution of (1.3) with initial data given by the weak derivative $\nabla \mathbf{h}(0, \cdot)$.*

Theorem 1.5 claims that we can approximate the regularization $\mathbf{h}^N$ of the KPZ equation *uniformly* in $N$ by a solution $\mathbf{h}^{N,1,\varepsilon}$ to (1.7), which we will ultimately analyze via the work in [21]. The residual term $\mathbf{h}^{N,2,\varepsilon}$ solves (1.8); we emphasize the nontrivial aspect of this statement is two-fold. First, we must find suitable approximations to $\mathbf{h}^N$ that can be analyzed by using energy solution theory [19, 20, 21, 22]. Since energy solution theory is mostly dependent on initial data and invariant measures, this approximation will be made at the level of initial data. The problem we are left with is to *globally propagate* this approximation uniformly in $N$. In particular, we must control $\mathbf{h}^{N,2,\varepsilon}$, which solves a type of Hamilton-Jacobi equation with divergent Hamiltonian. Note that we must also extend [21] to initial conditions that are not exactly Brownian bridge but relative entropy perturbations as well, though this will not be difficult provided the work of [17, 22]. We clarify Theorem 1.5 gives a first application of the energy solution theory to a very general class of non-stationary SPDE growth models; in particular, we only require the modest requirements in [20] for the models of interest that are much more general than the required conditions on the nonlinearity in [28, 29]. In particular, one can *define* the large-$N$ limit of $\mathbf{u}^N = \nabla \mathbf{h}^N$ in Theorem 1.5 as an *intrinsic* solution to (1.3) for such general continuous initial data.

Note we have assumed $\mathbf{E}^{\mathrm{G}}\mathbf{F}'' \neq 0$. It is certainly possible that this does not hold, i.e. $\mathbf{E}^{\mathrm{G}}\mathbf{F}'' = 0$. In this case, the limit SPDE is not the Cole-Hopf solution of (1.3) but rather the solution to the linear SPDE given by (1.3) with $\beta = 0$. The proof of Theorem 1.5 given in this paper covers this case as well, since our method is by comparison to very-close-to-stationary initial data, and the result for stationary initial data in [21] includes the case $\beta = 0$.

## 1.2 Convergence to invariant measure

As a byproduct of our proof of Theorem 1.5, we will be able to show that the stochastic Burgers equation (SBE), starting with the gradient of any continuous initial data, converges to the Gaussian white noise measure on $\mathscr{D}(\mathbb{T})$ in a quantitative fashion. Moreover, we provide an explicit rate of convergence; this is new as [27] gives only uniqueness and ergodicity of white noise invariant measure for SBE, and [23] provides a spectral gap estimate, which only provides quantitative bounds for initial data to SBE that is very close to the white noise invariant measure.

We now state the main result of this subsection. In Theorem 1.6, the metric W denotes Wasserstein distance in Appendix B, for which we take the Polish space $\mathscr{X}$ to be the topological dual of the Sobolev space on $\mathbb{T}$ of bounded functions with bounded derivative. The object H is the relative entropy of Appendix C, in which the state space is $\mathscr{S} = \mathscr{D}(\mathbb{T})$. We clarify the statement of the following result afterwards.

**Theorem 1.6.** *Take any deterministic function* $\mathbf{h}(0, \cdot) \in \mathscr{C}(\mathbb{T})$ *and let* $\mathbf{u}$ *be the Cole-Hopf solution to (1.3) with initial data* $\mathbf{u}(0, \cdot) = \nabla \mathbf{h}(0, \cdot)$. *If* $\mathbf{L}(\mathfrak{t})$ *denotes the law of* $\mathbf{u}(\mathfrak{t}, \cdot)$ *as a probability measure on* $\mathscr{D}(\mathbb{T})$ *and if* $\eta$ *denotes Gaussian white noise invariant measure, then for any positive* $\varepsilon$, *there exists a measure* $\mathbf{L}^\varepsilon(\mathfrak{t})$ *such that*

$$\mathsf{W}(\mathbf{L}(\mathfrak{t}), \mathbf{L}^\varepsilon(\mathfrak{t})) \leqslant \varepsilon \quad \text{and} \quad \mathsf{H}(\mathbf{L}^\varepsilon(\mathfrak{t})|\eta) \leqslant \exp(-C\mathfrak{t})\kappa. \tag{1.10}$$

*The constant* $C$ *is independent of all other parameters in this statement. The constant* $\kappa = \kappa(\varepsilon, \mathbf{h})$ *depends only on the modulus of continuity of* $\mathbf{h}(0, \cdot)$ *and* $\varepsilon$. *For example, if* $\mathbf{h}(0, \cdot)$ *is Hölder continuous, then* $\kappa(\varepsilon, \mathbf{h})$ *would depend only on the Hölder norm of* $\mathbf{h}(0, \cdot)$ *and on* $\varepsilon$.

We emphasize the upper bounds in (1.10) both vanish if we take $\mathfrak{t} \to \infty$ then $\varepsilon \to 0$.

Theorem 1.6 builds on the spectral gap of [23], first by improving to a log-Sobolev inequality (LSI) via relatively straightforward considerations. Indeed, if $\nabla \mathbf{h}(0, \cdot)$ has a measure with an $\mathscr{L}^2$-Radon-Nikodym derivative with respect to white noise measure, then $\kappa$ equals the $\mathscr{L}^2$-norm of said Radon-Nikodym derivative and $\varepsilon = 0$ in (1.10) would be okay. The point is that having an $\mathscr{L}^2$-Radon-Nikodym derivative is an extremely strong condition that, although is satisfied for a "dense" class of initial data, is not satisfied for "generic" initial data to SBE. The Wasserstein bound in (1.10) says to first approximate gradients of general continuous initial data by a member of the aforementioned dense class. As with Theorem 1.5, we then globally propagate this comparison. This will give the Wasserstein estimate in (1.10). Ultimately, Theorem 1.6 says that the law of the Cole-Hopf solution to (1.3) with initial data given by the (weak) derivative of any general continuous function is within $\varepsilon$ of a solution that is statistically close to white noise after long times via LSI, but the comparison with general data is in a different (Wasserstein) topology.

### 1.3 Organization

We first prove Theorem 1.5; the proof of Theorem 1.6 will follow from steps in the proof of Theorem 1.5 except for a log-Sobolev inequality for the stochastic Burgers equation and its $\Pi_N$-UV cutoffs. Lastly, before the appendix sections, which give definitions and basic properties of key functionals we use for the reader's convenience, we comment on extensions of our methods to fractional stochastic Burgers equations.

## 2 Proof of Theorem 1.5

The main objectives of this section, which are the main ingredients to proving Theorem 1.5, are listed below.

- We begin with a construction of $\mathbf{h}^{N,1,\varepsilon}$, in which $\varepsilon$ is a fixed positive parameter. In particular, we specify the initial condition for $\mathbf{h}^{N,1,\varepsilon}$, which then determines it uniquely because it must then solve (1.7). Observe that $\mathbf{h}^{N,2,\varepsilon}$ is automatically determined once we construct $\mathbf{h}^{N,1,\varepsilon}$ and its initial condition, because the initial condition for $\mathbf{h}^{N,2,\varepsilon}$ is then given by the difference between the initial data $\mathbf{h}^N(0, \cdot) = \Pi_N \mathbf{h}(0, \cdot)$ and the initial data $\mathbf{h}^{N,1,\varepsilon}(0, \cdot)$. Again, once we specify initial data of $\mathbf{h}^{N,2,\varepsilon}$, we specify $\mathbf{h}^{N,2,\varepsilon}$ uniquely because it must solve the equation (1.8), which has a unique solution because it is a classical parabolic Hamilton-Jacobi-type equation [12].

- We then proceed to establish the two bullet points in Theorem 1.5; the last statement therein will follow by a standard argument. For $\mathbf{h}^{N,2,\varepsilon}$, we employ a maximum principle whereas for $\mathbf{h}^{N,1,\varepsilon}$, we employ the energy solution theory in [21] combined with relative entropy estimates as with [17, 22]; for this, we must prove a relative

entropy estimate for the law of $\mathbf{h}^{N,1,\varepsilon}(0,\cdot)$ with respect to Brownian bridge initial data.

- First assume Theorem 1.5 were true for *deterministic* continuous initial data $\mathbf{h}(0,\cdot)$ for which $\Pi_N \mathbf{h}(0,\cdot)$ converges to $\mathbf{h}(0,\cdot)$ uniformly on $\mathbb{T}$. We would be able to show Theorem 1.5 for the allowed random continuous initial data in Assumption 1.4 by conditioning on said random initial data, tossing out probability zero events. Indeed, convergence as measures is a deterministic statement, namely a statement about convergence of expectations of test functions. Convergence almost surely is also deterministic if we toss out all probability zero events. We will therefore assume that $\mathbf{h}(0,\cdot)$ is deterministic and continuous, and that $\Pi_N \mathbf{h}(0,\cdot)$ converges to $\mathbf{h}(0,\cdot)$ uniformly on $\mathbb{T}$.

## 2.1 Construction of $\mathbf{h}^{N,1,\varepsilon}$

Let us recall the initial data $\mathbf{h}^N(0,\cdot) = \Pi_N \mathbf{h}(0,\cdot)$ converges uniformly to $\mathbf{h}(0,\cdot)$ with probability 1. Thus, to approximate $\mathbf{h}^N(0,\cdot)$ initial data within $\varepsilon$, which we emphasize is deterministic, it suffices to approximate the limit $\mathbf{h}(0,\cdot)$ itself. We will do so in the sequel for which we provide explanation afterwards.

**Definition 2.1.** *Given any deterministic data $\mathbf{h}(0,\cdot)$, define the Fourier smoothing $\mathbf{h}^{N,1,\varepsilon}(0,\cdot) = \Pi_N \mathbf{h}^{1,\varepsilon}(0,\cdot)$, where $\mathbf{h}^{1,\varepsilon}(0,\cdot)$ is the Brownian bridge $\mathbf{b}$ independent of $\xi$ conditioned to be within $\varepsilon$ of $\mathbf{h}(0,\cdot)$ uniformly on $\mathbb{T}$; observe that this construction implicitly chooses the $k = 0$ Fourier coefficient for Brownian bridge so that this event on which we condition has positive probability. Let $\mathbf{P}^\varepsilon$ be the law of $\mathbf{h}^{1,\varepsilon}$ as a measure on $\mathscr{C}(\mathbb{T})$, and let $\mathbf{P}^\infty$ be Brownian bridge measure on $\mathscr{C}(\mathbb{T})$.*

Definition 2.1 can be viewed as constructing $\mathbf{h}^{N,1,\varepsilon}$ to automatically satisfy the first required bullet point in Theorem 1.5 at the level of initial data. We will show in this section that this comparison from Theorem 1.5 propagates globally in time. However, to clarify why we choose Brownian approximations, we appeal to the following result, which shows that the associated probability measure $\mathbf{P}^\varepsilon$ is uniformly "stable" with respect to the law of Brownian bridge itself free of any conditioning. This stability is at the level of relative entropy; see Definition C.1 for its definition and Lemma C.2 for the key properties of relative entropy that we will use to study $\mathbf{h}^{N,1,\varepsilon}$.

**Lemma 2.2.** *There exists a constant $\kappa = \kappa(\varepsilon, \mathbf{h}(0,\cdot))$ depending only on $\varepsilon$ and $\mathbf{h}(0,\cdot)$, in particular independent of $N$, such that the relative entropy of $\mathbf{P}^\varepsilon$ with respect to $\mathbf{P}^\infty$ is bounded by $\kappa$:*

$$\mathsf{H}(\mathbf{P}^\varepsilon | \mathbf{P}^\infty) \;\leqslant\; \kappa. \tag{2.1}$$

*The same is true is we replace $(\mathbf{P}^\varepsilon, \mathbf{P}^\infty)$ by $(\mathscr{T}_* \mathbf{P}^\varepsilon, \mathscr{T}_* \mathbf{P}^\infty)$, where $\mathscr{T}_*$ is the pushforward on probability measures induced by differentiation $\nabla$, Fourier smoothing $\Pi_N$, or compositions of these two. The same is also true if we replace $(\mathbf{P}^\varepsilon, \mathbf{P}^\infty)$ with path-space measures induced by (1.9), when viewed as a finite-dimensional SDE parameterized by the image of $\Pi_N$, with initial measures $(\mathscr{T}_* (\Pi_N)_* \mathbf{P}^\varepsilon, \mathscr{T}_* (\Pi_N)_* \mathbf{P}^\infty)$ for the same choices of $\mathscr{T}_*$.*

*Proof.* It suffices to prove the relative entropy bound for $\mathbf{P}^\varepsilon$ with respect to $\mathbf{P}^\infty$, as the other relative entropy estimates follow for free by Lemma C.2. To this end, observe the function $\mathsf{x} \log \mathsf{x}$ is continuous on $\mathbb{R}_{\geqslant 0}$. Because relative entropy in Definition C.1 is controlled by the maximum of $\mathsf{x} \log \mathsf{x}$ for $\mathsf{x}$ equal to the Radon-Nikodym derivative of $\mathbf{P}^\varepsilon$ with respect to $\mathbf{P}^\infty$, it then suffices to estimate this Radon-Nikodym derivative uniformly in the randomness in the expectation defining relative entropy. Moreover, since $\mathbf{P}^\varepsilon$ is defined as conditioning $\mathbf{P}^\infty$ on an event that we denote by $\mathcal{E}(\varepsilon)$, it suffices to estimate

$\mathcal{E}(\varepsilon)$-probability under $\mathbf{P}^\infty$ from below by a $\varepsilon, \mathbf{h}(0, \cdot)$-dependent constant. To make this precise, we note:

$$\mathsf{H}(\mathbf{P}^\varepsilon | \mathbf{P}^\infty) \;\leqslant\; \|\mathbf{R}^\varepsilon \log \mathbf{R}^\varepsilon\|_{\omega;\infty} \;\leqslant\; \sup_{0 \leqslant \mathrm{x} \leqslant \mathbf{P}^\infty(\mathcal{E}(\varepsilon))^{-1}} |\mathrm{x} \log \mathrm{x}| \;\lesssim_{\mathbf{P}^\infty(\mathcal{E}(\varepsilon))^{-1}}\; 1. \qquad (2.2)$$

In (2.2), $\mathbf{R}^\varepsilon$ is the Radon-Nikodym derivative of $\mathbf{P}^\varepsilon$ with respect to $\mathbf{P}^\infty$ where $\|\|_{\omega;\infty}$ is the $\infty$-norm for functions on $\mathscr{D}(\mathbb{T})$. The last bound in (2.2) follows from the fact that $\mathrm{x} \log \mathrm{x}$ restricted to $[0, C]$ is uniformly controlled by a $C$-dependent constant; to clarify, we note $\mathbf{R}^\varepsilon \leqslant \mathbf{P}^\infty(\mathcal{E}(\varepsilon))^{-1}$ by construction via conditioning in Definition 2.1.

In particular, to control the far LHS of (2.2) in terms of $\varepsilon, \mathbf{h}(0, \cdot)$, it suffices to bound $\mathbf{P}^\infty(\mathcal{E}(\varepsilon))^{-1}$ from above in terms of $\varepsilon, \mathbf{h}(0, \cdot)$, and therefore $\mathbf{P}^\infty(\mathcal{E}(\varepsilon))$ from below in terms of $\varepsilon, \mathbf{h}(0, \cdot)$. To this end, take a deterministic smooth function $\psi_\varepsilon$ that is uniformly within $\varepsilon/2$ of $\mathbf{h}(0, \cdot)$; because $\mathbf{h}(0, \cdot)$ is uniformly continuous, we may take $\psi_\varepsilon$ whose derivatives are controlled by $\varepsilon, \mathbf{h}(0, \cdot)$-dependent constants. As $\mathcal{E}(\varepsilon)$ is the event on which a Brownian bridge $\mathbf{b}$ is uniformly within $\varepsilon$ of $\mathbf{h}(0, \cdot)$, we have containment $\mathcal{E}(\varepsilon) \supseteq \mathcal{E}'(\varepsilon)$ where $\mathcal{E}'(\varepsilon)$ is the event on which $\psi_\varepsilon$ and $\mathbf{b}$ are uniformly within $\varepsilon/2$ of each other, or equivalently $\mathbf{b} - \psi_\varepsilon$ is uniformly bounded by $\varepsilon/2$ in absolute value. As $\psi_\varepsilon$ is smooth with $\varepsilon, \mathbf{h}(0, \cdot)$-dependent derivatives, the Girsanov theorem tells us that the Radon-Nikodym derivative of $\mathbf{b} - \psi_\varepsilon$ with respect to Brownian bridge measure is $\mathscr{L}^2$ as a function of $\mathscr{C}(\mathbb{T})$; here, we think of both $\mathbf{b} - \psi_\varepsilon$ and $\mathbf{b}$ as SDEs parameterized by $\mathrm{x} \in \mathbb{T}$. Thus, if $\widetilde{\mathbf{R}}^\varepsilon$ denotes this last Radon-Nikodym derivative, the Cauchy-Schwarz inequality tells us the following probability estimate in which $\|\|_{\omega;2}$ denotes the $\mathscr{L}^2$-norm for functions on $\mathscr{D}(\mathbb{T})$ with respect to the law of $\mathbf{b} - \psi_\varepsilon$:

$$\mathbf{P}^\infty\left(\sup_{x \in \mathbb{T}} |\mathbf{b}(x)| \leqslant \varepsilon/2\right) \;\leqslant\; \|\widetilde{\mathbf{R}}^\varepsilon\|_{\omega;2}\left(\mathbf{P}^\infty\left(\mathcal{E}'(\varepsilon)\right)\right)^{1/2}. \qquad (2.3)$$

The reflection principle for Brownian bridge implies the LHS of (2.3) is controlled by an $\varepsilon$-dependent constant from below. An upper bound on $\|\widetilde{\mathbf{R}}^\varepsilon\|_{\omega;2}$ therefore bounds $\mathbf{P}^\infty(\mathcal{E}'(\varepsilon))$ from below by $\varepsilon, \mathbf{h}(0, \cdot)$-dependent constants, which completes the proof via (2.2) and the sentence immediately after (2.2). $\qquad\square$

## 2.2 Convergence of $\mathbf{u}^{N,1,\varepsilon}$

In this subsection, we prove that $\mathbf{u}^{N,1,\varepsilon} = \nabla\mathbf{h}^{N,1,\varepsilon}$ converges to the solution of (1.3) with the initial data claimed in Theorem 1.5. Roughly speaking, this is done in [21] if the initial data of $\mathbf{h}^{N,1,\varepsilon}$ is a Brownian bridge independent of $\xi$ and free of any conditioning. This Brownian bridge initial data is ultimately used, however, to show that certain terms vanish and that certain limits exist in the large-$N$ limit. Therefore, the relative entropy estimate in Lemma 2.2 allows us to inherit all these for $\mathbf{h}^{N,1,\varepsilon}$ and $\mathbf{u}^{N,1,\varepsilon}$ initial data in Definition 2.1. We formally state this in the following.

**Lemma 2.3.** *Given any fixed positive $\varepsilon$, the process $\mathbf{u}^{N,1,\varepsilon} = \nabla\mathbf{h}^{N,1,\varepsilon}$ converges weakly in $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$ to the Cole-Hopf solution of (1.3) with initial data given by the weak derivative limit $\lim_{N\to\infty} \nabla\mathbf{h}^{N,1,\varepsilon}(0, \cdot) = \nabla\mathbf{h}^{1,\varepsilon}(0, \cdot)$, which is the weak derivative of Brownian bridge independent of $\xi$ and conditioned to stay uniformly within $\varepsilon$ on $\mathbb{T}$ of $\mathbf{h}(0, \cdot)$ constructed in Definition 2.1.*

*Proof.* Note that if the process $\mathbf{u}^{N,1,\varepsilon} = \nabla\mathbf{h}^{N,1,\varepsilon}$ converges weakly in $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$, its initial data must be given by the limit $\lim_{N\to\infty} \nabla\mathbf{h}^{N,1,\varepsilon}(0, \cdot) = \nabla\mathbf{h}^{1,\varepsilon}(0, \cdot)$. We clarify this limit is taken with respect to weak-$*$ topology on $\mathscr{D}(\mathbb{T})$. It can be computed by first recalling $\lim_{N\to\infty} \mathbf{h}^{N,1,\varepsilon}(0, \cdot) = \lim_{N\to\infty} \nabla\Pi_N \mathbf{h}^{1,\varepsilon}(0, \cdot)$. Because $\nabla$ is continuous with respect to weak-$*$ topology on $\mathscr{D}(\mathbb{T})$, we know $\lim_{N\to\infty} \nabla\Pi_N \mathbf{h}^{N,1,\varepsilon}(0, \cdot) = \nabla\lim_{N\to\infty} \Pi_N \mathbf{h}^{1,\varepsilon}(0, \cdot)$. Because the law of $\mathbf{h}^{1,\varepsilon}$ is absolutely continuous with respect to Brownian bridge $\mathbf{b}$ on

$\mathbb{T}$, and because $\Pi_N \mathbf{b} \to \mathbf{b}$ uniformly on $\mathbb{T}$ with probability 1, we deduce $\Pi_N \mathbf{h}^{1,\varepsilon} \to \mathbf{h}^{1,\varepsilon}$ uniformly on $\mathbb{T}$ with probability 1 as well. Since uniform topology is stronger than weak-$*$ topology, we deduce $\lim_{N\to\infty} \Pi_N \mathbf{h}^{1,\varepsilon}(0,\cdot) = \mathbf{h}^{1,\varepsilon}(0,\cdot)$. Ultimately, we obtain that the initial data of $\mathbf{u}^{\infty,1,\varepsilon}$, if it converges weakly in $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$, is equal to (in law) $\mathbf{h}^{1,\varepsilon}$.

It suffices to prove convergence of $\mathbf{u}^{N,1,\varepsilon} = \nabla \mathbf{h}^{N,1,\varepsilon}$ to the Cole-Hopf solution of (1.3). For this, we establish notation, setting $\langle , \rangle_{\mathbb{T}}$ as the integral pairing for functions on $\mathbb{T}$ and $\langle , \rangle_{t;\mathbb{T}} = \int_0^t \langle , \rangle \mathrm{d}\mathfrak{s}$. To prove convergence, we first will show tightness of $\mathbf{u}^{N,1,\varepsilon}$ and then identify limit points. To this end, we have the proposed tightness for Brownian bridge initial data in [21]. Lemma 2.2 gives $O(1)$-relative entropy for the path-space law of $\mathbf{u}^{N,1,\varepsilon}$ with respect to the path-space law with Brownian bridge initial data. The last paragraph Lemma C.2 then lets us inherit the proposed tightness from that for Brownian bridge initial data. Ultimately, we deduce tightness of $\mathbf{u}^{N,1,\varepsilon}$ in $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$.

To identify limit points, the first step here is to use the Itô formula to obtain the following in which $\mathbf{B}^N(t;\Phi)$ is a Brownian motion of quadratic variation $2t\|\nabla \Pi_N \Phi\|_{x;2}^2$, where $\|\|_{x;2}$ is the $\mathscr{L}^2(\mathbb{T})$-norm:

$$\langle \mathbf{u}^{N,1,\varepsilon}(t,\cdot), \Phi \rangle_{\mathbb{T}} = \langle \mathbf{u}^{N,1,\varepsilon}(0,\cdot), \Phi \rangle_{\mathbb{T}} + \langle \mathbf{u}^{N,1,\varepsilon}, \Delta \Phi \rangle_{t;\mathbb{T}} \qquad (2.4)$$
$$- \varepsilon_N^{-1} \langle \Pi_N \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1,\varepsilon}), \nabla \Phi \rangle_{t;\mathbb{T}} + \mathbf{B}^N(t;\Phi).$$

Observe that each term in (2.4) is a continuous functional of the $\mathscr{D}(\mathbb{T})$-valued process $\mathbf{u}^{N,1,\varepsilon}$. For example, the LHS of (2.4) and the first three terms on the RHS are given by integrating space-time values of $\mathbf{u}^{N,1,\varepsilon}$ against deterministic functions. The last martingale term on the RHS, being determined by the remaining terms, is therefore a functional of $\mathbf{u}^{N,1,\varepsilon}$. Therefore, because $\mathbf{u}^{N,1,\varepsilon}$ converges along subsequences, any continuous functional of $\mathbf{u}^{N,1,\varepsilon}$ must also converge along subsequences. We choose the continuous functional given by the joint process consisting of the terms in (2.4). We deduce by taking large-$N$ limits of (2.4) the following identity:

$$\langle \mathbf{u}^{\infty,1,\varepsilon}(t,\cdot), \Phi \rangle_{\mathbb{T}} = \langle \mathbf{u}^{\infty,1,\varepsilon}(0,\cdot), \Phi \rangle_{\mathbb{T}} + \langle \mathbf{u}^{\infty,1,\varepsilon}, \Delta \Phi \rangle_{t;\mathbb{T}}$$
$$- \lim_{N\to\infty} \varepsilon_N^{-1} \langle \Pi_N \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1,\varepsilon}), \nabla \Phi \rangle_{t;\mathbb{T}} + \lim_{N\to\infty} \mathbf{B}^N(t;\Phi).$$

Recall $\mathbf{B}^N(t;\Phi)$ is a Brownian motion with linear-in-time quadratic variation $2t\|\nabla \Pi_N \Phi\|_{x;2}^2$. Because $\Phi \in \mathscr{C}^\infty(\mathbb{T})$, we know $2t\|\nabla \Pi_N \Phi\|_{x;2}^2 \to 2t\|\nabla \Phi\|_{x;2}^2$ as $N \to \infty$. Therefore, $\mathbf{B}^N(t;\Phi)$ must converge weakly (as a process) to a Brownian motion of quadratic variation $2t\|\nabla \Phi\|_{x;2}^2$. (Technically, one needs uniform integrability of $|\mathbf{B}^N(t;\Phi)|^2$ as $N \to \infty$ to guarantee $\mathbf{B}^N(t;\Phi)$ converges to a martingale with the limit quadratic variation, but this follows from fourth-moment estimates readily available for Brownian motion.) In [21], it is shown that for Brownian bridge initial data $\mathbf{h}^{N,1,\infty}$ (that is independent of $\xi$), the remaining limit above converges to the process

$$\lim_{N\to\infty} \varepsilon_N^{-1} \langle \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1,\infty}), \nabla \Phi \rangle_{t;\mathbb{T}} = 2^{-1} \mathbf{E}^{\mathrm{G}} \mathbf{F}'' \cdot \lim_{N\to\infty} \varepsilon_N^{-1} \langle (\mathbf{u}^{N,1,\infty})^2, \nabla \Phi \rangle_{t;\mathbb{T}} \overset{\bullet}{=} \mathbf{A}(t;\Phi).$$
$$(2.5)$$

Again, since $\varepsilon_N^{-1} \langle \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1,\infty}), \nabla \Phi \rangle_{t;\mathbb{T}}$ and $2^{-1} \mathbf{E}^{\mathrm{G}} \mathbf{F}'' \varepsilon_N^{-1} \langle (\mathbf{u}^{N,1,\infty})^2, \nabla \Phi \rangle_{t;\mathbb{T}}$ are continuous functionals of the $\mathscr{D}(\mathbb{T})$-valued process $\mathbf{u}^{N,1,\varepsilon}$, if they both converge (along subsequences) and their difference converges to zero in probability for Brownian bridge initial data $\mathbf{h}^{N,1,\infty}$, then the same must be true for $\mathbf{h}^{N,1,\varepsilon}$ initial data. This last claim follows from the path-space relative entropy estimate in Lemma 2.2 and (C.2) (these show that asymptotically probability-zero events for Brownian bridge initial data $\mathbf{h}^{N,1,\infty}$ are also asymptotically probability-zero for $\mathbf{h}^{N,1,\varepsilon}$ initial data). Thus, we deduce, for any $\Phi \in \mathscr{C}^\infty(\mathbb{T})$, that

$$\langle \mathbf{u}^{\infty,1,\varepsilon}(t,\cdot), \Phi \rangle_{\mathbb{T}} = \langle \mathbf{u}^{\infty,1,\varepsilon}(0,\cdot), \Phi \rangle_{\mathbb{T}} + \langle \mathbf{u}^{\infty,1,\varepsilon}, \Delta \Phi \rangle_{t;\mathbb{T}} - \mathbf{A}(t;\Phi) + \mathbf{B}(t;\Phi). \qquad (2.6)$$

By Theorem 2.8 in [22], in order to show that $\mathbf{u}^{\infty,1,\varepsilon}$ are Cole-Hopf solutions, it suffices to show that:

- The limit point $\mathbf{u}^{\infty,1,\varepsilon}$ of the $\mathscr{D}(\mathbb{T})$-valued processes $\mathbf{u}^{N,1,\varepsilon}$ satisfies (1.6).
- We have $\mathbf{E}|e^{\langle \mathbf{u}^{\infty,1,\varepsilon}(\mathfrak{t},\cdot),\Theta_{\mathrm{x}}\rangle_{\mathbb{T}}}|^2 < \infty$ for $\mathfrak{t} \geqslant 0$, where $\Theta_{\mathrm{x}}(\mathrm{y}) = \Theta(\mathrm{x} - \mathrm{y})$ and $\Theta$ is defined by its Fourier transform $\int_{\mathbb{T}} \Theta(\mathrm{y})e^{-2\pi i k\mathrm{y}}d\mathrm{y} = \mathbf{1}_{k\neq 0}$.

For these two points, note that Lemma 2.2 actually shows the Radon-Nikodym derivative of the path-space law of $\mathbf{u}^{N,1,\varepsilon}$ with respect to the law of the solution $\mathbf{u}^{N,1,\infty}$ to (1.9) with stationary initial data is $\mathrm{O}_\varepsilon(1)$ in $\mathscr{L}^\infty$, not just in relative entropy. Thus, the same is true by lower semicontinuity of $\mathscr{L}^\infty$-norm for the Radon-Nikodym derivative of the path-space law of $\mathbf{u}^{\infty,1,\varepsilon}$ with respect to the path-space law of $\mathbf{u}^{\infty,1,\infty}$ with white noise initial data. In particular, we have the following two estimates, which reduce showing the previous two bullet points to showing the same but for $\mathbf{u}^{\infty,1,\varepsilon}$ replaced by $\mathbf{u}^{\infty,1,\infty}$:

$$\mathbf{E}|\int_{\mathfrak{s}}^{\mathfrak{t}} (\mathbf{u}^{\infty,1,\varepsilon} * \rho^N)^2(-\nabla\Phi) - (\mathbf{u}^{\infty,1,\varepsilon} * \rho^M)^2(-\nabla\Phi)dr|^2 \tag{2.7}$$

$$\lesssim \mathbf{E}|\int_{\mathfrak{s}}^{\mathfrak{t}} (\mathbf{u}^{\infty,1,\infty} * \rho^N)^2(-\nabla\Phi) - (\mathbf{u}^{\infty,1,\infty} * \rho^M)^2(-\nabla\Phi)dr|^2 \tag{2.8}$$

$$\mathbf{E}|e^{\langle \mathbf{u}^{\infty,1,\varepsilon}(\mathfrak{t},\cdot),\Theta_{\mathrm{x}}\rangle_{\mathbb{T}}}|^2 \tag{2.9}$$

$$\lesssim_\varepsilon \mathbf{E}|e^{\langle \mathbf{u}^{\infty,1,\infty}(\mathfrak{t},\cdot),\Theta_{\mathrm{x}}\rangle_{\mathbb{T}}}|^2. \tag{2.10}$$

The term (2.8) is bounded the RHS of (1.6); this is shown in Lemma 1 of [19]. The term (2.10) is finite by Theorem 2.13 in [20] (see also Proposition 2.1 in [22]). This completes the proof. $\qquad\square$

## 2.3 Estimate for $\mathbf{h}^{N,2,\varepsilon}$

The result of this subsection is confirmation of the second bullet point in Theorem 1.5 for our construction of $\mathbf{h}^{N,2,\varepsilon}$ made in Definition 2.1. The proof follows by the parabolic maximum principle [12].

**Lemma 2.4.** *We first recall that $\mathbf{h}^{N,2,\varepsilon} = \mathbf{h}^N - \mathbf{h}^{N,1,\varepsilon}$ is defined to solve (1.8) with initial data $\mathbf{h}^N(0,\cdot) - \mathbf{h}^{N,1,\varepsilon}(0,\cdot)$. We have $|\mathbf{h}^{N,2,\varepsilon}(\mathfrak{t},\mathrm{x})| \leqslant \varepsilon + \mathrm{o}_N$ for all non-negative $\mathfrak{t}$ and $\mathrm{x} \in \mathbb{T}$, where $\mathrm{o}_N \to 0$ uniformly in $\mathfrak{t},\mathrm{x}$ with probability 1.*

*Proof.* We first prove the inequality for $\mathfrak{t} = 0$. To this end, we have $\mathbf{h}^{N,2,\varepsilon}(0,\mathrm{x}) = \mathbf{h}^N(0,\mathrm{x}) - \mathbf{h}^{N,1,\varepsilon}(0,\mathrm{x})$, so

$$|\mathbf{h}^{N,2,\varepsilon}(0,\mathrm{x})| \leqslant |\mathbf{h}^N(0,\mathrm{x}) - \mathbf{h}(0,\mathrm{x})| + |\mathbf{h}(0,\mathrm{x}) - \mathbf{h}^{1,\varepsilon}(0,\mathrm{x})| + |\mathbf{h}^{1,\varepsilon}(0,\mathrm{x}) - \mathbf{h}^{N,1,\varepsilon}(0,\mathrm{x})|. \tag{2.11}$$

The first term on the far RHS is $\mathrm{o}_N$ by assumption. Moreover, by construction in Definition 2.1, the second term on the far RHS is defined to be at most $\varepsilon$. The last term is $\mathrm{o}_N$, which vanishes uniformly in $\mathbb{T}$ with probability 1 by classical Brownian motion estimates. This gives the proposed estimate for the initial time. It now suffices to get the following for all $\mathfrak{t} \geqslant 0$ and $\mathrm{x} \in \mathbb{T}$ simultaneously with probability 1:

$$|\mathbf{h}^{N,2,\varepsilon}(\mathfrak{t},\mathrm{x})| \leqslant \sup_{\mathrm{y}\in\mathbb{T}}|\mathbf{h}^{N,2,\varepsilon}(0,\mathrm{y})|. \tag{2.12}$$

Indeed, this bound would propagate the validity of the proposed inequality at $\mathfrak{t} = 0$ globally in time with probability 1. To prove this last bound, we first note that $\mathbf{h}^{N,1,\varepsilon}$ is smooth with probability 1 by Lemma A.1. We will only use this fact to guarantee that all derivatives of $\mathbf{h}^{N,1,\varepsilon}$ are defined pointwise with probability 1. Now consider the following.

- Fix $\chi \in \mathscr{C}^{\infty}(\mathbb{R})$ such that its support is contained in $[-1,1]$, it is non-negative-valued, and $\int_{\mathbb{R}} \chi(\mathrm{x}) \mathrm{dx} = 1$. For any $\upsilon > 0$, define $\chi_{\upsilon}(\mathrm{x}) = \upsilon^{-1}\chi(\mathrm{x}/\upsilon)$ for $\mathrm{x} \in \mathbb{R}$.
- Define $\mathbf{F}_{\upsilon} = \mathbf{F} \star \chi_{\upsilon}$ to be a smoothing of $\mathbf{F}$. Define $\mathbf{h}^{N,2,\varepsilon,\upsilon}$ to be the solution to the following PDE on $\mathbb{R}_{\geqslant 0} \times \mathbb{T}$ with the same initial data $\mathbf{h}^{N,2,\varepsilon,\upsilon}(0,\mathrm{x}) = \mathbf{h}^{N,2,\varepsilon}(0,\mathrm{x})$:

$$\partial_{\mathrm{t}}\mathbf{h}^{N,2,\varepsilon,\upsilon} = \Delta\mathbf{h}^{N,2,\varepsilon,\upsilon} + \varepsilon_N^{-1}\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1,\varepsilon} + \mathbf{h}^{N,2,\varepsilon,\upsilon})) - \varepsilon_N^{-1}\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1,\varepsilon}). \tag{2.13}$$

Let us comment briefly on the above construction; this will be key to our proof of Lemma 2.4.

- Note the support of $\chi_{\upsilon}$ is contained in $[-\upsilon, \upsilon]$.
- Because $\mathbf{F} : \mathbb{R} \to \mathbb{R}$ is uniformly Lipschitz and $\chi_{\upsilon}$ is supported in $[-\upsilon, \upsilon]$, we know

$$|\mathbf{F}(\mathrm{x}) - \mathbf{F}_{\upsilon}(\mathrm{x})| = |\int_{\mathbb{R}} \chi_{\upsilon}(\mathrm{x}-\mathrm{y})(\mathbf{F}(\mathrm{x}) - \mathbf{F}(\mathrm{y}))\mathrm{dy}| \lesssim_{\mathbf{F}} \upsilon; \tag{2.14}$$

the dependence on $\mathbf{F}$ is through its Lipschitz constant. The bound above is independent of $\mathrm{x} \in \mathbb{R}$.
- The Lipschitz norm of $\nabla^k \mathbf{F}_{\upsilon}$ is bounded by $\mathrm{O}_{k,\upsilon}(1)$ times that of $\mathbf{F}$ for all $\upsilon > 0$. Indeed, for any $\mathrm{x}_1, \mathrm{x}_2$, we have

$$|\nabla^k \mathbf{F}_{\upsilon}(\mathrm{x}_1) - \nabla^k \mathbf{F}_{\upsilon}(\mathrm{x}_2)| = |\int_{\mathbb{R}} \left(\nabla^k \chi_{\upsilon}(\mathrm{x}_1 - \mathrm{y}) - \nabla^k \chi_{\upsilon}(\mathrm{x}_2 - \mathrm{y})\right)\mathbf{F}(\mathrm{y})\mathrm{dy}| \tag{2.15}$$

$$= |\int_{\mathbb{R}} \nabla^k \chi_{\upsilon}(\mathrm{y}) \left(\mathbf{F}(\mathrm{x}_1 - \mathrm{y}) - \mathbf{F}(\mathrm{x}_2 - \mathrm{y})\right)\mathrm{dy}| \tag{2.16}$$

$$\lesssim_{k,\upsilon} \|\mathbf{F}\|_{\mathrm{Lip}}|\mathrm{x}_1 - \mathrm{x}_2|, \tag{2.17}$$

where $\|\mathbf{F}\|_{\mathrm{Lip}}$ denotes the Lipschitz constant of $\mathbf{F}$.

To prove (2.12), it suffices to prove two ingredients. The first is that $\mathbf{h}^{N,2,\varepsilon,\upsilon} \to_{\upsilon\to 0} \mathbf{h}^{N,2,\varepsilon}$ locally uniformly in time and uniformly in space (with probability 1). The second is that (2.12) holds if we replace $\mathbf{h}^{N,2,\varepsilon}$ on both sides of the inequality by $\mathbf{h}^{N,2,\varepsilon,\upsilon}$. Note that the RHS does not change if we replace $\mathbf{h}^{N,2,\varepsilon}$ therein by $\mathbf{h}^{N,2,\varepsilon,\upsilon}$ as these two have the same initial data by construction. The second ingredient is almost immediate. Again, by Lemma A.2, (2.13) admits a classical solution. Now, let us rewrite (2.13) as

$$\partial_{\mathrm{t}}\mathbf{h}^{N,2,\varepsilon,\upsilon} = \Delta\mathbf{h}^{N,2,\varepsilon,\upsilon} \tag{2.18}$$

$$+ \varepsilon_N^{-1}\left(\frac{\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1,\varepsilon} + \mathbf{h}^{N,2,\varepsilon,\upsilon})) - \varepsilon_N^{-1}\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1,\varepsilon})}{\nabla\mathbf{h}^{N,2,\varepsilon,\upsilon}}\right)\nabla\mathbf{h}^{N,2,\varepsilon,\upsilon}.$$

The coefficient for $\nabla\mathbf{h}^{N,2,\varepsilon,\upsilon}$ on the RHS of (2.18) is uniformly bounded in x (for each $N$):

$$|\varepsilon_N^{-1}\frac{\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1} + \mathbf{h}^{N,2,\varepsilon,\upsilon})) - \varepsilon_N^{-1}\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1})}{\nabla\mathbf{h}^{N,2,\varepsilon,\upsilon}}| \tag{2.19}$$

$$\lesssim_{N,\mathbf{F}} |\frac{\nabla\mathbf{h}^{N,1,\varepsilon} + \nabla\mathbf{h}^{N,2,\varepsilon,\upsilon} - \nabla\mathbf{h}^{N,1,\varepsilon}}{\nabla\mathbf{h}^{N,2,\varepsilon,\upsilon}}| = 1. \tag{2.20}$$

We note the dependence of $\mathbf{F}$ above is through its Lipschitz constant. Thus, to (2.18), we can apply the comparison principle to deduce (2.12) holds for $\mathbf{h}^{N,2,\varepsilon,\upsilon}$ in place of $\mathbf{h}^{N,2,\varepsilon}$. (For the comparison principle, see Theorem 8 of Chapter 7.1.4 in [12] for $n = 1$ and $a^{11} = 1$ and $b^1 = \varepsilon_N^{-1}\frac{\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}+\mathbf{h}^{N,2,\varepsilon,\upsilon}))-\varepsilon_N^{-1}\mathbf{F}_{\upsilon}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1})}{\nabla\mathbf{h}^{N,2,\varepsilon,\upsilon}}$ and $c = 0$. (We note that although Theorem 8 of Chapter 7.1.4 in [12] assumes continuity of coefficients, the proof works for boundedness of the first-order coefficient.)

It now remains to prove the convergence $\mathbf{h}^{N,2,\varepsilon,\upsilon} \to_{\upsilon \to 0} \mathbf{h}^{N,2,\varepsilon}$ locally uniformly in time and uniformly in space (with probability 1). To this end, we start with their mild representations. In particular, starting with the definition of $\mathbf{h}^{N,2,\varepsilon}$ and Lemma A.2, we have the following calculation:

$$\mathbf{h}^{N,2,\varepsilon}(\mathfrak{t},\mathrm{x}) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{t},\mathrm{x}) \tag{2.21}$$

$$= \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.22}$$

$$- \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}_\upsilon(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.23}$$

$$= \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.24}$$

$$- \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.25}$$

$$+ \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.26}$$

$$- \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}_\upsilon(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s}. \tag{2.27}$$

Let us first estimate the contribution of (2.24)-(2.25). Because $\mathbf{F}$ is uniformly Lipschitz, we have

$$| \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.28}$$

$$- \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x})| \tag{2.29}$$

$$\lesssim_{\mathbf{F},N} \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\cdot) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot)| \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.30}$$

$$\leqslant \mathfrak{t} \sup_{0 \leqslant \mathfrak{s} \leqslant \mathfrak{t}} \sup_{\mathrm{y} \in \mathbb{T}} |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\mathrm{y}) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\mathrm{y})|. \tag{2.31}$$

The last estimate follows by contractivity of convolution with the heat kernel on $\mathbb{T}$ (with respect to the $\mathscr{L}^\infty$-norm on $\mathbb{T}$). As for (2.26)-(2.27), because $|\mathbf{F} - \mathbf{F}_\upsilon| \lesssim \upsilon$ uniformly, we have

$$| \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s} \tag{2.32}$$

$$- \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}_\upsilon(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right) (\mathrm{x}) \mathrm{d}\mathfrak{s}| \tag{2.33}$$

$$\lesssim_N \sup_{\mathrm{y} \in \mathbb{R}} |\mathbf{F}(\mathrm{y}) - \mathbf{F}_\upsilon(\mathrm{y})| \times \int_0^{\mathfrak{t}} \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta}(1) \mathrm{d}\mathfrak{s} \tag{2.34}$$

$$\lesssim \mathfrak{t}\upsilon. \tag{2.35}$$

Combining the previous three displays gives the following estimate for any $T \geqslant 0$ fixed:

$$\sup_{0 \leqslant \mathfrak{t} \leqslant T} \sup_{\mathrm{x} \in \mathbb{T}} |\mathbf{h}^{N,2,\varepsilon}(\mathfrak{t},\mathrm{x}) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{t},\mathrm{x})| \tag{2.36}$$

$$\lesssim_{N,\mathbf{F}} T \sup_{0 \leqslant \mathfrak{s} \leqslant \mathfrak{t}} \sup_{\mathrm{y} \in \mathbb{T}} |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\mathrm{y}) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\mathrm{y})| + T\upsilon.$$

Thus, it remains to estimate the double supremum on the RHS of (2.36). To this end, we proceed in very similar fashion, namely estimating gradients of (2.24)-(2.27). First, let

us denote by $\mathbf{G}_{\mathfrak{t}}(\mathrm{x})$ the heat kernel for $\mathrm{e}^{\mathfrak{t}\Delta}$ on the torus. Taking gradients of (2.24)-(2.25) (with respect to $\mathrm{x}$), we have

$$| \int_0^{\mathfrak{t}} \nabla \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\cdot))) - \right)(\mathrm{x})\mathrm{d}\mathfrak{s} \tag{2.37}$$

$$- \int_0^{\mathfrak{t}} \nabla \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right)(\mathrm{x})| \tag{2.38}$$

$$\lesssim_{\mathbf{F},N} \int_0^{\mathfrak{t}} \int_{\mathbb{T}} |\nabla \mathbf{G}_{\mathfrak{t}-\mathfrak{s}}(\mathrm{x}-\mathrm{y})| \cdot |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\mathrm{y}) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\mathrm{y})|\mathrm{dyd}\mathfrak{s} \tag{2.39}$$

$$\lesssim \int_0^{\mathfrak{t}} (1 + |\mathfrak{t}-\mathfrak{s}|^{-\frac{1}{2}}) \sup_{\mathrm{y}\in\mathbb{T}} |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\mathrm{y}) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\mathrm{y})|\mathrm{d}\mathfrak{s}, \tag{2.40}$$

where the last estimate follows by $\int_{\mathbb{T}} |\nabla \mathbf{G}_{\mathfrak{t}-\mathfrak{s}}(\mathrm{x}-\mathrm{y})\mathrm{dy}| \lesssim 1 + |\mathfrak{t}-\mathfrak{s}|^{-1/2}$; see Lemma A.2. Taking gradients of (2.26)-(2.27), we have

$$| \int_0^{\mathfrak{t}} \nabla \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right)(\mathrm{x})\mathrm{d}\mathfrak{s} \tag{2.41}$$

$$- \int_0^{\mathfrak{t}} \nabla \mathrm{e}^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \varepsilon_N^{-1} \mathbf{F}_\upsilon(\varepsilon_N^{1/2} \nabla(\mathbf{h}^{N,1,\varepsilon}(\mathfrak{s},\cdot) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\cdot))) - \right)(\mathrm{x})\mathrm{d}\mathfrak{s}| \tag{2.42}$$

$$\lesssim_N \sup_{\mathrm{y}\in\mathbb{R}} |\mathbf{F}(\mathrm{y}) - \mathbf{F}_\upsilon(\mathrm{y})| \times \int_0^{\mathfrak{t}} \int_{\mathbb{T}} |\nabla \mathbf{G}_{\mathfrak{t}-\mathfrak{s}}(\mathrm{x}-\mathrm{y})|\mathrm{dyd}\mathfrak{s} \tag{2.43}$$

$$\lesssim \sup_{\mathrm{y}\in\mathbb{R}} |\mathbf{F}(\mathrm{y}) - \mathbf{F}_\upsilon(\mathrm{y})| \times \int_0^{\mathfrak{t}} 1 + |\mathfrak{t}-\mathfrak{s}|^{-\frac{1}{2}}\mathrm{d}\mathfrak{s} \tag{2.44}$$

$$\lesssim (\mathfrak{t} + \mathfrak{t}^{\frac{1}{2}})\upsilon. \tag{2.45}$$

Combining the previous two displays, we deduce the following for any $\mathfrak{t} \geqslant 0$:

$$\sup_{\mathrm{x}\in\mathbb{T}} |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{t},\mathrm{x}) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{t},\mathrm{x})| \tag{2.46}$$

$$\lesssim_{N,\mathbf{F}} \int_0^{\mathfrak{t}} (1 + |\mathfrak{t}-\mathfrak{s}|^{-\frac{1}{2}}) \sup_{\mathrm{y}\in\mathbb{T}} |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{s},\mathrm{y}) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{s},\mathrm{y})|\mathrm{d}\mathfrak{s} + (\mathfrak{t} + \mathfrak{t}^{\frac{1}{2}})\upsilon. \tag{2.47}$$

We can now apply the Gronwall inequality to deduce

$$\sup_{\mathrm{x}\in\mathbb{T}} |\nabla \mathbf{h}^{N,2,\varepsilon}(\mathfrak{t},\mathrm{x}) - \nabla \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{t},\mathrm{x})| \lesssim (\mathfrak{t} + \mathfrak{t}^{\frac{1}{2}})\upsilon \times \exp\left( \int_0^{\mathfrak{t}} 1 + |\mathfrak{t}-\mathfrak{s}|^{-\frac{1}{2}}\mathrm{d}\mathfrak{s} \right) \tag{2.48}$$

$$\lesssim (\mathfrak{t} + \mathfrak{t}^{\frac{1}{2}})\exp(\mathfrak{t} + \mathfrak{t}^{\frac{1}{2}})\upsilon. \tag{2.49}$$

Combining this last display with (2.36), we obtain

$$\sup_{0 \leqslant \mathfrak{t} \leqslant T} \sup_{\mathrm{x}\in\mathbb{T}} |\mathbf{h}^{N,2,\varepsilon}(\mathfrak{t},\mathrm{x}) - \mathbf{h}^{N,2,\varepsilon,\upsilon}(\mathfrak{t},\mathrm{x})| \lesssim_{N,\mathbf{F}} T(T^{\frac{1}{2}} + T)\mathrm{e}^{T^{1/2}+T}\upsilon + T\upsilon, \tag{2.50}$$

which shows, for each fixed $N > 0$, the convergence $\mathbf{h}^{N,2,\varepsilon,\upsilon} \to_{\upsilon\to0} \mathbf{h}^{N,2,\varepsilon}$ locally uniformly in time and uniformly in space (with probability 1). This completes the proof. $\square$

## 2.4 Proof of Theorem 1.5

According to Lemma 2.3 and Lemma 2.4, the two bullet points in Theorem 1.5 are satisfied for our particular choice of $\mathbf{h}^{N,1,\varepsilon}$ and $\mathbf{h}^{N,2,\varepsilon}$ given by choosing their initial data via Definition 2.1 and requiring them to solve (1.7) and (1.8), respectively. To compute the large-$N$ limit of $\mathbf{u}^N$, similar to the proof of Lemma 2.3, it suffices to show that the

one-dimensional process $\mathbf{u}^N(\mathfrak{t}, \Phi) \overset{\bullet}{=} \langle \mathbf{u}^N(\mathfrak{t}, \cdot), \Phi \rangle_{\mathbb{T}}$ converges to the action of the solution to (1.3) on $\Phi$ provided any generic smooth function $\Phi \in \mathscr{C}^\infty(\mathbb{T})$; this convergence must be in $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathbb{R})$ with the locally uniform topology, and it will be directly inherited from convergence of $\mathbf{u}^{N,1,\varepsilon}$ and estimates for $\mathbf{u}^{N,2,\varepsilon}$ from Lemma 2.4. Let us clarify that convergence in locally uniform topology on $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathbb{R})$ means convergence uniformly in $\mathscr{C}([0, \mathfrak{t}_+], \mathbb{R})$ for every $\mathfrak{t}_+ \geqslant 0$. By integration-by-parts, we have

$$\mathbf{u}^N(\mathfrak{t}, \Phi) = \langle \mathbf{u}^{N,1,\varepsilon}(\mathfrak{t}, \cdot), \Phi \rangle_{\mathbb{T}} - \langle \mathbf{h}^{N,2,\varepsilon}(\mathfrak{t}, \cdot), \nabla \Phi \rangle_{\mathbb{T}}. \tag{2.51}$$

The second term on the RHS of (2.51) is bounded by order $\varepsilon$ times $\Phi$-data; this follows by the height function estimate in Lemma 2.4. Because this comparison holds uniformly on finite time-horizons, for any $\mathfrak{t}_+ \geqslant 0$, uniformly in $\mathfrak{t} \in [0, \mathfrak{t}_+]$, we have the following in which $\mathbf{u}(\mathfrak{t}, \Phi)$ denotes the action of the Cole-Hopf solution to (1.3) on $\Phi$:

$$|\mathbf{u}^N(\mathfrak{t}, \Phi) - \mathbf{u}(\mathfrak{t}, \Phi)| \leqslant |\langle \mathbf{u}^{N,1,\varepsilon}(\mathfrak{t}, \cdot), \Phi \rangle_{\mathbb{T}} - \mathbf{u}(\mathfrak{t}, \Phi)| + \kappa(\Phi)\varepsilon + \kappa(\Phi)\mathrm{o}_N. \tag{2.52}$$

Above, the constant $\kappa(\Phi)$ depends only on its argument $\Phi$. Observe the LHS of (2.52) does not depend on $\varepsilon$. Therefore, if we can show the first term on the RHS is bounded by $\kappa(\Phi)\varepsilon + \kappa(\Phi)\mathrm{o}_N$, then because $\varepsilon$ on the RHS of (2.52) is arbitrary, we would be able to deduce $\mathbf{u}^N(\mathfrak{t}, \Phi) \to \mathbf{u}(\mathfrak{t}, \Phi)$ uniformly on $[0, \mathfrak{t}_+]$ in probability provided any finite $\mathfrak{t}_+ \geqslant 0$ and $\Phi$. As noted prior to (2.51), this would complete the proof of convergence. We first note that by Lemma 2.2, it suffices to replace $\mathbf{u}^{N,1,\varepsilon}$ in (2.52) with the Cole-Hopf solution $\mathbf{u}^{1,\varepsilon}$ to (1.3) with initial data the weak derivative of Brownian bridge independent of $\xi$ and conditioned to be within $\varepsilon$ of $\mathbf{h}(0, \cdot)$. It then suffices to employ the following estimate that compares solutions to (1.3) with initial data whose antiderivatives are very close uniformly on $\mathbb{T}$; precisely, we choose $\alpha = \varepsilon$ for the result below, and we choose $\mathbf{h}^1(0, \cdot) = \mathbf{h}^{1,\varepsilon}(0, \cdot)$ from Definition 2.1 and $\mathbf{h}^2(0, \cdot) = \mathbf{h}(0, \cdot)$:

**Lemma 2.5.** *Suppose $\mathbf{u}^1$ and $\mathbf{u}^2$ are Cole-Hopf solutions to (1.3) with initial data $\mathbf{u}^1 = \nabla \mathbf{h}^1(0, \cdot)$ and $\mathbf{u}^2 = \nabla \mathbf{h}^2(0, \cdot)$, in which $\mathbf{h}^1(0, \cdot)$ and $\mathbf{h}^2(0, \cdot)$ are continuous on $\mathbb{T}$ and satisfy the following uniformly in $\mathrm{x} \in \mathbb{T}$ with probability 1:*

$$|\mathbf{h}^1(0, \mathrm{x}) - \mathbf{h}^2(0, \mathrm{x})| \leqslant \alpha. \tag{2.53}$$

*Provided any $\Phi \in \mathscr{C}^\infty(\mathbb{T})$, we have the following uniformly in $\mathfrak{t} \geqslant 0$ with probability 1, where $\kappa(\Phi)$ depends only on $\Phi$:*

$$|\langle \mathbf{u}^1(\mathfrak{t}, \cdot), \Phi \rangle_{\mathbb{T}} - \langle \mathbf{u}^2(\mathfrak{t}, \cdot), \Phi \rangle_{\mathbb{T}}| \leqslant \kappa(\Phi)\alpha. \tag{2.54}$$

*Proof.* We first rewrite the LHS of (2.54) by realizing $\mathbf{u}^i = \nabla \mathbf{h}^i$, where $\mathbf{h}^i$ is the Cole-Hopf solution to (1.1). We then move the gradient onto $\Phi$ and write $\mathbf{h}^i = \beta^{-1} \log \mathbf{Z}^i$, where $\mathbf{Z}^i$ solves the *SHE* $\partial_{\mathfrak{t}} \mathbf{Z}^i = \Delta \mathbf{Z}^i + \beta \mathbf{Z}^i \xi$. This gives

$$\langle \mathbf{u}^1(\mathfrak{t}, \cdot), \Phi \rangle_{\mathbb{T}} - \langle \mathbf{u}^2(\mathfrak{t}, \cdot), \Phi \rangle_{\mathbb{T}} = -\beta^{-1} \left( \langle \log \mathbf{Z}^1(\mathfrak{t}, \cdot), \nabla \Phi \rangle_{\mathbb{T}} - \langle \log \mathbf{Z}^2(\mathfrak{t}, \cdot), \nabla \Phi \rangle_{\mathbb{T}} \right). \tag{2.55}$$

Because $\mathbf{h}^i$ are continuous and satisfy an a priori global estimate for the initial data, it suffices to compare the solutions SHE with globally close initial data. This follows by the comparison principle for the SHE equations with initial data $\exp(\beta(\mathbf{h}^2(0, \mathrm{x}) \pm \alpha))$ and $\exp(\beta \mathbf{h}^1(0, \mathrm{x}))$. In particular, we know $\exp(\beta \mathbf{h}^2(0, \mathrm{x}) - \beta\alpha) \leqslant \exp(\beta \mathbf{h}^1(0, \mathrm{x})) \leqslant \exp(\beta \mathbf{h}^2(0, \mathrm{x}) + \beta\alpha)$ given our initial data assumption since $\beta$ is positive. If we run the SHE with these three initial data, the comparison principle [32] for the SHE implies this two-sided bound remains true not just for time 0 but for any positive time $\mathfrak{t} \geqslant 0$ because $\exp(\pm\beta\alpha)$ are constants and SHE is linear. Because exponential and logarithm functions preserve order, we deduce $\mathbf{h}^2(\mathfrak{t}, \mathrm{x}) - \alpha \leqslant \mathbf{h}^1(\mathfrak{t}, \mathrm{x}) \leqslant \mathbf{h}^2(\mathfrak{t}, \mathrm{x}) + \alpha$ for all $\mathfrak{t} \geqslant 0$ and $\mathrm{x} \in \mathbb{T}$. Therefore, the RHS of (2.55) is controlled by the supremum of $|\nabla \Phi|$ times $\alpha|\mathbb{T}| = \alpha$ as $\mathbb{T}$ is unit-length, so we are done. $\qquad \square$

# 3 Proof of Theorem 1.6

For the sake of clarity, we will establish all the notation to be used in this section in the following list. In particular, we will not inherit any notation used in the previous sections unless explicitly mentioned in order to reset notation and avoid any possible confusion in presentation.

- Consider a deterministic continuous function $\mathbf{h}(0, \cdot)$ thought of as initial data in Theorem 1.6. We define $\mathbf{h}(\mathbf{t}, \mathbf{x})$ as the Cole-Hopf solution to (1.1) with initial data $\mathbf{h}(0, \cdot)$ and $\mathbf{u}(\mathbf{t}, \cdot) = \nabla \mathbf{h}(\mathbf{t}, \cdot)$ the Cole-Hopf solution to (1.3).

- Provided any positive $\varepsilon$, we define $\mathbf{h}^\varepsilon(0, \cdot)$ to be Brownian bridge independent of $\xi$ and conditioned to be within $\varepsilon$ of $\mathbf{h}(0, \cdot)$ uniformly on $\mathbb{T}$. We define $\mathbf{h}^\varepsilon(\mathbf{t}, \mathbf{x})$ as the Cole-Hopf solution to (1.1) with initial data $\mathbf{h}^\varepsilon(0, \cdot)$ and $\mathbf{u}^\varepsilon(\mathbf{t}, \cdot) = \nabla \mathbf{h}^\varepsilon(\mathbf{t}, \cdot)$ the Cole-Hopf solution to (1.3). The noise in this and the previous bullet points are the same noise in a pathwise sense.

- Lastly, we define $\widetilde{\mathbf{h}}^\varepsilon$ to be difference $\mathbf{h} - \mathbf{h}^\varepsilon$. This is a non-regularized analog of $\mathbf{h}^{N,2,\varepsilon}$ from Theorem 1.5.

Theorem 1.6 will follow from the list of estimates provided below, as we illustrate after we present these estimates. Let us first note that we will use some notation not-yet-established, so as to match it soon with notation in Theorem 1.6.

**Lemma 3.1.** *We have $|\widetilde{\mathbf{h}}^\varepsilon(\mathbf{t}, \mathbf{x})| \leqslant \varepsilon$ for all $\mathbf{t} \geqslant 0$ and $\mathbf{x} \in \mathbb{T}$ with probability 1.*

**Lemma 3.2.** *Let $\mathbf{Q}^\varepsilon$ denote the law of $\mathbf{h}^\varepsilon$ as a probability measure on $\mathscr{C}(\mathbb{T})$. Recalling $\mathbf{P}^\infty$ as the law of Brownian bridge on $\mathscr{C}(\mathbb{T})$, we have the relative entropy estimate $\mathsf{H}(\mathbf{Q}^\varepsilon | \mathbf{P}^\infty) \leqslant \kappa(\varepsilon, \mathbf{h})$ for all $\varepsilon$, where $\kappa(\varepsilon, \mathbf{h})$ depends only on $\varepsilon$ and the modulus of continuity of the initial data $\mathbf{h}(0, \cdot)$. Therefore, by Lemma C.2, the same relative entropy estimate holds if we replace $(\mathbf{Q}^\varepsilon, \mathbf{P}^\infty)$ by their respective pushforward under any common map.*

**Lemma 3.3.** *We first let $\nabla \mathbf{Q}^\varepsilon(\mathbf{t})$ denote the law of $\mathbf{u}^\varepsilon(\mathbf{t}, \cdot) = \nabla \mathbf{h}^\varepsilon(\mathbf{t}, \cdot)$ as a probability measure on $\mathscr{D}(\mathbb{T})$, assuming the initial measure is given by $\nabla \mathbf{Q}^\varepsilon$, the pushforward of $\mathbf{Q}^\varepsilon$ from Lemma 3.2 under the gradient map $\mathscr{C}(\mathbb{T}) \to \mathscr{D}(\mathbb{T})$. Recalling from Theorem 1.6 that $\eta$ denotes the Gaussian white noise measure on $\mathscr{D}(\mathbb{T})$, then for any $\mathbf{t} \geqslant 0$, we have $\mathsf{H}(\nabla \mathbf{Q}^\varepsilon(\mathbf{t}) | \eta) \leqslant \exp(-C\mathbf{t}) \mathsf{H}(\nabla \mathbf{Q}^\varepsilon | \eta)$, in which $C$ is a positive constant that depends on nothing.*

Let us now prove Theorem 1.6. We will choose $\mathbf{L}^\varepsilon(\mathbf{t})$ therein to be $\nabla \mathbf{Q}^\varepsilon(\mathbf{t})$ constructed in Lemma 3.3. By Lemma 3.1, the Wasserstein estimate needed in Theorem 1.6 follows immediately; the coupling between $\mathbf{L}(\mathbf{t})$ and $\mathbf{L}^\varepsilon(\mathbf{t}) = \nabla \mathbf{Q}^\varepsilon(\mathbf{t})$ used in the Wasserstein metric is the coupling of $\mathbf{u}(\mathbf{t}, \cdot)$ and $\mathbf{u}^\varepsilon(\mathbf{t}, \cdot)$ used in the bullet points prior to Lemma 3.1, namely coupling the noises that define $\mathbf{u}(\mathbf{t}, \cdot)$ and $\mathbf{u}^\varepsilon(\mathbf{t}, \cdot)$. Indeed, Lemma 3.1 implies that for this coupling, $\mathbf{u}(\mathbf{t}, \cdot)$ and $\mathbf{u}^\varepsilon(\mathbf{t}, \cdot)$ are within $\varepsilon$ of each other with respect to the metric used in the Wasserstein distance; below, we adopt the integration notation introduced and used in the proof of Lemma 2.3:

$$
\begin{aligned}
|\langle \mathbf{u}(\mathbf{t}, \cdot), \Phi \rangle_{\mathbb{T}} - \langle \mathbf{u}^\varepsilon(\mathbf{t}, \cdot), \Phi \rangle_{\mathbb{T}}| &= |\langle \mathbf{h}(\mathbf{t}, \cdot), \nabla \Phi \rangle_{\mathbb{T}} - \langle \mathbf{h}^\varepsilon(\mathbf{t}, \cdot), \nabla \Phi \rangle_{\mathbb{T}}| \\
&\leqslant \langle |\widetilde{\mathbf{h}}^\varepsilon(\mathbf{t}, \cdot)|, |\nabla \Phi| \rangle_{\mathbb{T}} \leqslant \varepsilon \sup_{\mathbf{x} \in \mathbb{T}} |\nabla \Phi(\mathbf{x})|.
\end{aligned}
$$

The relative entropy estimate in Theorem 1.6 follows from combining Lemma 3.2 and Lemma 3.3 with the observation that $\eta = \nabla \mathbf{P}^\infty$ as probability measures, where $\nabla \mathbf{P}^\infty$ denotes the pushforward of the Brownian bridge measure $\mathbf{P}^\infty$ on $\mathscr{C}(\mathbb{T}) \subseteq \mathscr{D}(\mathbb{T})$ under the gradient map. This completes the proof. □

*Proof of* Lemma 3.1. This follows from the proof of Lemma 2.5 for $\mathbf{h}^1 = \mathbf{h}(0, \cdot)$ and $\mathbf{h}^2 = \mathbf{h}^\varepsilon(0, \cdot)$ and $\alpha = \varepsilon$. □

*Proof of* Lemma 3.2. As in the proof of Lemma 2.2, it amounts to estimate from below the probability that Brownian bridge is within $\varepsilon$ of a deterministic continuous function on $\mathbb{T}$; said estimate can only depend on $\varepsilon$ and properties of said deterministic continuous function. But this was estimated in the proof of Lemma 2.2, so we are done. □

*Proof of* Lemma 3.3. Define $\mathbf{u}^{N,\varepsilon,1}$ as the solution to (1.9) for $\mathbf{F}(\mathrm{x}) = \mathrm{x}^2$ with initial data given by $\Pi_N \mathbf{u}^\varepsilon$. We then have $\lim_{N\to\infty} \mathbf{u}^{N,\varepsilon,1} = \mathbf{u}^\varepsilon$ as measures on $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{D}(\mathbb{T}))$ because the relative entropy of $\mathbf{u}^\varepsilon$ with respect to white noise $\eta$ measure is uniformly bounded in $N$ by Lemma 3.2; this convergence follows as in the proof of Lemma 2.3. Because the relative entropy is lower semi-continuous in both of its arguments jointly (see Appendix 1.8 in [31]), we then deduce the following relative entropy estimate in which $\nabla \mathbf{Q}^{\varepsilon,N}(\mathfrak{t})$ is the law of $\mathbf{u}^{N,\varepsilon,1}(\mathfrak{t}, \cdot)$ and $(\Pi_N)_* \eta$ is the Fourier smoothing of white noise measure $\eta$:

$$\mathsf{H}\left(\nabla \mathbf{Q}^\varepsilon(\mathfrak{t})|\eta\right) \ \leqslant \ \limsup_{N\to\infty} \mathsf{H}\left(\nabla \mathbf{Q}^{\varepsilon,N}(\mathfrak{t})|(\Pi_N)_* \eta\right). \tag{3.1}$$

We are left with estimating the RHS of (3.1). To this end, we differentiate the RHS in $\mathfrak{t}$. As $(\Pi_N)_* \eta$ is an invariant Gaussian measure for (1.9) when viewed as a multi-dimensional SDE (see Section 2 of [21]), we get the classical entropy production inequality below (see [24] or Appendix 1.9 in [31], which extends to the continuous state-space case easily); the RHS below is non-positive, namely the negative of a Dirichlet form, as Markov generators are non-positive definite:

$$\partial_{\mathfrak{t}} \mathsf{H}\left(\nabla \mathbf{Q}^{\varepsilon,N}(\mathfrak{t})|(\Pi_N)_* \eta\right) \ \lesssim \ \mathbf{E}^{(\Pi_N)_* \eta} \mathbf{R}^{\varepsilon,N}(\mathfrak{t}) \mathscr{L}_N \mathbf{R}^{\varepsilon,N}(\mathfrak{t}), \tag{3.2}$$

where $\mathbf{R}^{\varepsilon,N}(\mathfrak{t})$ is the square root of the Radon-Nikodym derivative of $\nabla \mathbf{Q}^{\varepsilon,N}(\mathfrak{t})$ with respect to $(\Pi_N)_* \eta$ and $\mathscr{L}_N$ is the generator for the SDEs (1.9), again for $\mathbf{F}(\mathrm{x}) = \mathrm{x}^2$. Let us now observe:

- As noted in [19, 21, 23], the symmetric part of the generator $\mathscr{L}_N$, with respect to $(\Pi_N)_* \eta$, can be written as $\sum_{|k| \leqslant N} \mathbf{1}_{k \neq 0} \mathscr{L}_{\mathrm{OU}}^{(k)}$, where $\mathscr{L}_{\mathrm{OU}}^{(k)}$ is the generator of a one-dimensional Ornstein-Uhlenbeck process of speed $\gtrsim 1$. Indeed, (1.9) without the asymmetry $\mathbf{F}$ diagonalizes via Fourier transform into independent Ornstein-Uhlenbeck processes as in [19, 21, 23].

- The measure $(\Pi_N)_* \eta$ is a product measure on Fourier coordinates $k \in \mathbb{Z}$ such that $|k| \leqslant N$ and $k \neq 0$, since $\eta$ is a mean-zero space-time white noise on $\mathbb{T}$. Moreover, its marginal on the $k$-th Fourier coordinate is the invariant measure for $\mathscr{L}_{\mathrm{OU}}^{(k)}$; see [19, 21, 23].

- Each $\mathscr{L}_{\mathrm{OU}}^{(k)}$ satisfies a log-Sobolev inequality with respect to its invariant measure, the $k$-th marginal of the product measure $(\Pi_N)_* \eta$, with constant $\mathrm{O}(1)$. That is, for any probability density $f$ with respect to the $k$-th marginal of $(\Pi_N)_* \eta$, we have

$$\mathbf{E}^{(\Pi_N)_* \eta} f \log f \ \lesssim \ -\mathbf{E}^{(\Pi_N)_* \eta} \sqrt{f} \mathscr{L}_{\mathrm{OU}}^{(k)} \sqrt{f}. \tag{3.3}$$

Thus, the sum over all $\{|k| \leqslant N, k \neq 0\}$, which gives the symmetric part of $\mathscr{L}_N$, satisfies a log-Sobolev inequality with constant $\mathrm{O}(1)$ with respect to the product measure $(\Pi_N)_* \eta$ (so-called "tensorization" of log-Sobolev inequalities). For this, see Corollary 4.2 in [18].

This resulting LSI from the above bullet points gives us the following from (3.2), where $C$ is some positive universal constant that depends on nothing except the constant 1 in

front of the Laplacian in (1.9) and would otherwise scale with the constants in front of the Laplacian and noise therein:

$$\partial_t \mathsf{H}\left(\nabla \mathbf{Q}^{\varepsilon,N}(\mathsf{t})|(\Pi_N)_*\eta\right) \;\leqslant\; -C\mathsf{H}\left(\nabla \mathbf{Q}^{\varepsilon,N}(\mathsf{t})|(\Pi_N)_*\eta\right). \tag{3.4}$$

From the previous differential inequality, we straightforwardly deduce

$$\mathsf{H}\left(\nabla \mathbf{Q}^{\varepsilon,N}(\mathsf{t})|(\Pi_N)_*\eta\right) \;\leqslant\; \exp(-C\mathsf{t})\cdot \mathsf{H}\left(\nabla \mathbf{Q}^{\varepsilon,N}(0)|(\Pi_N)_*\eta\right). \tag{3.5}$$

The relative entropy factor on the RHS of (3.5) is bounded by itself without the $N$-superscripts, again because the $N$-superscript means pushforward under $\Pi_N$ projections, and relative entropy is contractive with respect to pushforwards; see Lemma C.2. Combining this observation with (3.1) completes the proof. $\qquad\square$

## 4  Fractional stochastic Burgers equation

Instead of (1.3), one can study the following stochastic PDE, in which we take $\alpha \in (1/2, 1]$:

$$\partial_t \mathbf{u}_\alpha \;=\; -(-\Delta)^\alpha \mathbf{u}_\alpha + \beta \nabla(\mathbf{u}_\alpha^2) + (-\Delta)^{\alpha/2}\xi. \tag{4.1}$$

Equations of this type were studied in [19, 23] for stationary initial data, but for $\alpha > 3/4$ well-posedness and fairly robust analytic properties of (4.1) should be derivable via regularity structures [26] or paracontrolled distributions [14] ( $\alpha = 3/4$ is "critical" for these theories). We note, however, that the anti-derivative of (4.1) satisfies the same type of comparison principle that (1.2) does that we have used throughout this paper. Moreover, for $\alpha > 3/4$, equations of the following type should converge to solutions of (4.1), in the same sense as in Theorem 1.5 and with the same assumptions on the nonlinearity $\mathbf{F}$ in Assumption 1.3 and the initial data in Assumption 1.4; in the following, $\xi^N$ is a smoothing of the noise, for example via Fourier smoothing $\Pi_N$, and like in Definition 1.1, we may include $\Pi_N$ in front of the nonlinearity $\mathbf{F}$ below as well:

$$\partial_t \mathbf{u}_\alpha^N \;=\; -(-\Delta)^\alpha \mathbf{u}_\alpha^N + \varepsilon_N^{-1} \nabla \mathbf{F}(\varepsilon_N^{1/2}\mathbf{u}_\alpha^N) + (-\Delta)^{\alpha/2}\xi^N. \tag{4.2}$$

This last convergence claim can be checked for stationary white noise initial data by following the arguments of [21]; only the assumption $\alpha > 1/2$ is needed for the analysis therein, but $\alpha > 3/4$ is needed for uniqueness of limit points that was established in [23]. For the reader's convenience, let us note that the only estimate that needs be checked in [21] is the following, used in the proof of the Boltzmann-Gibbs principle (Proposition 11 of [21]):

$$\sum_{1 \leqslant |k_1|, |k_2| \leqslant N} (|k_1|^\alpha + |k_2|^\alpha)^{-2} \ll N \quad \text{where} \quad \alpha > 1/2. \tag{4.3}$$

Since we have a maximum principle for equations of the type (4.2) (the fractional Laplacian in (4.2) is the generator for a Markov process), we may use our methods to prove a fractional version of Theorem 1.5 if $\alpha > 3/4$, and if $\alpha > 1/2$ more generally if we settle for tightness, not convergence and identification of limit points. This depends on the invariant measure of (4.2) models to be Gaussian white noise. In the same spirit, we can also prove convergence to white noise invariant measure for arbitrary continuous initial data similar to Theorem 1.6 but for the fractional stochastic Burgers equation (4.1). Indeed, if we take an approximation to (4.1) by Fourier smoothing in the same way as we did for (1.3) in the proof of Theorem 1.6, the symmetric part of the resulting finite-dimensional SDE is again given by independent Ornstein-Uhlenbeck processes with uniformly bounded from below speed. We conclude by emphasizing our

methods provide a *first*, intrinsic or extrinsic, notion of solutions to (4.1) with general continuous initial data; again, regularity structures [26] and paracontrolled distributions [14] and energy solutions [19, 23] address Hölder regularity initial data, and for $\alpha \neq 1$ the Cole-Hopf transform does not apply to make sense of (4.1). On the other hand, our method depends crucially on regularizing the fractional stochastic Burgers equation in a Markovian fashion with an explicit invariant measure, unlike regularity structures or paracontrolled distributions.

## A PDE estimates

**Lemma A.1.** *For any $N > 0$ and any initial data $\mathbf{u}^{N,1}(0, \cdot) \in \Pi_N \mathscr{L}^2(\mathbb{T})$, with probability 1 there exists a unique process $\mathbf{u}^{N,1} \in \mathscr{C}(\mathbb{R}_{\geqslant 0}, \Pi_N \mathscr{L}^2(\mathbb{T}))$ such that*

$$
\begin{aligned}
\mathbf{u}^{N,1}(\mathfrak{t}, \mathrm{x}) \; = \; & \mathbf{u}^{N,1}(0, \mathrm{x}) + \int_0^{\mathfrak{t}} \Delta \mathbf{u}^{N,1}(\mathfrak{s}, \mathrm{x}) \mathrm{d}\mathfrak{s} + \int_0^{\mathfrak{t}} \varepsilon_N^{-1} \Pi_N \nabla \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1}(\mathfrak{s}, \mathrm{x})) \mathrm{d}\mathfrak{s} \quad \text{(A.1)} \\
& + \; \nabla \Pi_N \mathbf{B}^N(\mathfrak{t}, \mathrm{x}).
\end{aligned}
$$

*where for each $\mathrm{x} \in \mathbb{T}$, the process $\mathbf{B}^N(\mathfrak{t}, \mathrm{x})$ is a continuous martingale with quadratic variation $2\mathfrak{t}$. In particular, $\mathbf{u}^{N,1}(\mathfrak{t}, \mathrm{x})$ is smooth in $\mathrm{x}$ for all $\mathfrak{t} \geqslant 0$ with probability 1.*

*Proof.* This follows from the argument at the beginning of Section 4 of [19], but $\mathbf{F}$ is not necessarily quadratic as it was in [19]. However, all we need from $\mathbf{F}$ is that $\int_{\mathbb{T}} \mathbf{u}^{N,1}(\mathfrak{t}, \mathrm{x}) \nabla \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1}(\mathfrak{t}, \mathrm{x})) \mathrm{dx} = 0$ in order to use said argument from [19]. This relation is proved in Lemma 5 of [21]. (Technically, Lemma 5 of [21] shows that we can move $\nabla \mathbf{F}(\varepsilon_N^{1/2} \cdot)$ onto the other copy of $\mathbf{u}^{N,1}$ if we add a sign to the $\mathbb{T}$-integral, so that $\nabla \mathbf{F}(\varepsilon_N^{1/2} \cdot)$ is "asymmetric" with respect to Lebesgue measure on $\mathbb{T}$. This certainly implies $\int_{\mathbb{T}} \mathbf{u}^{N,1}(\mathfrak{t}, \mathrm{x}) \nabla \mathbf{F}(\varepsilon_N^{1/2} \mathbf{u}^{N,1}(\mathfrak{t}, \mathrm{x})) \mathrm{dx} = 0$.) This completes the proof. $\quad\square$

**Lemma A.2.** *Suppose $\partial_{\mathfrak{t}} \mathbf{G}_{\mathfrak{t}}(\mathrm{x}) = \Delta \mathbf{G}_{\mathfrak{t}}(\mathrm{x})$ for $\mathfrak{t} > 0$ and $\mathrm{x} \in \mathbb{T}$, and suppose $\mathbf{G}_0(\mathrm{x}) = \delta_{\mathrm{x}}$ as measures. Then for any $\Phi \in \mathscr{L}^\infty(\mathbb{T})$ and for any $\mathfrak{t} \geqslant 0$, we have*

$$
|(\mathrm{e}^{\mathfrak{t}\Delta} \Phi(\cdot))(\mathrm{x})| \; = \; \left| \int_{\mathbb{T}} \mathbf{G}_{\mathfrak{t}}(\mathrm{x} - \mathrm{y}) \Phi(\mathrm{y}) \mathrm{dy} \right| \; \leqslant \; \sup_{\mathrm{y} \in \mathbb{T}} |\Phi(\mathrm{y})| \quad \text{(A.2)}
$$

$$
\int_{\mathbb{T}} |\nabla \mathbf{G}_{\mathfrak{t}}(\mathrm{x})| \mathrm{dx} \; \lesssim \; 1 + \mathfrak{t}^{-\frac{1}{2}}. \quad \text{(A.3)}
$$

*Proof.* Let $\mathbf{H}$ solve $\partial_{\mathfrak{t}} \mathbf{H}_{\mathfrak{t}}(\mathrm{x}) = \Delta \mathbf{H}_{\mathfrak{t}}(\mathrm{x})$ and $\mathbf{H}_0(\mathrm{x}) = \delta_{\mathrm{x}}$ on $\mathbb{R}_{\geqslant 0} \times \mathbb{R}$. By the method of images, we have the following representation of $\mathbf{G}$ in terms of the full-line heat kernel $\mathbf{H}$:

$$
\mathbf{G}_{\mathfrak{t}}(\mathrm{x}) \; = \; \sum_{k \in \mathbb{Z}} \mathbf{H}_{\mathfrak{t}}(\mathrm{x} + k). \quad \text{(A.4)}
$$

Above, we identify $\mathbb{T} \simeq \mathbb{R}/\mathbb{Z} \simeq [0, 1)$. Indeed, the RHS is 1-periodic. It converges because $\mathbf{H}$ is Gaussian and decays sub-exponentially in space. Moreover, it is a sum of terms that vanish under $\partial_{\mathfrak{t}} - \Delta$. Lastly, for $\mathrm{x} \in \mathbb{T} \simeq [0, 1)$, we have $\mathrm{x} + k \notin \mathbb{T}$ if $k \neq 0$, so every $k \neq 0$-term on the RHS vanishes at $\mathfrak{t} = 0$ and we recover just $\mathbf{H}_0(\mathrm{x}) = \delta_{\mathrm{x}}$. By uniqueness of solutions to the heat equation on $\mathbb{T}$ (the proof is the same as in the full line case), we

prove the above relation holds. Using this relation, let us first estimate

$$|\int_{\mathbb{T}} \mathbf{G}_{\mathrm{t}}(\mathrm{x} - \mathrm{y})\Phi(\mathrm{y})\mathrm{dy}| \leqslant \sup_{\mathrm{y}\in\mathbb{T}} |\Phi(\mathrm{y})| \times \int_{\mathbb{T}} |\mathbf{G}_{\mathrm{t}}(\mathrm{x} - \mathrm{y})|\mathrm{dy} \tag{A.5}$$

$$\leqslant \sup_{\mathrm{y}\in\mathbb{T}} |\Phi(\mathrm{y})| \times \int_{\mathbb{T}} \sum_{k\in\mathbb{Z}} |\mathbf{H}_{\mathrm{t}}(\mathrm{x} + k)|\mathrm{dx} \tag{A.6}$$

$$\leqslant \sup_{\mathrm{y}\in\mathbb{T}} |\Phi(\mathrm{y})| \int_{\mathbb{R}} |\mathbf{H}_{\mathrm{t}}(\mathrm{x})|\mathrm{dx} = \sup_{\mathrm{y}\in\mathbb{T}} |\Phi(\mathrm{y})|. \tag{A.7}$$

Indeed, the last line follows by writing $\mathbb{R}$ as a disjoint union (over $k \in \mathbb{Z}$) of $k + \mathbb{T}$ and a standard Gaussian integration. We also have

$$\int_{\mathbb{T}} |\nabla\mathbf{G}_{\mathrm{t}}(\mathrm{x})|\mathrm{dx} \leqslant \int_{\mathbb{T}} \sum_{k\in\mathbb{Z}} |\nabla\mathbf{H}_{\mathrm{t}}(\mathrm{x} + k)|\mathrm{dx} \leqslant \int_{\mathbb{R}} |\nabla\mathbf{H}_{\mathrm{t}}(\mathrm{x})|\mathrm{dx} \lesssim 1 + \mathrm{t}^{-\frac{1}{2}}, \tag{A.8}$$

where the last bound is a standard Gaussian estimate. This finishes the proof. $\qquad\square$

**Lemma A.3.** *Suppose* $\mathbf{h}^{N,2}(0, \cdot) \in \mathscr{C}^1(\mathbb{T})$. *With probability 1, there exists a unique* $\mathbf{h}^{N,2} \in \mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{C}^1(\mathbb{T}))$ *such that for all* $\mathrm{t} \geqslant 0$ *and* $\mathrm{x} \in \mathbb{T}$, *we have*

$$\mathbf{h}^{N,2}(\mathrm{t}, \mathrm{x}) = (\mathrm{e}^{\mathrm{t}\Delta}\mathbf{h}^{N,2}(0, \cdot))(\mathrm{x}) \tag{A.9}$$

$$+ \int_0^{\mathrm{t}} \left(\mathrm{e}^{(\mathrm{t}-\mathfrak{s})\Delta}\left(\varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s}, \cdot) + \mathbf{h}^{N,2}(\mathfrak{s}, \cdot))) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}(\mathfrak{s}, \cdot))\right)\right)\mathrm{d}\mathfrak{s}. \tag{A.10}$$

*Above,* $\nabla\mathbf{h}^{N,1} = \mathbf{u}^{N,1}$ *from* Lemma A.1. *If, in addition to* Assumption 1.3, *we know* $\mathbf{F}$ *is smooth and satisfies* $|\nabla^k\mathbf{F}(\mathrm{x})| \lesssim_k 1$, *then* $\mathbf{h}^{N,2} \in \mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{C}^2(\mathbb{T}))$ *and, with probability 1, satisfies the following PDE pointwise in* $(\mathrm{t}, \mathrm{x})$:

$$\partial_{\mathrm{t}}\mathbf{h}^{N,2} = \Delta\mathbf{h}^{N,2} + \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1} + \mathbf{h}^{N,2})) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}). \tag{A.11}$$

*Proof.* Let us first note that by Lemma A.1, all derivatives of $\mathbf{h}^{N,1}$ are defined pointwise with probability 1. We now start with a contraction mapping argument. Let $\mathscr{X}$ be the set of functions $\mathbf{f} \in \mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{C}^1(\mathbb{T}))$ such that $\mathbf{f}(0, \cdot) = \mathbf{h}^{N,2}(0, \cdot)$. Define the nonlinear map $\mathfrak{S}$ acting on $\mathscr{X}$ by

$$\mathfrak{S}\mathbf{f}(\mathrm{t}, \mathrm{x}) = (\mathrm{e}^{\mathrm{t}\Delta}\mathbf{f}(0, \cdot))(\mathrm{x}) \tag{A.12}$$

$$+ \int_0^{\mathrm{t}} \left(\mathrm{e}^{(\mathrm{t}-\mathfrak{s})\Delta}\left(\varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s}, \cdot) + \mathbf{f}(\mathfrak{s}, \cdot)) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}(\mathfrak{s}, \cdot)))\right)\right)\mathrm{d}\mathfrak{s}. \tag{A.13}$$

Given $\mathbf{f}_1, \mathbf{f}_2 \in \mathscr{X}$, because $\mathbf{f}_1(0, \cdot) = \mathbf{f}_2(0, \cdot)$ by definition of $\mathscr{X}$, we have

$$\mathfrak{S}\mathbf{f}_1(\mathrm{t}, \mathrm{x}) - \mathfrak{S}\mathbf{f}_2(\mathrm{t}, \mathrm{x}) \tag{A.14}$$

$$= \varepsilon_N^{-1}\int_0^{\mathrm{t}} \left(\mathrm{e}^{(\mathrm{t}-\mathfrak{s})\Delta}\left(\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s}, \cdot) + \mathbf{f}_1(\mathfrak{s}, \cdot))) - \mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s}, \cdot) + \mathbf{f}_2(\mathfrak{s}, \cdot)))\right)\right)(\mathrm{x})\mathrm{d}\mathfrak{s} \tag{A.15}$$

$$\nabla\mathfrak{S}\mathbf{f}_1(\mathrm{t}, \mathrm{x}) - \mathfrak{S}\mathbf{f}_2(\mathrm{t}, \mathrm{x}) = \tag{A.16}$$

$$\varepsilon_N^{-1}\int_0^{\mathrm{t}} \left(\nabla\mathrm{e}^{(\mathrm{t}-\mathfrak{s})\Delta}\left(\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s}, \cdot) + \mathbf{f}_1(\mathfrak{s}, \cdot))) - \mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s}, \cdot) + \mathbf{f}_2(\mathfrak{s}, \cdot)))\right)\right)(\mathrm{x})\mathrm{d}\mathfrak{s}. \tag{A.17}$$

Because $\mathbf{F}$ is uniformly Lipschitz, by Lemma A.2 and elementary manipulations, we have the following bounds in which the implied constant depends on $\mathbf{F}$ only through its Lipschitz constant:

$$| \,(\text{A.15})| \ \lesssim_{N,\mathbf{F}} \ \int_0^{\mathfrak{t}} \int_{\mathbb{T}} |\mathbf{G}_{\mathfrak{t}-\mathfrak{s}}(x-y)| \times |\nabla \mathbf{f}_1(\mathfrak{s},y) - \nabla \mathbf{f}_2(\mathfrak{s},y)| \mathrm{d}y\mathrm{d}\mathfrak{s} \tag{A.18}$$

$$\leqslant \ \mathfrak{t} \sup_{0\leqslant \mathfrak{s}\leqslant \mathfrak{t}} \sup_{y\in\mathbb{T}} |\nabla \mathbf{f}_1(\mathfrak{s},y) - \nabla \mathbf{f}_2(\mathfrak{s},y)| \tag{A.19}$$

$$| \,(\text{A.17})| \ \lesssim_{N,\mathbf{F}} \ \int_0^{\mathfrak{t}} \int_{\mathbb{T}} |\nabla\mathbf{G}_{\mathfrak{t}-\mathfrak{s}}(x-y)| \times |\nabla \mathbf{f}_1(\mathfrak{s},y) - \nabla \mathbf{f}_2(\mathfrak{s},y)| \mathrm{d}y\mathrm{d}\mathfrak{s} \tag{A.20}$$

$$\lesssim \ \int_0^{\mathfrak{t}} (1 + |\mathfrak{t}-\mathfrak{s}|^{-\frac{1}{2}})\mathrm{d}\mathfrak{s} \sup_{0\leqslant \mathfrak{s}\leqslant \mathfrak{t}} \sup_{y\in\mathbb{T}} |\nabla \mathbf{f}_1(\mathfrak{s},y) - \nabla \mathbf{f}_2(\mathfrak{s},y)| \tag{A.21}$$

$$\lesssim \ (\mathfrak{t} + \mathfrak{t}^{\frac{1}{2}}) \sup_{0\leqslant \mathfrak{s}\leqslant \mathfrak{t}} \sup_{y\in\mathbb{T}} |\nabla \mathbf{f}_1(\mathfrak{s},y) - \nabla \mathbf{f}_2(\mathfrak{s},y)|. \tag{A.22}$$

By the above two displays, there exists $c = c(N, \mathbf{F}) > 0$ such that if $T_0 \leqslant c$, then $\mathfrak{S}$ is a contraction on the space of functions $\mathbf{f}$ in $\mathscr{C}([0,T_0], \mathscr{C}^1(\mathbb{T}))$ such that $\mathbf{f}(0,\cdot) = \mathbf{h}^{N,2}(0,\cdot)$. This is a closed subspace of a Banach space $\mathscr{C}([0,T_0], \mathscr{C}^1(\mathbb{T}))$, so it itself is a Banach space. Therefore, there is a unique fixed point of $\mathfrak{S}$. This gives mild solutions until time $T_0$ with probability 1. Because the life-time $T_0$ does not depend on the initial data, we can iterate and obtain mild solutions for infinite life-time with probability 1.

It remains to show that if $\mathbf{F}$ is smooth, then we have classical $\mathscr{C}^2(\mathbb{T})$-valued solutions. To this end, consider the same map $\mathfrak{S}$ but acting on $\mathscr{Y}$, where $\mathscr{Y}$ is the space of functions $\mathbf{f} \in \mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{C}^2(\mathbb{T}))$ such that $\mathbf{f}(0,\cdot) = \mathbf{h}^{N,2}(0,\cdot)$. For any $\mathbf{f}_1, \mathbf{f}_2 \in \mathscr{Y}$, we have

$$\nabla^2 \mathfrak{S}\mathbf{f}_1(\mathfrak{t},x) - \nabla^2 \mathfrak{S}\mathbf{f}_2(\mathfrak{t},x) \ = \ \varepsilon_N^{-1} \times \tag{A.23}$$

$$\int_0^{\mathfrak{t}} \left( \nabla e^{(\mathfrak{t}-\mathfrak{s})\Delta} \left( \nabla\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{f}_1(\mathfrak{s},\cdot))) - \nabla\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{f}_2(\mathfrak{s},\cdot))) \right) \right)(x)\mathrm{d}\mathfrak{s}.$$

We now compute the $\nabla\mathbf{F}$-terms. By the chain rule, for $i = 1, 2$ we have

$$\nabla\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{f}_i(\mathfrak{s},\cdot))) \tag{A.24}$$

$$= \ \varepsilon_N^{1/2}\mathbf{F}'(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{f}_i(\mathfrak{s},\cdot)))(\nabla^2\mathbf{h}^{N,1}(\mathfrak{s},\cdot) - \nabla^2\mathbf{f}_i(\mathfrak{s},\cdot)). \tag{A.25}$$

Thus, because $\mathbf{F}$ has uniformly bounded derivatives by assumption, we deduce

$$|\nabla\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{f}_1(\mathfrak{s},\cdot))) - \nabla\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{f}_2(\mathfrak{s},\cdot)))| \tag{A.26}$$

$$\lesssim_{N,\mathbf{F}} \ |\nabla\mathbf{f}_1(\mathfrak{s},\cdot) - \nabla\mathbf{f}_2(\mathfrak{s},\cdot)| + |\nabla^2\mathbf{f}_1(\mathfrak{s},\cdot) - \nabla^2\mathbf{f}_2(\mathfrak{s},\cdot)|. \tag{A.27}$$

Combining the previous three displays with Lemma A.2, we deduce

$$| \,(\text{A.23})| \tag{A.28}$$

$$\lesssim_{N,\mathbf{F}} \ \int_0^{\mathfrak{t}} \int_{\mathbb{T}} |\nabla\mathbf{G}_{\mathfrak{t}-\mathfrak{s}}(x-y)| \times \left(|\nabla\mathbf{f}_1(\mathfrak{s},y) - \nabla\mathbf{f}_2(\mathfrak{s},y)| + |\nabla^2\mathbf{f}_1(\mathfrak{s},y) - \nabla^2\mathbf{f}_2(\mathfrak{s},y)|\right)\mathrm{d}y\mathrm{d}\mathfrak{s}$$
$$\tag{A.29}$$

$$\lesssim \ \int_0^{\mathfrak{t}} (1 + |\mathfrak{t}-\mathfrak{s}|^{-\frac{1}{2}})\mathrm{d}\mathfrak{s} \times \sup_{0\leqslant \mathfrak{s}\leqslant \mathfrak{t}} \sup_{y\in\mathbb{T}} \left(|\nabla\mathbf{f}_1(\mathfrak{s},y) - \nabla\mathbf{f}_2(\mathfrak{s},y)| + |\nabla^2\mathbf{f}_1(\mathfrak{s},y) - \nabla^2\mathbf{f}_2(\mathfrak{s},y)|\right)$$
$$\tag{A.30}$$

$$\lesssim \ (\mathfrak{t} + \mathfrak{t}^{\frac{1}{2}}) \sup_{0\leqslant \mathfrak{s}\leqslant \mathfrak{t}} \sup_{y\in\mathbb{T}} \left(|\nabla\mathbf{f}_1(\mathfrak{s},y) - \nabla\mathbf{f}_2(\mathfrak{s},y)| + |\nabla^2\mathbf{f}_1(\mathfrak{s},y) - \nabla^2\mathbf{f}_2(\mathfrak{s},y)|\right). \tag{A.31}$$

In particular, there exists $c' = c'(N, \mathbf{F}) > 0$ such that if $T_0 \leqslant c'$, then $\mathfrak{S}$ is a contraction on $\mathscr{Y}$, which is a closed subspace of the Banach space $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{C}^2(\mathbb{T}))$ is therefore a

Banach space itself. This gives a fixed point and thus classical $\mathscr{C}^2(\mathbb{T})$-valued solutions for life-time $T_0$, which we can again bootstrap to infinite life-time, all with probability 1. To show this solution $\mathbf{h}^{N,2}$ satisfies (A.11) pointwise with probability 1, we compute by using $\partial_t e^{t\Delta} = \Delta e^{t\Delta}$ and the fundamental theorem of calculus as follows:

$$\partial_t \mathbf{h}^{N,2}(\mathfrak{t},\mathrm{x}) = \partial_t(e^{t\Delta}\mathbf{h}^{N,2}(0,\cdot))(\mathrm{x}) \tag{A.32}$$

$$+ \partial_t \int_0^t \left( e^{(t-\mathfrak{s})\Delta}\left( \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{h}^{N,2}(\mathfrak{s},\cdot))) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}(\mathfrak{s},\cdot)) \right) \right) \mathrm{d}\mathfrak{s} \tag{A.33}$$

$$= \Delta(e^{t\Delta}\mathbf{h}^{N,2}(0,\cdot))(\mathrm{x}) \tag{A.34}$$

$$+ \int_0^t \partial_t \left( e^{(t-\mathfrak{s})\Delta}\left( \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{h}^{N,2}(\mathfrak{s},\cdot))) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}(\mathfrak{s},\cdot)) \right) \right) \mathrm{d}\mathfrak{s} \tag{A.35}$$

$$+ \left( e^{(t-\mathfrak{s})\Delta}\left( \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{h}^{N,2}(\mathfrak{s},\cdot))) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}(\mathfrak{s},\cdot)) \right) \right)|_{\mathfrak{s}=t} \tag{A.36}$$

$$= \Delta(e^{t\Delta}\mathbf{h}^{N,2}(0,\cdot))(\mathrm{x}) \tag{A.37}$$

$$+ \Delta \int_0^t \left( e^{(t-\mathfrak{s})\Delta}\left( \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1}(\mathfrak{s},\cdot) + \mathbf{h}^{N,2}(\mathfrak{s},\cdot))) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}(\mathfrak{s},\cdot)) \right) \right) \mathrm{d}\mathfrak{s} \tag{A.38}$$

$$+ \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla(\mathbf{h}^{N,1} + \mathbf{h}^{N,2})) - \varepsilon_N^{-1}\mathbf{F}(\varepsilon_N^{1/2}\nabla\mathbf{h}^{N,1}). \tag{A.39}$$

We note that setting $\mathfrak{s} = \mathfrak{t}$ in (A.36) is, rigorously, equal to removing $e^{(t-\mathfrak{s})\Delta}$ (because it is a delta function at $\mathfrak{s} = \mathfrak{t}$) and evaluating the rest at $\mathfrak{s} = \mathfrak{t}$ and $\cdot = \mathrm{x}$. This is because what is inside $e^{(t-\mathfrak{s})\Delta}$ (and thus what is being integrated against an approximation to the identity) is continuous in space-time with probability 1 (because of almost sure regularity of $\mathbf{F}$ and $\nabla\mathbf{h}^{N,1}$ and $\nabla\mathbf{h}^{N,2}$). Now, note (A.37)-(A.38) is equal to $\Delta\mathbf{h}^{N,2}(\mathfrak{t},\mathrm{x})$, and (A.39) is the rest of (A.11). This completes the proof. $\square$

## B  Wasserstein distance

In this section, we specialize to the Polish space given by the space of differentiable functions on $\mathbb{T}$ that are bounded and whose derivatives are bounded. We will let $\mathscr{X}$ denote its Banach dual, equipped with the operator norm topology.

**Definition B.1.** *Provided any probability measures $\mathbf{P}(1), \mathbf{P}(2)$ on $\mathscr{X}$, the Wasserstein metric is given by the following in which the infimum is given over all couplings of $\mathbf{P}(1)$ and $\mathbf{P}(2)$, namely all measures $\mathbf{P}(1,2)$ on $\mathscr{X} \times \mathscr{X}$ such that the projection onto the first coordinate is $\mathbf{P}(1)$ and the projection onto the second coordinate is $\mathbf{P}(2)$:*

$$\mathsf{W}(\mathbf{P}(1), \mathbf{P}(2)) = \inf_{(\mathbf{u}^1, \mathbf{u}^2) \sim \mathbf{P}(1,2)} \mathbf{E}^{\mathbf{P}(1,2)}\|\mathbf{u}^1 - \mathbf{u}^2\|_{\mathscr{X}}. \tag{B.1}$$

## C  Relative entropy

**Definition C.1.** *Provided any complete separable metric space $\mathscr{S}$ and any two probability measures $\mathbf{P}(1)$ and $\mathbf{P}(2)$, the relative entropy of $\mathbf{P}(1)$ with respect to $\mathbf{P}(2)$ is equal to $\infty$ if $\mathbf{P}(1)$ is not absolutely continuous with respect to $\mathbf{P}(2)$, and if it is (so that $\mathbf{P}(1) \ll \mathbf{P}(2)$), it is the following in which $\mathbf{R}(1,2)$ is the Radon-Nikodym derivative of $\mathbf{P}(1)$ with respect to $\mathbf{P}(2)$ and in which the expectation below is with respect to $\mathbf{P}(2)$:*

$$\mathsf{H}(\mathbf{P}(1)|\mathbf{P}(2)) \overset{\bullet}{=} \mathbf{E}^{\mathbf{P}(2)}\mathbf{R}(1,2)\log\mathbf{R}(1,2). \tag{C.1}$$

We now collect a set of standard relative entropy estimates that are used crucially in this paper. These estimates are used in [17, 21], for example, and are rather standard, so we only provide brief proofs.

**Lemma C.2.** *Consider any complete separable metric space $\mathscr{S}$ with any two probability measures $\mathbf{P}^1$ and $\mathbf{P}^2$ and any event $\mathcal{E} \subseteq \mathscr{S}$. We first have the following entropy inequality:*

$$\mathbf{P}^1(\mathcal{E}) \;\leqslant\; \left(\log 2 + \mathsf{H}(\mathbf{P}^1|\mathbf{P}^2)\right)\left(\log\left(1 + \mathbf{P}^2(\mathcal{E})^{-1}\right)\right)^{-1}. \tag{C.2}$$

*Retain the context prior to (C.2). For another complete separable metric space $\mathscr{S}'$ and continuous map $\Pi : \mathscr{S} \to \mathscr{S}'$, if $\Pi_*$ denotes the pushforward on probability measures on $\mathscr{S}$, we have the following contraction principle:*

$$\mathsf{H}(\Pi_*\mathbf{P}^1|\Pi_*\mathbf{P}^2) \;\leqslant\; \mathsf{H}(\mathbf{P}^1|\mathbf{P}^2). \tag{C.3}$$

*We now let $\mathbf{P}^{1,\mathrm{dyn}}$ and $\mathbf{P}^{2,\mathrm{dyn}}$ be path-space measures on $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{S})$ given by the same Markov process valued in $\mathscr{S}$ but with initial law given by $\mathbf{P}^1$ and $\mathbf{P}^2$, respectively. Then the relative entropy of $\mathbf{P}^{1,\mathrm{dyn}}$ with respect to $\mathbf{P}^{2,\mathrm{dyn}}$ is controlled by that of $\mathbf{P}^1$ with respect to $\mathbf{P}^2$, so that we have the following inequality:*

$$\mathsf{H}(\mathbf{P}^{1,\mathrm{dyn}}|\mathbf{P}^{2,\mathrm{dyn}}) \;\leqslant\; \mathsf{H}(\mathbf{P}^1|\mathbf{P}^2). \tag{C.4}$$

*Lastly, if a sequence of path-space measures $\{\mathbf{P}^{2,\mathrm{dyn},N}\}_{N \geqslant 0}$ on $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{S})$ is tight and the relative entropies of another set of path-space measures $\{\mathbf{P}^{1,\mathrm{dyn},N}\}_{N \geqslant 0}$ with respect to $\{\mathbf{P}^{2,\mathrm{dyn},N}\}_{N \geqslant 0}$ are uniformly bounded in $N$, then the sequence $\{\mathbf{P}^{1,\mathrm{dyn},N}\}_{N \geqslant 0}$ on $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{S})$ is also tight.*

*Proof.* The estimate (C.2) is (2.6) in [17]. Proving (C.3) starts with the duality formula below for relative entropy, which itself follows by noting the convex conjugate of $\mathrm{x}\log\mathrm{x}$ is $\mathrm{e}^{\mathrm{x}}$ (see Corollary 4.15 in [6]):

$$\mathsf{H}(\Pi_*\mathbf{P}^1|\Pi_*\mathbf{P}^2) \;=\; \sup_{\psi \in \mathscr{L}^\infty(\mathscr{S}')} \left(\mathbf{E}^{\Pi_*\mathbf{P}^1}\psi - \log \mathbf{E}^{\Pi_*\mathbf{P}^2}\mathrm{e}^\psi\right). \tag{C.5}$$

For any $\psi \in \mathscr{L}^\infty(\mathscr{S}')$, we have the following by definition of pushforward:

$$\mathbf{E}^{\Pi_*\mathbf{P}^1}\psi - \log \mathbf{E}^{\Pi_*\mathbf{P}^2}\mathrm{e}^\psi \;=\; \mathbf{E}^{\mathbf{P}^1}\psi \circ \Pi - \log \mathbf{E}^{\mathbf{P}^2}\psi \circ \Pi. \tag{C.6}$$

If $\psi \in \mathscr{L}^\infty(\mathscr{S}')$, then $\psi \circ \Pi \in \mathscr{L}^\infty(\mathscr{S})$. Thus,

$$\sup_{\psi \in \mathscr{L}^\infty(\mathscr{S}')} \left(\mathbf{E}^{\Pi_*\mathbf{P}^1}\psi - \log \mathbf{E}^{\Pi_*\mathbf{P}^2}\mathrm{e}^\psi\right) \;=\; \sup_{\psi \in \mathscr{L}^\infty(\mathscr{S}')} \left(\mathbf{E}^{\mathbf{P}^1}\psi \circ \Pi - \log \mathbf{E}^{\mathbf{P}^2}\psi \circ \Pi\right) \tag{C.7}$$

$$\leqslant\; \sup_{\psi \in \mathscr{L}^\infty(\mathscr{S})} \left(\mathbf{E}^{\mathbf{P}^1}\psi - \log \mathbf{E}^{\mathbf{P}^2}\mathrm{e}^\psi\right), \tag{C.8}$$

which gives (C.3) when combined with (C.5). The estimate (C.4) follows from conditioning the path-space measures $\mathbf{P}^{i,\mathrm{dyn}}$ as the same path-space measure with initial conditions sampled via $\mathbf{P}^i$ data, respectively, and then employing the chain rule for relative entropy (see Section 4.2 in [6]), which cancels the contribution of the shared path-space part of each $\mathbf{P}^{i,\mathrm{dyn}}$ measure. In particular, by said chain rule for relative entropy, we decompose $\mathsf{H}(\mathbf{P}^{1,\mathrm{dyn}}|\mathbf{P}^{2,\mathrm{dyn}})$ into relative entropy of initial data laws plus "worst-case relative entropy" of the dynamics themselves:

$$\mathsf{H}(\mathbf{P}^{1,\mathrm{dyn}}|\mathbf{P}^{2,\mathrm{dyn}}) \;\leqslant\; \mathsf{H}(\mathbf{P}^1|\mathbf{P}^2) + \sup_{s \in \mathscr{S}} \mathsf{H}(\mathbf{P}^{1,\mathrm{dyn}}|_s|\mathbf{P}^{2,\mathrm{dyn}}|_s), \tag{C.9}$$

where $\mathsf{H}(\mathbf{P}^{1,\mathrm{dyn}}|_s|\mathbf{P}^{2,\mathrm{dyn}}|_s)$ is the relative entropy of the path-space measure $\mathbf{P}^{1,\mathrm{dyn}}$ conditioning on the initial data to be $s \in \mathscr{S}$, with respect to $\mathbf{P}^{2,\mathrm{dyn}}$ conditioning on the same initial data $s \in \mathscr{S}$. But these two path-space measures are the same because they have the same initial data and same dynamics, so the sup on the RHS above is equal to zero. The last paragraph follows by (C.2), which estimates complements of compact sets in $\mathscr{C}(\mathbb{R}_{\geqslant 0}, \mathscr{S})$ under $\mathbf{P}^{1,\mathrm{dyn},N}$ by their probabilities under $\mathbf{P}^{2,\mathrm{dyn},N}$ and using tightness of $\mathbf{P}^{2,\mathrm{dyn},N}$. $\qquad\square$

# References

[1] G. Amir, I. Corwin, J. Quastel, "Probability distribution of the free energy of the continuum directed polymer model in $(1 + 1)$-dimensions". *Communications in Pure and Applied Math*, 64:466-537 (2011). MR2796514

[2] G. Barraquand, P. Le Doussal, "Steady state of the KPZ equation on an interval and Liouville quantum mechanics". arXiv:2105.15178. 2021.

[3] Ben-Artzi, Matania. "Lectures on viscous Hamilton-Jacobi equations". http://math.huji.ac.il/~mbartzi/recent-publications/PIMS-course.pdf.

[4] L. Bertini, G. Giacomin, "Stochastic Burgers and KPZ Equations from Particle Systems". *Communications in Mathematical Physics* Volume 183, Issue 3, p. 571-606, (1997). MR1462228

[5] P. Billingsley. *Convergence of Probability Measures*. New York: John Wiley & Sons, Inc. ISBN 0-471-19745-9. MR0233396

[6] S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Print ISBN-13: 9780199535255. DOI: 10.1093/acprof:oso/9780199535255.001.0001. 2013. MR3185193

[7] G. Cannizzaro, D. Erhard, P. Schonbauer, "2D Anisotropic KPZ at stationarity: scaling, tightness and non triviality". arXiv:1907.01530. 2019.

[8] I. Corwin, "The Kardar-Parisi-Zhang equation and universality class". Random Matrices: Theory and Applications Vol. 01, No. 01, 1130001 (2012) MR2930377

[9] I. Corwin, A. Knizel. "Stationary measure for the open KPZ equation". arXiv:2103.12253. 2021.

[10] I. Corwin, H. Shen, "Open ASEP in the weakly asymmetric regime". Communications on Pure and Applied Mathematics 71 (10), 2065-2128. MR3861074

[11] A. Dembo, L.-C. Tsai, "Weakly asymmetric non-simple exclusion process and the KPZ equation". *Communications in Mathematical Physics*, 341 (1), 219-261, (2016). MR3439226

[12] L. C. Evans, *Partial differential equations*, American Mathematical Society, Providence, R.I. 2010. MR2597943

[13] T. Funaki, J. Quastel. "KPZ equation, its renormalization and invariant measures". *Stochastic Partial Differential Equations: Analysis and Computations* 3, 159-220. 2015. MR3350451

[14] M. Gubinelli, P. Imkeller, N. Perkowski, "Paracontrolled distributions and singular PDEs". *Forum of Mathematics, Pi*, vol. 3, p. e6., doi: 10.1017/fmp.2015.2. 2015. MR3406823

[15] P. Goncalves, M. Jara, "Nonlinear Fluctuations of Weakly Asymmetric Interacting Particle Systems". *Archive for Rational Mechanics and Analysis*, (2014) 212:597. MR3176353

[16] P. Goncalves, M. Jara, "Stochastic Burgers equation from long range exclusion interactions". *Stochastic Processes and their Applications*, Volume 127, Issue 12, December 2017, Pages 4029-4052. MR3718105

[17] P. Goncalves, M. Jara, S. Sethuraman, "A stochastic Burgers equation from a class of microscopic interactions". *Annals of Probability*, Volume 43, Number 1 (2015), 286-338. MR3298474

[18] L. Gross, "Logarithmic Sobolev Inequalities". *American Journal of Mathematics*, Winter, 1975, Vol. 97, No. 4, pp. 1061-1083. 1975. MR0420249

[19] M. Gubinelli, M. Jara. "Regularization by noise and stochastic Burgers equations". *Stochastic Partial Differential Equations: Analysis and Computations* Volume 1. 325-350. 2013. MR3327509

[20] M. Gubinelli, N. Perkowski, "Energy solutions of KPZ are unique". *Journal of the AMS*. 31 (2018), 427-471. MR3758149

[21] M. Gubinelli, N. Perkowski, "The Hairer-Quastel universality result at stationarity (Stochastic Analysis on Large Scale Interacting Systems)". *RIMS Kokyuroku Bessatsu* B59, 101-115. 2016. MR3675927

[22] M. Gubinelli, N. Perkowski, "Probabilistic approach to the stochastic Burgers equation". arXiv:1701.07373 [math.PR], Jan. 25, 2017. MR3828193

[23] M. Gubinelli, N. Perkowsi, "The infinitesimal generator of the stochastic Burgers equation". *Prob. Theory and Rel. Fields*. 178, pages 1067-1124. 2020. MR4168394

[24] M.Z. Guo, G.C. Papnicolaou, and S.R.S. Varadhan. "Nonlinear diffusion limit for a system with nearest neighbor interactions". *Commun. Math. Phys.* (1988) 118:31. MR0954674

[25] M. Hairer, "Solving the KPZ equation". *Annals of Mathematics*, Volume 178, Issue 2, (2013), p. 559-664. MR3071506

[26] M. Hairer, "A Theory of Regularity Structures". *Invent. Math.* 198, no. 2 (2014), 269-504. MR3274562

[27] M. Hairer, J. Mattingly. "The strong Feller property for singular stochastic PDEs". *Ann. Inst. H. Poincare Probab. Statist.* 54(3): 1314-1340. 2018. MR3825883

[28] M. Hairer, J. Quastel, "A class of growth models rescaling to KPZ". *Forum of Mathematics, Pi*, 6, E3. doi: 10.1017/fmp.2018.2. 2018. MR3877863

[29] M. Hairer, W. Xu. "Large scale limit of interface fluctuation models". Ann. Probab., Volume 47, Number 6 (2019), 3478-3550. MR4038037

[30] M. Kardar, G. Parisi, Y.-C. Zhang, "Dynamic scaling of growing interfaces", *Physical Review Letters* 56 (1986), no. 9, 889.

[31] C. Kipnis, C. Landim, *Scaling Limits of Interacting Particle Systems*, Springer-Verlig Berlin Heidelberg, Volume 320, (1999). MR1707314

[32] C. Mueller, "On the support of solutions to the heat equation with noise". *Stochastics and Stochastics Reports*. 1991. MR1149348

[33] I. Mitoma, "Tightness of Probabilities on $C([0,1]; \mathscr{Y}')$ and $D([0,1]; \mathscr{Y}')$". *Annals of Probability*, Volume 11 No. 4 (1983), 989-999. MR0714961