Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Hölder regularity for a non-linear parabolic equation driven by space-time white noise

by

Felix Otto and Hendrik Weber

Preprint no.: 31 2015

Hölder regularity for a non-linear parabolic equation driven by space-time white noise

F. Otto, H. Weber

Abstract. We consider the non-linear equation $u + \partial_t u - \partial_x^2 \pi(u) = \xi$ driven by space-time white noise ξ, which is uniformly parabolic because we assume that π' is bounded away from zero and infinity. Under the further assumption of Lipschitz continuity of π' we show that the stationary solution is — as for the linear case — almost surely Hölder continuous with exponent α for any $\alpha < \frac{1}{2}$ w. r. t. the parabolic metric. More precisely, we show that the corresponding local Hölder norm has almost Gaussian moments.

On the stochastic side, we use a combination of martingale arguments to get second moment estimates with concentration of measure arguments to upgrade to Gaussian moments. On the deterministic side, we appeal to finite and infinitesimal versions of the H^{-1}-contraction principle and a Campanato iteration.

1 Introduction and main result

We are interested in the stochastic nonlinear parabolic equation

$$T^{-1}u + \partial_t u - \partial_x^2 \pi(u) = \xi, \tag{1}$$

where ξ denotes space-time white noise. The nonlinear character of (1) is that of a fully nonlinear equation rather than a quasi-linear equation, since rewriting (1) as the quasi-linear equation (7) is not helpful as we explain below, and since the deterministic estimates we need are related to the linearization of a fully nonlinear equation, cf. (12), rather than to the linearization of a quasi-linear equation (this distinction would be more pronounced in a multi-dimensional case). We assume that the nonlinearity π is uniformly elliptic in the sense that there exists a $\lambda > 0$ such that

$$\lambda \leq \pi'(u) \leq 1 \quad \text{for all } u \in \mathbb{R}. \tag{2}$$

In particular, this rules out the degenerate case that goes under the name of porous medium equation. Furthermore, we assume some regularity of π in the sense that there exists $L < \infty$ such that

$$|\pi''(u)| \leq L \quad \text{for all } u \in \mathbb{R}. \tag{3}$$

We are interested in Hölder regularity of solutions of (1); the simplest solution to (1) is the space-time stationary solution u of (1) on which we shall focus in this paper. The main reason for including the massive term in (1) (i. e. assuming $T < \infty$) is to ensure existence and uniqueness of this object; the only other role is to provide a large-scale estimate through Lemma 1. In this first version of the paper, we will be *completely informal* about why and in which sense (1) is well-posed, and why the martingale and concentration of measure arguments can be carried out (we will just motivate them when we first need them).

A crucial insight is that the law of the (unique) stationary random field u is invariant under the rescaling

$$x = R\hat{x}, \quad t = R^2\hat{t}, \quad u = R^{\frac{1}{2}}\hat{u}, \tag{4}$$

provided, the nonlinearity and the massive term are adjusted according to

$$\hat{\pi}(\hat{u}) = R^{-\frac{1}{2}}\pi(R^{\frac{1}{2}}\hat{u}), \quad \hat{T} = R^{-2}T. \tag{5}$$

For this observation we used that in view of its defining relation $\langle \big(\int \zeta\xi dxdt\big)^2 \rangle = \int \zeta^2 dxdt$ for a test function ζ (that is, loosely speaking $\int \zeta\langle \xi(t,x)\xi(0,0)\rangle dxdt = \zeta(0,0)$), space-time white noise rescales as $\xi = \frac{1}{\sqrt{RR^2}}\hat{\xi} = R^{-\frac{3}{2}}\hat{\xi}$. From this invariance property we learn that as we go to small scales (i. e. $R \ll 1$), the effective nonlinearity as measured by the Lipschitz constant L of π' in (3) decreases according to

$$\hat{L} = R^{\frac{1}{2}}L. \tag{6}$$

This suggests that on small scales, u has the same regularity as if (1) were replaced by its linear version (without massive term) $\partial_t u - a_0\partial_x^2 u = \xi$ for some constant $a_0 \in [\lambda, 1]$. Hence we expect that on small scales, u is Hölder continuous with exponents α (in the parabolic Carnot-Carathéodory geometry) for any $\alpha < \frac{1}{2}$. This is exactly what we show, making crucial use of the above scale invariance.

We note in passing that it is *not* helpful to write the elliptic operator in the more symmetric form

$$T^{-1}u + \partial_t u - \partial_x(\pi'(u)\partial_x u) = \xi, \tag{7}$$

since even in case of the stochastic heat equation, u (and thus $\pi'(u)$) is a function in the Hölder space with exponent $\frac{1}{2}-$ so that $\partial_x u$ would be a distribution in the (negative) Hölder space with exponent $-\frac{1}{2}-$, so that there is no standard distributional definition of the product $\pi'(u)\partial_x u$. In fact, rather than appealing to regularity theory for linear but non-constant coefficient equations of the form $T^{-1}u + \partial_t u - \partial_x(a\partial_x u) = \partial_x g$, we have to appeal to the theory for $T^{-1}w + \partial_t w - \partial_x^2(aw) = \partial_x^2 g$, cf. Proposition 3.

Let us now briefly comment on existing regularity theory for non-linear parabolic stochastic differential equation of the type of (1). There is a large body of literature on stochastic equations of the type (1), but mostly with a quite different focus: The focus there is to tackle on the one hand more nonlinear situations, like the case of a degenerate ellipticity (i. e. $\lambda = 0$ in (2)) or the case of multiplicative noise, but on the other hand to assume "whatever it takes" on the spatial covariance structure of the noise. Sometimes, structural assumptions allow to mimic an approach that is obvious in the semi-linear case, namely the approach of decomposing the solution into a rough part w that solves a more explicitly treatable stochastic differential equation and a more regular part v that solves a parabolic equation with random coefficients and/or right-hand-side described through w, and then allows for an application of deterministic regularity theory. We refer to [6] for an example with a multiplicative decomposition of this type. The recent work by Debussche et. al. on quasi-linear parabolic stochastic equations, i. e. equations of the form (7) or more generally with an elliptic operator of the form $-\nabla \cdot a(u)\nabla u$, refines this approach to a fixed point argument, and appeals to Nash's Hölder a priori bound on linear parabolic equations with just uniformly elliptic coefficients as a starting point to bootstrap to the optimal Hölder continuity via Schauder theory, see [3, Introduction]. However, cf. the above discussion of (7), this treatment seems limited to situations where the noise ξ is so regular that in the case of the linear equation, ∇u is at least locally integrable in time-space (to be more quantitative: $\int |\nabla u|^p dxdt < \infty$ for some $p > 3$ on the level of space-time isotropic L^p-norms).

By the equivalence of Campanato and Hölder spaces, see for instance [9, Theorem 5.5], Hölder continuity can be expressed in terms of a localized L^2-modulus of continuity. Because of the eventual conditioning on the distant noise, it is more convenient to replace a sharp spatial localization on parabolic cylinders by a soft localization via an exponentially decaying function

$$\eta(x) = \frac{1}{2}\exp(-|x|), \quad \eta_r(x) := \frac{1}{r}\eta(\frac{x}{r}), \tag{8}$$

note that the normalization imply that $\int \eta_r \cdot dx$ corresponds to a spatial average that is localized near the origin on scale r. We note that while the exponential form of the cut-off is probably not essential (any thicker than Gaussian tails should suffice), it is convenient at many places of the proof. Abbreviating the L^2-modulus of continuity at the origin and on parabolic scale r by

$$D^2(u,r) := \fint_{-r^2}^{0} \int \eta_r(u - \fint_{-r^2}^{0} \int \eta_r u)^2 dxdt,$$

our main result reads as follows:

Theorem 1. *Let u be the unique (in law) stationary solution to* (1). *W. l. o. g. suppose that $T = 1$ in (1). For any Hölder exponent $\alpha \in (0, \frac{1}{2})$ we have almost a Gaussian bound for the α-Hölder L^2-averaged modulus of continuity at the origin in the sense that for any exponent $\epsilon > 0$ we have*

$$\Big\langle \exp\Big(\big(\sup_{r\leq 1} \frac{1}{r^\alpha} D(u,r)\big)^{2(1-\epsilon)}\Big)\Big\rangle \leq C, \tag{9}$$

with a constant $C < \infty$ only depending on $L < \infty$, $\alpha < \frac{1}{2}$ and $\epsilon > 0$.

Theorem 1 implies stretched exponential – in fact, almost Gaussian – bounds for the more conventional local Hölder semi-norms of the random field u. For any $\alpha \in (0,1)$ we set

$$[u]_\alpha = \sup_{R\in(0,1)} \frac{1}{R^\alpha} \sup_{\substack{(t,x),(s,y)\in(-1,0)\times(-1,1) \\ \sqrt{|t-s|}+|x-y|<R}} |u(t,x) - u(s,y)| \,.$$

Theorem 1 implies the following:

Corollary 1. *Under the assumptions of Theorem 1 we have*

$$\big\langle \exp\big([u]_\alpha^{2(1-\epsilon)}\big)\big\rangle \leq C$$

for a constant $C < \infty$ which only depends on $L < \infty$, $\alpha < \frac{1}{2}$ and $\epsilon > 0$.

2 Strategy of proof and ingredients

Theorem 1, like Lemma 1 below, relies on a concentration of measure argument for Lipschitz random variables: For any a random variable F that is 1-Lipschitz when considered as a path-wise functional of the white noise ξ, one has $\langle\exp(\lambda F)\rangle \leq \exp(\lambda\langle F\rangle + \frac{1}{2}\lambda^2)$ for any number λ. In particular, if $F \geq 0$ is 1-Lipschitz and satisfies $\langle F\rangle \leq 1$, it has Gaussian moments $\langle\exp(\frac{1}{C}F^2)\rangle \leq 1$, for some universal constant $C < \infty$. Here the norm underlying the Lipschitz property is the norm of the Cameron-Martin space, which simply means that infinitesimal variations $\delta\xi$ of the space-time white noise are measured in the space-time L^2-norm. To continue with the name-dropping, this type of Lipschitz continuity means that the carré-du-champs $|\nabla F|^2$ of the Malliavin derivative is bounded independently of the given realization of the noise, where for a given realization ξ of the noise, $|\nabla F|$ is the smallest constant Λ in

$$|\delta F| \leq \Lambda\Big(\int(\delta\xi)^2 dxdt\Big)^{\frac{1}{2}}. \tag{10}$$

Here δF denotes the infinitesimal variation of F generated by the infinitesimal variation $\delta\xi$ of the noise ξ, a linear relation captured by the Fréchet derivative (a linear form) of F w. r. t. ξ. For those not confident in this continuum version of concentration of measure we derive it from the discrete case in the proof of Lemma 1, where also the type of martingale arguments entering Proposition 1 via Lemma 2 (and Lemmas 3 and 4 again) is explained for the non-expert.

Concentration of measure will be applied to the random variable $F = D(u,r)$. It is Proposition 2 which provides the bound on the Malliavin derivative w. r. t. to the ensemble $\langle\cdot\rangle_1$ that describes the space time white noise ξ *restricted* to the time slice $(t,x) \in (-1,0)\times\mathbb{R}$. In particular, this means that the admissible variations $\delta\xi$ in (10) are supported in $(t,x)\in(-1,0)\times\mathbb{R}$; we denote by $|\nabla F|_1^2$ the corresponding carré-du-champs. Proposition 1 in turn provides the estimate of the (conditional) expectation, that is, the expectation in $\langle\cdot\rangle_1$.

So the combination of Propositions 1 and 2 yield Gaussian moments for $\frac{1}{r^\alpha}D(u,r)$, however only up to $1 + r^{\frac{1}{2}}D'(u,1)$, which roughly behaves as $1 + r^{\frac{1}{2}}D(u,1)$, and modulo the multiplicative (and nonlinear) error of $\frac{L}{r^{\frac{3}{2}}}(1 + D(u,1))$. Evoking the scale invariance (4) & (5), this estimate will be used for small scales, where thanks to the behavior (6) of L, the multiplicative error fades away. This amounts to a (stochastic) Campanato iteration which ultimately yields Theorem 1. Since both propositions will be applied to small scales, so that in view of (5) also the massive term fades away, we cannot expect help from it; as a matter of fact, we will ignore the massive term in the proof (besides in Lemma 1).

Proposition 1. *Pick a Hölder exponent $\alpha \in (0,\frac{1}{2})$. Then we have all $r \le 1$*

$$\langle D(u,r)\rangle_1 \lesssim r^\alpha\big(1 + \frac{L}{r^{\frac{3}{2}}}(D'(u,1)+1)\big)\big(1 + r^{\frac{1}{2}}D'(u,1)\big),$$

where $D'(u,1)$ depends only on $u(t=-1,\cdot)$:

$$D'^2(u,1) := \int \eta(u - \int \eta u)^2 dx_{|t=-1}.$$

Here and in the proof, $\lesssim$ means up to a constant only depending on $\lambda > 0$ and $\alpha < \frac{1}{2}$.

Proposition 2. *We have for the carré-du-champs of the Malliavin derivative*

$$|\nabla D(u,r)|_1 \lesssim r^{\frac{1}{2}} + \frac{L}{r^{\frac{3}{2}}}D(u,1),$$

where here and in the proof, $\lesssim$ means up to a constant only depending on $\lambda > 0$.

The only purpose of the presence of the massive term is that in the original scale, it provides control of the L^2-averaged Hölder continuity on scales 1, and thus the anchoring for the Campanato iteration:

Lemma 1. *Suppose that $T = 1$ in (1). Then we have*

$$\big\langle \exp\big(\frac{1}{C}D^2(u,1)\big)\big\rangle \le C$$

for some constant C only depending on λ.

In order to derive Propositions 1 and 2, we will consider differences of solutions to (1) for Proposition 1, or infinitesimal perturbations of solutions for Proposition 2. Finite or infinitesimal differences of solutions satisfy a formally linear parabolic equation with an inhomogeneous coefficient field a, which in view of (2) is uniformly elliptic:

$$\lambda \le a(t,x) \le 1 \quad \text{for all } (t,x) \in (-1,0)\times\mathbb{R}. \tag{11}$$

The linearized operator comes in the conservative form of $\partial_t u - \partial_x^2(au)$. For a priori estimates of the corresponding initial value problem, it is most natural to write the r. h. s. also in conservative form:

$$\partial_t w - \partial_x^2(aw) = \partial_t h + \partial_x^2 g. \tag{12}$$

The L^2-estimates on solutions of (12) from Proposition 3 might be seen as an infinitesimal version of the $\dot{H}^{-1}$-contraction principle for the deterministic counterpart of (1), which will be explicitly used in Lemma 2, see the proof of Lemma 1, which is a good starting point for the PDE arguments, too.

Following a standard approach in Schauder theory for parabolic (and elliptic) equations, we also consider (12) with constant coefficients $a_0 \in [\lambda, 1]$, which will arise from locally "freezing" the variable coefficient field a:

$$\partial_t v - a_0\partial_x^2 v = f. \tag{13}$$

Proposition 4 states classical L^∞ and Hölder-$\frac{1}{2}$ estimates for (13), the only difficulty coming from the low regularity of the initial data $v_{|t=-1}$ and the moderate regularity of the r. h. s. f assumed in Proposition 4. We give a self-contained proof.

Proposition 3. *Consider a solution w of (12) with r. h. s. described by (g,h) and vanishing initial data:*

$$w = h = 0 \quad for\, t = -1.$$

Then we have both the local estimate

$$\int_{-1}^{0}\int \eta w^2 dxdt \lesssim \int_{-1}^{0}\int \eta(g^2+h^2)dxdt \tag{14}$$

and the global estimate

$$\int_{-1}^{0}\int w^2 dxdt \lesssim \int_{-1}^{0}\int (g^2+h^2)dxdt. \tag{15}$$

Here and in the proof $\ll$ and $\lesssim$ refer just to λ.

Proposition 4. *Consider a solution v of (13) with r. h. s. f. Then we have a localized L^∞-estimate*

$$\sup_{(t,x)\in(-1,0)\times\mathbb{R}} (t+1)^{\frac{1}{2}}\eta v^2 \lesssim \int_{-1}^{0}\int \eta f^2 dxdt + \int \eta v^2 dx_{|t=-1}. \tag{16}$$

In case v has vanishing initial data in the sense of $v(t=-1,\cdot)=0$, we also claim the L^2-averaged Hölder-$\frac{1}{2}$ estimate

$$\sup_{r\le 1}\frac{1}{r}D^2(v,r) \lesssim \int_{-1}^{0}\int \eta f^2 dxdt. \tag{17}$$

Here and in the proof $\ll$ and $\lesssim$ refer just to λ.

We'd like to point out a synergy in terms of methods between this approach to regularity for stochastic partial differential equations driven by stationary noise, and an approach to regularity for elliptic partial differential equations with stationary random coefficient field that is emerging over the past years [10, 1, 7]. At first glance, the differences dominate: Here, we have a *nonlinear* and *parabolic* partial differential equation driven by a random right-hand-side ξ, and we hope for almost-sure *small-scale* regularity *despite the short-range decorrelation* of ξ, which implies its roughness. There, the main features already appear on the level of a *linear* and *elliptic* equation, for instance on the level of the harmonic coordinates or the corrector ϕ_i given by $-\nabla\cdot a(\nabla\phi_i+e_i)=0$ where e_i is the i-th unit vector, and one hopes for almost-sure *large scale* regularity *thanks to the long-range decorrelation* of the coefficient field a. In the first case, *randomness limits* Hölder regularity, whereas in the second case, *randomness improves* Hölder regularity: In fact, for almost every realization of a, a-harmonic functions u satisfy a first-order Liouville principle [7], and even Liouville principles of any order [5], which is the simplest way to encode large-scale Hölder regularity. Even the lowest-order Liouville principle is known to fail for some uniformly elliptic and smooth coefficient fields a, so that these results indeed show a regularizing effect of randomness.

Despite these obvious differences, the approach is very similar: Both here and there (in [10] and, more explicitly, in [7]) one is appealing to the combination of sensitivity estimates (how do certain functionals of the solution depend on the right hand side here, or on the coefficient field there?) measured in terms of a carré du champs (of the Malliavin derivative here, or of a suitable vertical derivative that is compatible with the correlation structure there), and then appeals to concentration of measure (on the Gaussian level here, or via the intermediate of a Logarithmic Sobolev Inequality there).

Such a synergy in methods that treat models with thermal noise like in high- or infinite dimensional stochastic differential equations with reversible invariant (Gibbs) measure and those that treat models with quenched noise like in stochastic homogenization is not new: In their seminal work on Gradient Gibbs measures, a model in statistical mechanics that describes thermally fluctuating surfaces, Naddaf and Spencer appeal to stochastic homogenization to characterize the large-scale correlation structure of the field [11]. Their analysis can also be interpreted as considering the infinite-dimensional stochastic differential equation of which the measure is the reversible invariant measure, an equation which can be seen as a spatial discretization of a stochastic nonlinear parabolic partial differential equation, and to consider the Malliavin derivative of its solution with respect to the (discrete) space-time white noise [4]. Again, the nonlinearity is rather of the symmetric form (7) and Naddaf and Spencer appeal to Nash's heat kernel bounds."

We close this parenthesis by noting that for stochastic partial differential equations and stochastic homogenization, even the deterministic ingredients are similar: In both cases, the sensitivity estimate leads to a *linear* partial differential equation (parabolic here, elliptic there) with a priori only uniformly elliptic coefficient field (in space-time here, in space there), that is, without any a priori modulus of continuity. In both cases, a buckling argument is needed to obtain bounds on Hölder norms with high stochastic integrability. While here, the need of a buckling estimate is obvious since the small-scale regularity of the coefficient field $a = \pi'(u)$ in the sensitivity equation is determined by the small-scale regularity of the solution u around which one is linearizing, the buckling is less obvious there: It turns out that the large-scale regularity properties of the operator $-\nabla\cdot a\nabla$ are determined by the large-scale properties of the harmonic coordinates $x_i+\phi_i$, the special solution mentioned above. Here, buckling proceed by showing that the linear operator $\partial_t-\partial_x^2 a$ is close to a constant coefficient operator $\partial_t-a_0\partial_x^2$ on small scales, there, it proceeds by showing that it is close to a constant coefficient operator on large scales, namely the homogenized operator $-\nabla\cdot a_{hom}\nabla$. In both cases, a Campanato-type iteration is the appropriate deterministic tool for the buckling. Here, this is not surprising since Campanato iteration is a robust way of deriving Hölder estimates (see for instance [9, Chapter 5]); there, the use of Campanato iteration to push the constant-coefficient regularity theory from the infinite scale to large but finite scales was first introduced in [2] in case of periodic homogenization, then transferred to stochastic homogenization in [1], and refined in [7] in a way that brings it very close to its small-scale application.

After this aside, we turn back to our proof. Next to these deterministic ingredients, Proposition 1 also requires a couple of classical, second moment stochastic estimates. The first lemma provides such a low-stochastic moment estimate on the L^2-Hölder-$\frac{1}{2}$ modulus of continuity, which however is restricted to a spatial modulus and is only localized to scales 1. This spatial L^2-Hölder modulus of continuity is expressed in terms of the L^2-difference of spatial shifts (which are then exponentially averaged over the shifts); this form arises naturally from a martingale version of the (deterministic) $\dot H^{-1}$-contraction principle for equations of the form (1) with uniform ellipticity (2). In fact, we use a spatially localized version of the $\dot H^{-1}$-contraction principle.

Lemma 2. *Let u denote the stationary solution of (1) and denote by u^h its spatial translation by the shift $h\in\mathbb{R}$. Then we have for $r\ll 1$*

$$\begin{aligned}&\Big\langle\int\eta_r(h)\int_{-\frac{1}{2}}^0\int\eta(u^h-u)^2dxdtdh\Big\rangle_1\\&\lesssim\ r+r^2\Big\langle\int_{-1}^0\int\eta(u-\int_{-1}^0\int\eta u)^2dxdt\Big\rangle_1=r+r^2\langle D^2(u,1)\rangle_1.\end{aligned}$$

Here and in the proof $\lesssim$ and $\ll$ just refer to λ.

The second (very similar) step is to estimate the "bulk" L^2-norm on the r. h. s. of Lemma 2 by the boundary L^2-norm of the initial data $u(t=-1,\cdot)$.

Lemma 3. *The stationary solution u of (1) satisfies*

$$\begin{aligned}\langle D^2(u,1)\rangle_1&=\Big\langle\int_{-1}^0\int\eta(u-\int_{-1}^0\int\eta u)^2dxdt\Big\rangle_1\\&\lesssim\ 1+\int\eta(u-\int\eta u)^2dx_{|t=-1}=1+D'^2(u,1).\end{aligned}\tag{18}$$

Here and in the proof $\lesssim$ and $\ll$ just refer to λ.

The third step is to upgrade the purely spatial L^2-averaged Hölder-$\frac{1}{2}$ modulus of continuity into a space-*time* modulus of continuity.

Lemma 4. *The stationary solution u of (1) satisfies for $r \ll 1$*

$$\Big\langle\Big(\fint_{-r^2}^{0}\big(\int\eta_r u-\fint_{-r^2}^{0}\int\eta_r u\big)^2dxdt\Big)^{\frac{1}{2}}\Big\rangle_1$$
$$\lesssim\ r^{\frac{1}{2}}+\Big\langle\Big(\fint_{-r^2}^{0}\int\eta_r\big(u-\int\eta_r u\big)^2udxdt\Big)^{\frac{1}{2}}\Big\rangle_1.$$

Here and in the proof $\lesssim$ and $\ll$ just refer to λ.

The crucial ingredient for Proposition 1 is the passage from measuring the Hölder-α L^2-modulus of continuity on scales 1 down to scales r. It is here that we need the deterministic ingredients of Propositions 4 and 3. Not surprisingly, we will need in this argument that solutions g to the stochastic *linear constant coefficient* parabolic equation, around which we perturb, have this localization property. This is provided by the following localized space-time *supremum* estimate of the Hölder-α modulus of continuity of g.

Lemma 5. *Let g be the solution of*

$$\partial_t g-a_0\partial_x^2 g=\xi\ \ \text{for}\ t>-1,\quad g(t=-1,\cdot)=0$$

for some constant coefficient $a_0\in[\lambda,1]$. Then for any Hölder exponent $\alpha<\frac{1}{2}$ and all shifts $h\in\mathbb{R}$

$$\Big\langle\sup_{(-1,0)\times\mathbb{R}}\eta(g^h-g)^2\Big\rangle_1\lesssim\min\{|h|^{2\alpha},1\}.$$

Here and in the proof $\lesssim$ and $\ll$ refer to λ and α.

3 Proofs

We start the string of proofs with Lemma 1, since it contains the other arguments *in nuce*.

Proof of Lemma 1. We will establish the lemma in the stronger version where instead of $D^2(u,1)$, we control the Gaussian moments of $E^2(u,1):=\int_{-1}^{0}\int\eta u^2dxdt\geq D^2(u,1)$:

$$\Big\langle\exp\big(\frac{1}{C}E^2(u)\big)\Big\rangle\lesssim 1.$$

By concentration of measure, cf. beginning of Section 2, this is a consequence of the bound on the expectation

$$\langle E^2(u,1)\rangle \lesssim 1 \tag{19}$$

and the uniform bound on the carré-du-champs of the Malliavin derivative

$$|\nabla E(u,1)|^2 \lesssim 1. \tag{20}$$

In order gain confidence in this principle of concentration of measure, let us relate it to the discrete case, that is, the case of countably many independent normal Gaussian random variables, see for instance [8, p.135] for a proof of concentration of measure by an efficient and short semi-group argument. In order to make the connection, let us divide space-time into squares Q of side-length h (no parabolic scaling needed here), which we think of being small. Assume that we are dealing with a function F of the space-time white noise ξ that depends on ξ only through the average of ξ on the cubes Q; which amounts to saying that F only depends on $\{\xi_Q\}_Q$, where $\xi_Q := \frac{1}{h}\int \xi dxdt$ (any reasonable function F can be approximated by such functions F_h for $h \downarrow 0$). The reason for using this normalization by the *square-root* of the space-time volume h^2 is that the application $\xi \mapsto \{\xi_Q\}_Q$ pushes the space-time white-noise ensemble $\langle\cdot\rangle$ into the normal Gaussian ensemble $\langle\cdot\rangle_h$. In particular $\langle F\rangle = \langle F\rangle_h$ and $\langle \exp(\frac{1}{C}F^2)\rangle = \langle \exp(\frac{1}{C}F^2)\rangle_h$. Hence by the discrete theory, we have concentration of measure provided we have a uniform bound on the squared Euclidean (rather Hilbertian) norm $|\nabla_h F|^2 := \sum_Q (\frac{\partial F}{\partial \xi_Q})^2$ of the (infinite-dimensional) vector of partial derivatives. Therefore it remains to argue that $|\nabla_h F|^2$ is dominated by the carré-du-champs $|\nabla F|^2$ of the continuum Malliavin derivative. By definition (10) of the latter we have

$$\lim_{\epsilon\downarrow 0}\frac{1}{\epsilon}(F(\xi+\epsilon\delta\xi)-F(\xi)) \le |\nabla F|\Big(\int(\delta\xi)^2 dxdt\Big)^{\frac{1}{2}} \tag{21}$$

for any field $\delta\xi$, hence in particular for a field $\delta\xi$ which is piecewise constant on the cubes. More precisely, we may assume that $\delta\xi$ is of the form $\delta\xi_{|Q} = \frac{1}{h}\delta\xi_Q$ for some $\{\delta\xi_Q\}_Q$ so that $\frac{1}{h}\int_Q \delta\xi dxdt = \delta\xi_Q$. Because of this normalization, the l. h. s. of (21) turns into $\sum_Q \frac{\partial F}{\partial\xi_Q}\delta\xi_Q$ by definition of the partial derivatives, whereas the r. h. s. turns into $(\sum_Q \delta\xi_Q^2)^{\frac{1}{2}}$, so that by the arbitrariness of $\{\delta\xi_Q\}_Q$, (21) indeed implies $|\nabla_h F| \le |\nabla F|$ (in fact, there is equality).

We start with the first half of the proof, that is, the bound (19) on the expectation. In fact, we shall establish that

$$\langle E^2(u,R)\rangle \lesssim 1,$$

provided the scale $R \sim 1$ is sufficiently large (larger than a constant only depending on λ). This indeed implies (19) since by definition of the average $\int_{-R^2}^{0} \int \eta_R \cdot dxdt$, $E^2(u,1) \leq R^3 E^2(u,R)$, where the power three represents the parabolic dimension. By the scale invariance (4) & (5), we might as well show

$$\langle E^2(u,1) \rangle \lesssim 1, \tag{22}$$

provided the massive term is sufficiently strong, that is, $T \sim 1$ is sufficiently small. In fact, it will be convenient for the upcoming calculation to replace the exponential cut-off η by $\tilde{\eta}^2$, where $\tilde{\eta}$ is a smoothened version of η_2, to fix ideas

$$\tilde{\eta}(x) := \exp(-\frac{1}{2}\sqrt{x^2+1}). \tag{23}$$

In order to establish (22), we will use a martingale argument based on the stochastic (partial) differential equation with (nonlinear) damping

$$\partial_t u = -(\frac{1}{T}u + (-\partial_x^2)\pi(u)) + \xi. \tag{24}$$

As is constitutive for a martingale argument, we shall monitor a symmetric and semi-definite expression, in our case $\int \tilde{\eta} u (1-\partial_x^2)^{-1} \tilde{\eta} u dx$, where we use physicist's notation in the sense that an operator, here $(1-\partial_x^2)^{-1}$, acts on everything to its right, here the product $\tilde{\eta} u$. This quadratic expression, which amounts to a version of the $\dot{H}^{-1}$-norm that is localized (thanks to the inclusion of $\tilde{\eta}$) and endowed with an infra-red cut-off (the effect of the 1 in $(1-\partial_x^2)^{-1}$), is motivated by the $\dot{H}^{-1}$ contraction principle, a well-known property of the deterministic versions of (1); in this language, we monitor here the (modified) $\dot{H}^{-1}$ distance to the trivial solution $u = 0$. In general terms, the time derivative of such quadratic expression under a stochastic equation comes in three contributions: the contribution solely of the deterministic r. h. s. of (24), the contribution solely from the stochastic r. h. s. ξ, and a mixed contribution. In this set-up, the space-time white noise ξ is viewed as a white noise in time with a spatial (and thus infinite-dimensional) covariance structure expressing white noise in space. The mixed contribution is a martingale, and thus vanishes when taking the expectation: This cancellation can best be understood when considering a time discretization of (24) that is explicit in the drift $-(\frac{1}{T}u + (-\partial_x^2)\pi(u))$ (of course, an explicit time discretization is not well-posed for an infinite dimensional dynamical system coming from a *parabolic* equation, so one better combines it in one's mind with a spatial discretization). The contribution which solely comes from ξ is the so-called quadratic variation, and its expectation can be computed based on the operator defining the quadratic expression, here $\tilde{\eta}(1-\partial_x^2)^{-1}\tilde{\eta}$, and the spatial covariance structure of the noise (provided it is white in time). Since the spatial covariance structure is the one coming from (spatial) white noise, it is given by the integral of the diagonal of the kernel (i. e. the trace-norm of the operator). In case of $\tilde\eta(1-\partial_x^2)^{-1}\tilde\eta$, the kernel is given by $\tilde\eta(x)\frac{1}{2}\exp(-|x-y|)\tilde\eta(y)$. Hence the expectation of the quadratic variation is given by $\int\frac{1}{2}\tilde\eta^2dx$. Altogether, the martingale argument thus yields

$$\begin{aligned}&\frac{d}{dt}\frac{1}{2}\Big\langle\int\tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta u dx\Big\rangle\\ &= -\Big\langle\int\tilde\eta u(1-\partial_x^2)^{-1}\big(\frac{1}{T}\tilde\eta u+\tilde\eta(-\partial_x^2)\pi(u)\big)dx\Big\rangle+\frac{1}{2}\int\frac{1}{2}\tilde\eta^2dx. \quad (25)\end{aligned}$$

We rewrite this identity as

$$\begin{aligned}&\frac{d}{dt}\exp(\frac{t}{T})\Big\langle\int\tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta u dx\Big\rangle\\ &= \exp(\frac{t}{T})\Big(-\Big\langle\int\tilde\eta u(1-\partial_x^2)^{-1}\big(\frac{1}{T}\tilde\eta u+2\tilde\eta(-\partial_x^2)\pi(u)\big)dx\Big\rangle+\int\frac{1}{2}\tilde\eta^2dx\Big),\end{aligned}$$

and integrate over $t\in(-\infty,0)$:

$$\Big\langle\int_{-\infty}^0\exp(\frac{t}{T})\int\tilde\eta u(1-\partial_x^2)^{-1}\big(\frac{1}{T}\tilde\eta u+2\tilde\eta(-\partial_x^2)\pi(u)\big)dxdt\Big\rangle\le\frac{T}{2}\int\tilde\eta^2dx.$$

Hence in order to arrive at (22), it is enough to show that for $T\ll 1$, we have the deterministic estimate

$$\int\tilde\eta u(1-\partial_x^2)^{-1}\big(\frac{1}{T}\tilde\eta u+2\tilde\eta(-\partial_x^2)\pi(u)\big)dx\gtrsim\int\tilde\eta^2u^2dx. \quad (26)$$

We have a closer look at the elliptic term $\tilde\eta(-\partial_x^2)\pi(u)$ in (26), whose contribution would be positive by the monotonicity of π if it weren't for the spatial cut-off and the infra-red cut off. Using Leibniz' rule, we rewrite it as (in our physicist's way of omitting parentheses)

$$\tilde\eta(-\partial_x^2)\pi(u)=(1-\partial_x^2)\pi(u)\tilde\eta+2\partial_x\pi(u)\partial_x\tilde\eta-\pi(u)(\tilde\eta-\partial_x^2\tilde\eta), \quad (27)$$

where we w. l. o. g. assume that $\pi(0)=0$. Hence by the symmetry of $(1-\partial_x^2)^{-1}$ we obtain

$$\begin{aligned}&\int\tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)\pi(u)dx\\ &= \int\tilde\eta^2u\pi(u)dx-2\int(\partial_x\tilde\eta)\pi(u)\partial_x(1-\partial_x^2)^{-1}\tilde\eta u dx\\ &\quad-\int(\tilde\eta-\partial_x^2\tilde\eta)\pi(u)(1-\partial_x^2)^{-1}\tilde\eta u dx.\end{aligned}$$

Using that the operators $\partial_x(1-\partial_x^2)^{-\frac{1}{2}}$ and $(1-\partial_x^2)^{-\frac{1}{2}}$ have operator norm 1 w. r. t. to L^2, we deduce the inequality

$$\begin{aligned}\int \tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)\pi(u)dx \;\geq\; & \int \tilde\eta^2 u\pi(u)dx \\ & -\Big(2\big(\int(\partial_x\tilde\eta)^2\pi^2(u)dx\big)^{\frac{1}{2}}+\big(\int(\tilde\eta-\partial_x^2\tilde\eta)^2\pi^2(u)dx\big)^{\frac{1}{2}}\Big) \\ & \times\Big(\int \tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta u dx\Big)^{\frac{1}{2}}. \end{aligned} \tag{28}$$

By the monotonicity properties (2) of π and our gratuitous assumption $\pi(0)=0$, this yields

$$\begin{aligned}&\int \tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)\pi(u)dx \\ &\geq\; \lambda\int\tilde\eta^2u^2dx-\Big(2\big(\int(\partial_x\tilde\eta)^2u^2dx\big)^{\frac{1}{2}}+\big(\int(\tilde\eta-\partial_x^2\tilde\eta)^2u^2dx\big)^{\frac{1}{2}}\Big) \\ &\quad\times\Big(\int\tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta u dx\Big)^{\frac{1}{2}}.\end{aligned}$$

Our smoothing out of the exponential cut-off function, cf. (23), has the sole purpose of making sure that

$$|\partial_x\tilde\eta|+|\partial_x^2\tilde\eta|\lesssim\tilde\eta, \tag{29}$$

so that we obtain by Young's inequality for the elliptic term,

$$\int \tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)\pi(u)dx \;\geq\; \frac{1}{C}\int\tilde\eta^2u^2dx - C\int\tilde\eta u(1-\partial_x^2)^{-1}\tilde\eta u dx.$$

We thus see that thanks to the massive term, (26) holds for $T\ll 1$.

We now turn to the second half of the proof, the estimate of the carré-du-champs (20). We first argue that (20) follows from the deterministic estimate

$$E^2(\delta u,1)\lesssim\int(\delta\xi)^2dxdt, \tag{30}$$

where δu and $\delta\xi$ are related via

$$\delta u+\partial_t\delta u-\partial_x^2(a\delta u)=\delta\xi \tag{31}$$

with $a = \pi'(u)$. Indeed, we note that by duality w. r. t. to the inner product $(g, f) \mapsto \int_{-1}^{0} \int \eta g f dx dt$,

$$E(u,1) = \sup \Big\{ E(u,f) := \int_{-1}^{0} \int \eta u f dx dt \,\Big|\, \int_{-1}^{0} \int \eta f^2 dx dt = 1, \ \ \mathrm{supp} f \subset (-1,0) \times \mathbb{R} \Big\}. \tag{32}$$

By the chain rule for the Malliavin derivative we thus obtain

$$|\nabla E(\cdot,1)| \le \sup_f |\nabla E(\cdot,f)|,$$

where the supremum runs over the set implicitly defined in (32), so that it is enough to show for a fixed f

$$|\nabla E(u,f)|^2 \lesssim 1.$$

By definition of the carré-du-champs of the Malliavin derivative in case of the *linear* functional $u \mapsto E(u,f)$, cf. (10), this amounts to showing

$$\int_{-1}^{0} \int \eta \delta u f dx dt \lesssim 1,$$

where the infinitesimal perturbation δu of the solution is related to the infinitesimal perturbation $\delta\xi$ of the noise via (31). By the characterizing properties of the f's, cf. (32), this estimate in turn amounts to establishing (30).

We now turn to the proof of the deterministic estimate (30). To ease notation and make the connection to Proposition 3, we rephrase (and strengthen) the goal: For w and f related via

$$w + \partial_t w - \partial_x^2 (a w) = f, \tag{33}$$

with uniformly elliptic coefficient field a in the sense of (11), we seek the estimate

$$\int_{-\infty}^{0} \int w^2 dx dt \lesssim \int_{-\infty}^{0} \int f^2 dx dt. \tag{34}$$

Like for (22), our Ansatz for (34) is motivated by the $\dot{H}^{-1}$-contraction principle. Again, we consider a version of $\dot{H}^{-1}$-norm with ultra-red cut-off, but this time without cut-off function η, namely $\int w(L^{-2} - \partial_x^2)^{-1} w dx$, where the length scale L for the ultra-red cut-off will be chosen later. We obtain from the equation (34)

$$\begin{aligned}
&\frac{d}{dt}\frac{1}{2}\int w(L^{-2}-\partial_x^2)^{-1}w dx\\
&= -\int w(L^{-2}-\partial_x^2)^{-1}(w-f+(-\partial_x^2)(aw))dx\\
&= -\int w(L^{-2}-\partial_x^2)^{-1}(w-f-L^{-2}aw)dx-\int aw^2dx.
\end{aligned}$$

We apply Cauchy-Schwarz' inequality and use the uniform ellipticity of a, cf. (11), to obtain the estimate

$$\begin{aligned}
&\frac{d}{dt}\frac{1}{2}\int w(L^{-2}-\partial_x^2)^{-1}w dx\\
&\leq -\int w(L^{-2}-\partial_x^2)^{-1}w dx-\lambda\int w^2dx\\
&\quad +\Big(\int((L^{-2}-\partial_x^2)^{-1}w)^2dx\Big)^{\frac{1}{2}}\Big(\big(\int f^2dx\big)^{\frac{1}{2}}+L^{-2}\big(\int w^2dx\big)^{\frac{1}{2}}\Big).
\end{aligned}$$

Thanks to the operator inequality $(L^{-2}-\partial_x^2)^{-1}\leq L(L^{-2}-\partial_x^2)^{-\frac{1}{2}}$ we have

$$\Big(\int((L^{-2}-\partial_x^2)^{-1}w)^2dx\Big)^{\frac{1}{2}}\leq L\Big(\int w(L^{-2}-\partial_x^2)^{-1}w dx\Big)^{\frac{1}{2}},$$

so that we may absorb the term $(\int((L^{-2}-\partial_x^2)^{-1}w)^2dx)^{\frac{1}{2}}L^{-2}(\int w^2dx)^{\frac{1}{2}}$ by Young's inequality for $L\gg 1$, obtaining

$$\frac{d}{dt}\int w(L^{-2}-\partial_x^2)^{-1}w dx \;\leq\; -\frac{1}{C}\int w^2dx+CL^2\int f^2dx.$$

Integration in time yields (34).

Proof of Proposition 3. We first note that (15) follow easily from (14): Indeed, by translation invariance (14) also holds with η replaced by the shift η^y, summation over $y\in\mathbb{Z}$ gives (15). We next note that w. l. o. g. we may assume $h=0$, since we may rewrite (12) as $\partial_t(w-h)-\partial_x^2(a(w-h))=\partial_x^2(g+ah)$. The proof of this Proposition is very close to the deterministic part of the proof of Lemma 1; in fact, it might be seen as an infinitesimal version of it. Like there, we substitute η by $\tilde\eta^2$, cf. (23), and start from monitoring the localized H^{-1}-norm of w with infra-red cut-off:

$$\frac{d}{dt}\frac{1}{2}\int\tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta w dx=-\int\tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x)^2(aw+g)dx.$$

As in (27), we write

$$\tilde\eta(-\partial_x^2)(aw+g) = (1-\partial_x^2)(aw+g)\tilde\eta + 2\partial_x(aw+g)\partial_x\tilde\eta - (aw+g)(\tilde\eta - \partial_x^2\tilde\eta),$$

which yields

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\int \tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta w dx \;&=\; -\int \tilde\eta w \tilde\eta(aw+g)dx \\ &\quad +2\int \tilde\eta w(1-\partial_x^2)^{-1}\partial_x(aw+g)\partial_x\tilde\eta dx \\ &\quad -\int \tilde\eta w(1-\partial_x^2)^{-1}(aw+g)(\tilde\eta-\partial_x^2\tilde\eta)dx.\end{aligned}$$

Using symmetry and boundedness properties of $(1-\partial_x^2)^{-1}$, and the estimates (29) on our mollified exponential cut-off $\tilde\eta$, the two last terms are estimated as

$$\begin{aligned}&\int \tilde\eta w(1-\partial_x^2)^{-1}\partial_x(aw+g)\partial_x\tilde\eta dx \\ &\quad\lesssim\; \Big(\int \tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta w dx \int \tilde\eta^2(aw+g)^2 dx\Big)^{\frac{1}{2}}\end{aligned}$$

and

$$\begin{aligned}&-\int \tilde\eta w(1-\partial_x^2)^{-1}(aw+g)(\tilde\eta-\partial_x^2\tilde\eta)dx \\ &\quad\lesssim\; \Big(\int \tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta w dx \int \tilde\eta^2(aw+g)^2 dx\Big)^{\frac{1}{2}}.\end{aligned}$$

Hence we obtain by the uniform ellipticity (11) of a together with the triangle inequality to break up $aw+g$ and Young's inequality

$$\begin{aligned}&\frac{d}{dt}\int \tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta w dx + \frac{1}{C}\int \tilde\eta^2 w^2 dx \\ &\quad\le\; C\Big(\int \tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta w dx + \int \tilde\eta^2 g^2 dx\Big),\end{aligned}$$

which we rewrite as

$$\begin{aligned}&\frac{d}{dt}\exp(-Ct)\int \tilde\eta w(1-\partial_x^2)^{-1}\tilde\eta w dx \\ &\quad+\frac{1}{C}\exp(-Ct)\int \tilde\eta^2 w^2 dx \;\le\; C\exp(-Ct)\int \tilde\eta^2 g^2 dx.\end{aligned}$$

Appealing to the initial condition $w(t=-1)=0$ yields the desired

$$\int_{-1}^{0}\int \tilde\eta^2 w^2 dxdt \ \lesssim \ \int_{-1}^{0}\int \tilde\eta^2 g^2 dxdt.$$

Proof of Proposition 4. We start by observing

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\int \eta(\partial_x v)^2 dx &= -\int \partial_x(\eta\partial_x v)(a_0\partial_x^2 v + f)dx\\ &= -\int \eta(a_0(\partial_x^2 v)^2 + f\partial_x^2 v)dx - \int \partial_x\eta\partial_x v(a_0\partial_x^2 v + f)dx,\end{aligned}$$

so that because of $a_0\in[\lambda,1]$ and $|\partial_x\eta|\le\eta$ we obtain by Young's inequality

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\int \eta(\partial_x v)^2 dx &\le -\int \eta(\lambda(\partial_x^2 v)^2 + f\partial_x^2 v)dx + \int \eta|\partial_x v|(|\partial_x^2 v| + |f|)dx\\ &\le -\frac{1}{C}\int \eta(\partial_x^2 v)^2 dx + C\int \eta((\partial_x v)^2 + f^2)dx. \qquad (35)\end{aligned}$$

Dropping the good r. h. s. term, we rewrite this as

$$\frac{d}{dt}(t+1)\int \eta(\partial_x v)^2 dx \ \lesssim \ \int \eta((\partial_x v)^2 + f^2)dx,$$

so that we obtain from integration in $t\in(-1,0)$

$$\sup_{t\in(-1,0)}(t+1)\int \eta(\partial_x v)^2 dx \lesssim \int_{-1}^{0}\int \eta((\partial_x v)^2 + f^2)dxdt. \qquad (36)$$

Thanks to the constant coefficients, also the (localized) L^2-norm is well-behaved. Indeed, from (13) we obtain

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\int \eta v^2 dx &= \int \eta v(a_0\partial_x^2 v + f)dx\\ &= \int \eta(-a_0(\partial_x v)^2 + vf)dx - a_0\int \partial_x\eta v\partial_x v dx,\end{aligned}$$

so that because of $a_0\in[\lambda,1]$ and $|\partial_x\eta|\le\eta$, we obtain by Young's inequality

$$\frac{d}{dt}\int \eta v^2 dx \ \le \ -\frac{1}{C}\int \eta(\partial_x v)^2 dx + \int \eta(v^2 + f^2)dx.$$

From the integration in t of this differential inequality for $\int \eta v^2 dx$ we learn

$$\sup_{t\in(-1,0)}\int \eta v^2 dx + \int_{-1}^{0}\int \eta(\partial_x v)^2 dxdt \lesssim \int_{-1}^{0}\int \eta f^2 dxdt + \int \eta v^2 dx_{|t=-1}. \qquad (37)$$

The combination of this with (36) yields

$$\sup_{t\in(-1,0)}\big((1+t)\int\eta(\partial_x v)^2+\int\eta v^2dx\big)\lesssim\int_{-1}^0\int\eta f^2dxdt+\int\eta v^2dx_{|t=-1}. \quad (38)$$

In view of this a priori estimate (38), the first part (16) of this proposition follows from the embedding

$$\sup_{(t,x)\in(-1,0)\times\mathbb{R}}(1+t)^{\frac{1}{2}}\eta v^2\lesssim\sup_{t\in(-1,0)}\big((1+t)\int\eta(\partial_x v)^2dx+\int\eta v^2dx\big),$$

which easily is is seen to hold: Because of

$$\sup_x\eta v^2\lesssim\int|\partial_x(\eta v^2)|dx\lesssim\int(\eta|v\partial_x v|+|\partial_x\eta|v^2)dx\lesssim\int\eta(|\partial_x v|+|v|)|v|dx,$$

we obtain by Young's inequality for $t\in(-1,0)$

$$\begin{aligned}(t+1)^{\frac{1}{2}}\sup_x\eta v^2 &\lesssim (1+t)\int\eta((\partial_x v)^2+v^2)dx+\int\eta v^2dx\\ &\lesssim (1+t)\int\eta(\partial_x v)^2dx+\int\eta v^2dx.\end{aligned}$$

We now turn to to the second part (17) of the proposition. As for the first part, it is the consequence of an a priori estimate and an embedding. For the a priori estimate, we rewrite (35) as

$$\begin{aligned}&\frac{d}{dt}\exp(-Ct)\int\eta(\partial_x v)^2dx\\ &\leq -\frac{1}{C}\exp(-Ct)\int\eta(\partial_x^2 v)^2dx+C\exp(-Ct)\int\eta f^2dx,\end{aligned}$$

which in view of the vanishing initial data yields

$$\int_{-1}^0\int\eta((\partial_x^2 v)^2+(\partial_x v)^2)dxdt\lesssim\int_{-1}^0\int\eta f^2dxdt.$$

Thanks to the equation (13) and to (37), this can be upgraded to contain the time-derivative term and the zero-order term

$$\int_{-1}^0\int\eta((\partial_t v)^2+(\partial_x^2 v)^2+(\partial_x v)^2+v^2)dxdt\lesssim\int_{-1}^0\int\eta f^2dxdt. \quad (39)$$

In view of this a priori estimate (39), in order to establish (17), it remains to show the embedding

$$\begin{aligned}
&\fint_{-r^2}^{0}\int \eta_r(v-\fint_{-r^2}^{0}\int \eta_r v)^2 dxdt\\
&\quad\lesssim\ r\int_{-1}^{0}\int \eta((\partial_t v)^2+(\partial_x^2 v)^2+(\partial_x v)^2+v^2)dxdt, \qquad (40)
\end{aligned}$$

where thanks to the term ηv^2-term included in the r. h. s., this inequality is only non-trivial for $r \ll 1$. We split this estimate into the spatial modulus of continuity and the temporal modulus of continuity:

$$\begin{aligned}
&\fint_{-r^2}^{0}\int \eta_r(v-\fint_{-r^2}^{0}\int \eta_r v)^2 dxdt\\
&\quad\le\ \sup_{t\in(-r^2,0)}\int \eta_r(v-\int \eta_r v)^2 dx+\fint_{-r^2}^{0}(\int \eta_r v-\fint_{-r^2}^{0}\int \eta_r v)^2 dt.
\end{aligned}$$

The estimate on the spatial modulus of continuity follows from the combination of the estimate

$$\int \eta_r(v-\int \eta_r v)^2 dx \lesssim r^2\int \eta_{2r}(\partial_x v)^2 dx \lesssim r\int \eta(\partial_x v)^2 dx, \qquad (41)$$

where time is just a parameter, with

$$\sup_{t\in(-1,0)}\int \eta(\partial_x v)^2 dx \lesssim \int_{-1}^{0}\int \eta((\partial_t v)^2+(\partial_x^2 v)^2+(\partial_x v)^2)dxdt. \qquad (42)$$

For temporal continuity, we need

$$\begin{aligned}
\fint_{-r^2}^{0}(\int \eta_r v-\fint_{-r^2}^{0}\int \eta_r v)^2 dt\ &\lesssim\ r^4\fint_{-r^2}^{0}\int \eta_r(\partial_t v)^2 dxdt\\
&\lesssim\ r\int_{1}^{0}\int \eta(\partial_t v)^2 dxdt. \qquad (43)
\end{aligned}$$

Here comes the argument for the first inequality in (41) (the second one follows from the pointwise estimate on the weights $\eta_R \le \frac{1}{R}\eta$ for $R \le 1$): By scaling, we may assume w. l. o. g. that $r=1$, so that the we need to show

$$\int \eta(v-\int \eta v)^2 dx \lesssim \int \eta_2(\partial_x v)^2 dx,$$

which can be seen by rewriting the l. h. s. with help of Jensen's inequality

$$
\begin{aligned}
&\int \eta(x)\big(v(x) - \int \eta(y)v(y)dy\big)^2 dx \\
&\le \int\int \eta(x)\eta(y)(v(x)-v(y))^2 dydx \\
&\le \int\int\int_0^1 \eta(x)\eta(y)(\partial_x v)^2(\sigma x + (1-\sigma)y)d\sigma dydx \\
&= 2\int_0^{\frac12}\int\int \eta(x)\eta(y)(\partial_x v)^2(\sigma x+(1-\sigma)y)dydxd\sigma \\
&= 2\int_0^{\frac12}\frac{1}{1-\sigma}\int\int \eta(x)\eta(\tfrac{z-\sigma x}{1-\sigma})(\partial_x v)^2(z)dzdxd\sigma \\
&\le 2\sup_{\sigma\in(0,\frac12)}\int\int \eta(x)\eta(\tfrac{z-\sigma x}{1-\sigma})(\partial_x v)^2(z)dzdx,
\end{aligned}
$$

and then using the special properties of our averaging function in form of $\eta(x)\eta(\frac{z-\sigma x}{1-\sigma}) = \frac14\exp(-(|x|+\frac{|z-\sigma x|}{1-\sigma})) \le \frac14\exp(-\frac12(|x|+|z|)) = \exp(-\frac12|x|)\eta_2(z)$.
Now for the argument for (42): We start with the elementary inequality

$$
\sup_{t\in(-1,0)}\int \eta(\partial_x v)^2 dx \;\le\; \int_{-1}^0 \Big|\frac{d}{dt}\int \eta(\partial_x v)^2dx\Big|dt + \int_{-1}^0\int \eta(\partial_x v)^2dxdt
$$

and then note that by integration by parts, $|\partial_x\eta|\le\eta$, and Young's inequality

$$
\begin{aligned}
\int_{-1}^0 \Big|\frac{d}{dt}\int \eta(\partial_x v)^2dx\Big|dt &= 2\int_{-1}^0\Big|\int \eta\partial_x v\partial_t\partial_x v dx\Big|dt \\
&\lesssim \int_{-1}^0\int (\eta|\partial_x^2 v| + |\partial_x\eta||\partial_x v|)|\partial_t v|dxdt \\
&\lesssim \int_{-1}^0\int \eta((\partial_x^2 v)^2 + (\partial_x v)^2 + (\partial_t v)^2)dxdt.
\end{aligned}
$$

We finally turn to (43). The second estimate follows from $\fint_{-r^2}\int\eta_r\cdot dxdt \le \frac{1}{r^3}\fint_{-1}^0\int\eta\cdot dxdt$. By parabolic scaling, we may restrict ourselves to $r=1$ for the first estimate, which then takes the form

$$
\int_{-1}^0\Big(\int\eta v - \fint_{-1}^0\int\eta v\Big)^2dt \;\lesssim\; \int_{-1}^0\int\eta(\partial_t v)^2dxdt,
$$

and immediately follows from combining the Poincaré inequality with mean value zero

$$
\int_{-1}^0\Big(\int\eta v - \fint_{-1}^0\int\eta v\Big)^2dt \;\lesssim\; \int_{-1}^0\big(\partial_t\int\eta v dx\big)^2dt
$$

with Jensen's inequality $\big(\partial_t \int \eta v dx\big)^2 \le \int \eta(\partial_t v)^2 dx$.

PROOF OF PROPOSITION 1. For conciseness, we ignore the massive term in (1). The main object of this proposition is $\delta u := u^h - u$, where $u^h(t,x) = u(t,x+h)$ denotes a spatial shift of the stationary solution of (1). We note that δu satisfies the formally linear equation

$$\partial_t \delta u - \partial_x^2(a_h \delta u) = (\partial_t - a_0\partial_x^2)\delta g, \tag{44}$$

where we introduced the coefficient field

$$a_h = \int_0^1 \pi'(\sigma u^h + (1-\sigma)u)d\sigma, \tag{45}$$

which by (2) is uniformly elliptic in the sense of (11), and the set $\delta g := g^h - g$, where g is defined via the linear version of (1)

$$\partial_t g - a_0\partial_x^2 g = \xi \quad \text{for} \;\; t \in (-1,0), \qquad g = 0 \;\; \text{for}\; t = -1,$$

cf. Lemma 5, with a constant coefficient $a_0 \in [\lambda, 1]$ to be chosen below.

We start with the main deterministic ingredient for Proposition 1, which we need to go from scales of order one to scales of order $r \ll 1$ in the L^2-averaged Hölder-$\frac{1}{2}$ modulus of continuity. It is given by the estimate

$$\begin{aligned} \fint_{-r^2}^0 \int \eta_r(\delta u)^2 dxdt \;&\lesssim\; \Big(1 + \frac{1}{r^3}\int_{-1}^0 \int \eta(a_h - a_0)^2 dxdt\Big) \\ &\times \;\; \Big(\int_{-1}^0 \int \eta(\delta u)^2 dxdt + \sup_{(t,x)\in(-1,0)\times\mathbb{R}} \eta(\delta g)^2\Big), \end{aligned} \tag{46}$$

which we shall establish for all $r \ll 1$. We observe that it is enough to establish for any $R \in [\frac{1}{2}, 1]$ the estimate

$$\begin{aligned} \fint_{-r^2}^0 \int \eta_r(\delta u)^2 dxdt \;&\lesssim\; \Big(1 + \frac{1}{r^3}\int_{-R^2}^0 \int (t+R^2)^{-\frac{1}{2}}\eta(a_h - a_0)^2 dxdt\Big) \\ &\times \;\; \Big(\int \eta(\delta u)^2 dx_{|t=-R^2} + \sup_{(t,x)\in(-R^2,0)\times\mathbb{R}} (\delta g)^2\Big), \end{aligned}$$

since the the integral of this estimate over $R \in [\frac{1}{2}, 1]$ yields (46), using the integrability of $(t+R^2)^{-\frac{1}{2}}$ thanks to $\frac{1}{2} < 1$. To simplify notation, we replace $R \sim 1$ by unity, so that it remains to show

$$\begin{aligned} \fint_{-r^2}^0 \int \eta_r(\delta u)^2 dxdt \;&\lesssim\; \Big(1 + \frac{1}{r^3}\int_{-1}^0 \int (t+1)^{-\frac{1}{2}}\eta(a_h - a_0)^2 dxdt\Big) \\ &\times \;\; \Big(\int \eta(\delta u)^2 dx_{|t=-1} + \sup_{(t,x)\in(-1,0)\times\mathbb{R}} \eta(\delta g)^2\Big). \end{aligned} \tag{47}$$

To this purpose, we split the solution $\delta u = \delta g + v + w$, where v is defined through the constant-coefficient initial value problem

$$\partial_t v - a_0 \partial_x^2 v = 0 \ \text{ for } t \in (-1,0), \quad v = \delta u \ \text{ for } t = -1,$$

which is made such that $\delta g + v$ agrees with δu for $t=-1$ and satisfies for $t \in (-1,0)$ the following equation

$$\partial_t(\delta g + v) - a_0 \partial_x^2(\delta g + v) = (\partial_t - a_0\partial_x^2)\delta g = \xi, \tag{48}$$

and where w is defined through the initial value problem

$$\partial_t w - \partial_x^2(a_h w) = \partial_x^2((a_h - a_0)(\delta g + v)) \ \text{ for } t \in (-1,0), \quad w = 0 \ \text{ for } t = -1. \tag{49}$$

Taking the sum of (48) and (49), and comparing with (44), we see that this indeed gives $\delta u = \delta g + v + w$.

From the first part (16) of Proposition 4, we learn

$$\sup_{(t,x)\in(-1,0)\times\mathbb{R}} (1+t)^{\frac{1}{2}} \eta v^2 \lesssim \int \eta (\delta u)^2 dx_{|t=-1}, \tag{50}$$

which implies in particular for $r \le \frac{1}{2}$ (which amounts to $r \le \frac{1}{4}$ before setting setting $R=1$ above)

$$⨍_{-r^2}^{0} \int \eta_r v^2 \lesssim \int \eta(\delta u)^2 dx_{|t=-1}. \tag{51}$$

From the first part (14) of Proposition 3 (with η replaced by $\eta_{\frac{1}{2}}$), we gather that

$$\int_{-1}^{0}\int \eta_{\frac{1}{2}} w^2 dxdt \lesssim \int_{-1}^{0}\int \eta_{\frac{1}{2}} (a_h - a_0)^2((\delta g)^2 + v^2) dxdt,$$

which implies for $r \le \frac{1}{2}$ (by the obvious inequality $⨍_{-r^2}^{0}\int \eta_r \cdot dxdt \le (\frac{R}{r})^3 - ⨍_{-R^2}^{0}\int \eta_R \cdot dxdt$ for $r \le R$ and since $\eta_{\frac{1}{2}} \lesssim \eta^2$)

$$\begin{aligned} & r^3 ⨍_{-r^2}^{0}\int \eta_r w^2 dxdt \\ \lesssim & \int_{-1}^{0} (1+t)^{-\frac{1}{2}} \int \eta (a_h - a_0)^2 dxdt \sup_{(t,x)\in(-1,0)\times\mathbb{R}} (1+t)^{\frac{1}{2}} \eta((\delta g)^2 + v^2). \end{aligned} \tag{52}$$

Inserting (50) into (52) yields

$$\begin{aligned} ⨍_{-r^2}^{0}\int \eta_r w^2 dxdt \ &\lesssim\ \frac{1}{r^3}\int_{-1}^{0} (1+t)^{-\frac{1}{2}} \int \eta(a_h - a_0)^2 dxdt \\ &\quad \times \ \Big(\int \eta(\delta u)^2 dx_{|t=-1} + \sup_{(t,x)\in(-1,0)\times\mathbb{R}} \eta(\delta g)^2 \Big). \end{aligned}$$

Finally, because of $\fint_{-r^2}^0\int\frac{\eta_r}{\eta}dxdt\lesssim 1$ for $r\ll 1$, we have

$$\fint_{-r^2}^0\int\eta_r(\delta g)^2dxdt\lesssim\sup_{(t,x)\in(-1,0)\times\mathbb{R}}\eta(\delta g)^2.$$

Combining the two last estimates with (51) yields (47) for $\delta u=\delta g+v+w$.

We now post-process (47) and to that purpose make the choice of $a_0=\pi'(c)$ with $c:=\int_{-1}^0\int\eta udxdt$, so that in view of the definition (45) of a_h and the Lipschitz continuity (3) of π'

$$|a_h-a_0|\le|a_h-\pi'(u)|+|\pi'(u)-\pi'(c)|\le L(|\delta u|+|u-c|).$$

Therefore, (after replacing r by $2r$ in order to make η_{2r} appear, which is no problem thanks to $r\ll 1$) (46) turns into

$$\begin{aligned}&\fint_{-r^2}^0\int\eta_{2r}(u^h-u)^2dxdt\\ \lesssim\ &\Big(1+\frac{L^2}{r^3}\int_{-1}^0\int\eta(u^h-u)^2dxdt+\frac{L^2}{r^3}D^2(u,1)\Big)\\ \times\ &\Big(\int_{-\frac12}^0\int\eta(u^h-u)^2dxdt+\sup_{(t,x)\in(-1,0)\times\mathbb{R}}\eta(g^h-g)^2\Big).\end{aligned}\tag{53}$$

We now will integrate in h according to $\int\eta_{2r}(h)\cdot dh$. As we shall argue below, we have for the l. h. s. of (53)

$$\int\eta_{2r}(h)\fint_{-r^2}^0\int\eta_{2r}(u^h-u)^2dxdtdh\gtrsim\fint_{-r^2}^0\int\eta_r(u-\int\eta_r u)^2dxdt.\tag{54}$$

The r. h. s. of (53) comes in form of a product of two h-dependent functions we momentarily call $f_1(h)$ and $f_2(h)$. To this purpose we use that thanks to $4r\le 1$ we have $\eta_{2r}\lesssim\eta\eta_{4r}$ for our exponential cut-off so that $\int\eta_{2r}f_1f_2dh\lesssim\sup_h(\eta f_1)\int\eta_{4r}f_2dh$. We claim that for the first factor on the r. h. s. of (53) we have

$$\sup_h\eta(h)\int_{-1}^0\int\eta(u^h-u)^2dxdt\lesssim\int_{-1}^0\int\eta(u-\int_{-1}^0\int\eta u)^2dxdt=D(u,1).\tag{55}$$

Before inserting them, we give the easy arguments for (54) and (55): By scaling we may assume $r=1$ so that (54) follows from Jensen's inequality in form of

$$\int\eta(u-\int\eta u)^2dx\le\int\int\eta(x)\eta(x+h)(u^h(x)-u(x))^2dxdh$$

and the fact that for our exponential cut-off $\eta(x)\eta(x+h) = \frac{1}{2}\exp(-(|x| + |x+h|)) \le \frac{1}{2}\exp(-\frac{1}{2}(|h|+|x|)) = 4\eta_2(h)\eta_2(x)$. For (55), by the triangle inequality in L^2, it is enough to show for a constant c ($\int_{-1}^{0}\int \eta u dx dt$ in our case)

$$\sup_h \eta(h)\int \eta(u^h - c)^2 dx \le \int \eta(u-c)^2 dx. \tag{56}$$

This inequality follows from writing

$$\sup_h \eta(h)\int \eta(u^h - c)^2 dx = \sup_h \int \eta(h)\eta(x-h)(u(x)-c)^2 dx$$

and the fact that for our exponential cut-off $\eta(h)\eta(x-h) \le \eta(x)$. Inserting (54) and (55) into (53) we obtain

$$\begin{aligned} &\fint_{-r^2}^{0}\int \eta_r(u - \int \eta_r u)^2 dx dt \lesssim \Big(1 + \frac{L^2}{r^3}D^2(u,1)\Big) \\ &\quad\times \int \eta_{4r}(h)\Big(\int_{-\frac{1}{2}}^{0}\int \eta(u^h - u)^2 dx dt + \sup_{(t,x)\in(-1,0)\times\mathbb{R}} \eta(g^h - g)^2\Big) dh. \end{aligned} \tag{57}$$

Finally, we come to the probabilistic part of the proof: We take the (restricted) expectation of (the square root of) (57) and use Lemmas 2 and 5 on the two terms of the last factor

$$\begin{aligned} &\Big\langle\Big(\fint_{-r^2}^{0}\int \eta_r(u - \int \eta_r u)^2 dx dt\Big)^{\frac{1}{2}}\Big\rangle_1 \\ &\quad\lesssim \Big(1 + \frac{L}{r^{\frac{3}{2}}}\langle D^2(u,1)\rangle_1^{\frac{1}{2}}\Big)\Big(r^{\frac{1}{2}} + r\langle D^2(u,1)\rangle_1^{\frac{1}{2}} + r^{\alpha}\Big). \end{aligned}$$

We now appeal to the triangle inequality in form of

$$\begin{aligned} D(u,r) \le &\Big(\fint_{-r^2}^{0}\big(\int \eta_r u - \fint_{-r^2}^{0}\int \eta_r u\big)^2 dt\Big)^{\frac{1}{2}} \\ &+\Big(\fint_{-r^2}^{0}\int \eta_r(u - \int \eta_r u)^2 dx dt\Big)^{\frac{1}{2}} \end{aligned}$$

and Lemma 4 for the upgrade to

$$\langle D(u,r)\rangle_1 \lesssim r^{\frac{1}{2}} + \Big(1 + \frac{L}{r^{\frac{3}{2}}}\langle D^2(u,1)\rangle_1^{\frac{1}{2}}\Big)\Big(r^{\frac{1}{2}} + r\langle D^2(u,1)\rangle_1^{\frac{1}{2}} + r^{\alpha}\Big),$$

which we rewrite as

$$\langle D(u,r)\rangle_1 \lesssim r^{\alpha}\Big(1 + \frac{L}{r^{\frac{3}{2}}}\langle D^2(u,1)\rangle_1^{\frac{1}{2}}\Big)\Big(1 + r^{\frac{1}{2}}\langle D^2(u,1)\rangle_1^{\frac{1}{2}}\Big). \tag{58}$$

In this form, we see that (58) does not just hold for $r \ll 1$ but trivially for $r \le 1$ with $r \sim 1$, since $D(u,r) \le \frac{1}{r^3} D(u,1)$. It remains to appeal to Lemma 3.

Proof of Proposition 2. For conciseness, we ignore the massive term in (1) and fix $r \le 1$. Following the argument in the proof of Lemma 1, we first claim that the proposition reduces to the following deterministic estimate

$$D^2(\delta u, r) \lesssim \big(r + \frac{L^2}{r^3} D^2(u,1)\big) \int (\delta\xi)^2 dxdt \tag{59}$$

for any decaying δu and $\delta\xi$ supported for $t \in (-1,0)$ related via

$$\partial_t \delta u - \partial_x^2 (a \delta u) = \delta\xi, \tag{60}$$

where $a := \pi'(u)$ satisfies (11). Indeed, we note that by duality w. r. t. to the inner product $(g,f) \mapsto \fint_{-r^2}^0 \int \eta_r g f dxdt$,

$$D(u,r) = \sup\Big\{\ D(u,f) := \fint_{-r^2}^0 \int \eta_r u f dxdt \ \Big| \tag{61}$$
$$\fint_{-r^2}^0 \int \eta_r f^2 dxdt = 1, \ \ \mathrm{supp} f \subset (-r^2, 0) \times \mathbb{R}, \ \fint_{-r^2}^0 \int \eta_r f dxdt = 0 \ \Big\}.$$

By the chain rule for the Malliavin derivative we thus obtain

$$|\nabla D(u,r)|_1 \le \sup_f |\nabla D(u,f)|_1,$$

where the supremum runs over the set implicitly defined in (61), so that it is enough to show for a fixed f

$$|\nabla D(u,f)|_1^2 \le r + \frac{L^2}{r^3} D^2(u,1).$$

By definition (10) of the carré-du-champs of the Malliavin derivative applied to the linear functional $u \mapsto D(u,f)$, this amounts to show

$$\fint_{-r^2}^0 \int \eta_r \delta u f dxdt \lesssim \Big(\big(r + \frac{L^2}{r^3} D^2(u,1)\big) \int (\delta\xi)^2 dxdt\Big)^{\frac{1}{2}},$$

where the infinitesimal perturbation δu of the solution is related to the infinitesimal perturbation $\delta\xi$ of the noise supported on $(-1,0) \times \mathbb{R}$ via (60). By the characterizing properties of the f's, cf. (61), this estimate in turn amounts to (59).

We now turn to the proof of (59) and to this purpose split $\delta u = v + w$, where both functions have, like $\delta\xi$ and δu, vanishing initial data, that is $v = w \equiv 0$ for $t \le -1$, and are characterized by

$$\partial_t v - a_0 \partial_x^2 v = \delta\xi$$

and

$$\partial_t w - \partial_x^2(aw) = \partial_x((a-a_0)v),$$

where we will choose the constant coefficient $a_0 \in [\lambda, 1]$ below. Hence we see that by the first part (16) of Proposition 4, which we apply to all translations in space and backwards in time (we are allowed to do the latter because of the vanishing initial data), we have

$$\sup_{(t,x)\in(-\infty,0)\times\mathbb{R}} v^2 \lesssim \int_{-1}^{0}\int (\delta\xi)^2 dxdt. \tag{62}$$

By the second part (17) of Proposition 4, we have

$$D^2(v,r) \lesssim r \int_{-1}^{0}\int (\delta\xi)^2 dxdt. \tag{63}$$

Likewise, the first part (14) of Proposition 3 turns into

$$\int_{-1}^{0}\int \eta w^2 dxdt \lesssim \int_{-1}^{0}\int \eta (a-a_0)^2 v^2 dxdt,$$

which trivially implies for $r \le 1$ (the power three coming from the parabolic dimension):

$$r^3 \fint_{-r^2}^{0}\int \eta_r w^2 dxdt \lesssim \int_{-1}^{0}\int \eta (a-a_0)^2 dxdt \sup_{(t,x)\in(-1,0)\times\mathbb{R}} v^2,$$

into which we insert (62) and where we now choose $a_0 = \pi'(\int_{-1}^{0}\int \eta u)$ so that by the Lipschitz continuity (3) of π'

$$r^3 \fint_{-r^2}^{0}\int \eta_r w^2 dxdt \lesssim L^2 D^2(u,1) \int_{-1}^{0}\int (\delta\xi)^2 dxdt. \tag{64}$$

By the triangle inequality, (59) follows from (63) and (64).

Proof of Theorem 1. In this proof $\lesssim$ and $\ll$ refer to constants only depending on λ, α, and eventually ϵ. We consider the random variable

$$\bar D(u) := \min\{\max_{\rho\in[r,1]\text{dyadic}} \frac{1}{\rho^\alpha} D(u,\rho), \frac{r^2}{L}\}.$$

Because of $\bar{D} \le \sum_{\rho\in[r,1]\text{dyadic}} \frac{1}{\rho^\alpha} D(u,\rho)$ and since $\alpha < \frac{1}{2}$, we learn from Proposition 1 that

$$\langle \bar{D}(u)\rangle_1 \lesssim \big(1+\frac{L}{r^{\frac{3}{2}}}(1+D'(u,1))\big)\big(1+r^{\frac{1}{2}}D'(u,1)\big).$$

By the chain rule for the Malliavin derivative we have

$$|\nabla \bar{D}(u)|_1 \le \max_{\rho\in[r,1]\text{dyadic}} \frac{1}{\rho^\alpha}|\nabla D(u,\rho)|_1$$

and $|\nabla \bar{D}(u)|_1 = 0$ unless $D(u,1) \le \frac{r^2}{L}$. Hence we learn from Proposition 2

$$|\nabla \bar{D}(u)|_1 \lesssim 1.$$

By concentration of measure, cf. the beginning of Section 2, applied to the restricted ensemble $\langle\cdot\rangle_1$ this implies the existence of a random variable χ with Gaussian bounds, that is, $\langle \exp(\frac{1}{C}\chi^2)\rangle_1 \le 2$, and thus a fortiori $\langle \exp(\frac{1}{C}\chi^2)\rangle \le 2$ with respect to the full ensemble $\langle\cdot\rangle$, such that

$$\min\{\max_{\rho\in[\theta,1]} \frac{1}{\rho^\alpha} D(u,\rho), \frac{\theta^2}{L}\} \lesssim \chi + \big(1+\frac{L}{\theta^{\frac{3}{2}}}(1+D'(u,1))\big)\big(1+\theta^{\frac{1}{2}}D'(u,1)\big),$$

where we relabelled r by θ, which we suppose to be small $\theta \ll 1$. By the invariance in law under the scaling (4) & (5) & (6), this yields for any length scale R

$$\begin{aligned}&\min\{\max_{\rho\in[\theta,1]} \frac{1}{\rho^\alpha}\frac{1}{R^{\frac{1}{2}}} D(u,\rho R), \frac{\theta^2}{R^{\frac{1}{2}}}\} \\ &\lesssim \chi_R + \big(1+\frac{1}{\theta^{\frac{3}{2}}}(R^{\frac{1}{2}}+D(u,R))\big)\big(1+\frac{\theta^{\frac{1}{2}}}{R^{\frac{1}{2}}}D'(u,R)\big), \end{aligned} \tag{65}$$

with an R-dependent random variable χ_R of Gaussian moments $\langle \exp(\frac{1}{C}\chi_R^2)\rangle \lesssim 1$. Using the fact that $D'(u,R) = \int \eta_R(u - \int \eta_R u)^2 dx_{|t=-R^2}$ satisfies $\fint_{\frac{R}{2}}^{R} D'(u,R')dR' \lesssim D(u,R)$, we see that by replacing R by R' in (65) and by averaging over $R' \in (\frac{R}{2}, R)$ we obtain

$$\begin{aligned}&\min\{\max_{\rho\in[2\theta,1]} \frac{1}{\rho^\alpha}\frac{1}{R^{\frac{1}{2}}} D(u,\rho R), \frac{\theta^2}{R^{\frac{1}{2}}}\} \\ &\lesssim \chi_R' + \big(1+\frac{1}{\theta^{\frac{3}{2}}}(R^{\frac{1}{2}}+D(u,R))\big)\big(1+\frac{\theta^{\frac{1}{2}}}{R^{\frac{1}{2}}}D(u,R)\big),\end{aligned}$$

where $\chi'_R := \fint_{\frac{R}{2}}^R \chi_{R'} dR$ still has Gaussian moments $\langle \exp(\frac{1}{C}{\chi'_R}^2)\rangle \le 2$, since the latter property is preserved by convex combination. Changing the value of θ by a factor of two, the above implies

$$\begin{aligned}
&\min\{\frac{1}{\theta^\alpha}\frac{1}{R^{\frac12}}D(u,\theta R), \frac{\theta^2}{R^{\frac12}}\} \\
&\lesssim \ \chi'_R + \big(1+\frac{1}{\theta^{\frac32}}(R^{\frac12}+D(u,R))\big)\big(1+\frac{\theta^{\frac12}}{R^{\frac12}}D(u,R)\big).
\end{aligned}$$

Because of the trivial estimate $D(u,\theta R) \le \frac{1}{\theta^{\frac32}}D(u,R)$ we see that $D(u,R) \le \theta^4 \le \theta^{\frac72+\alpha}$ implies $D(u,\theta R) \le \theta^{2+\alpha}$ and trivially $D(u,R) \le \theta^{\frac32}$ so that the above yields

$$\left.\begin{array}{c} D(u,R) \le \theta^4 \\ R \le \theta^3 \end{array}\right\} \implies \frac{1}{\theta^\alpha}\frac{1}{R^{\frac12}}D(u,\theta R) \le \chi''_R + C\frac{\theta^{\frac12}}{R^{\frac12}}D(u,R),$$

where $\chi''_R \sim \chi'_R + 1$ still has Gaussian moments $\langle \exp(\frac{1}{C}{\chi''_R}^2)\rangle \lesssim 1$. Hence selecting $\theta \sim 1$ sufficiently small, we obtain

$$\left.\begin{array}{c} D(u,R) \le \theta^4 \\ R \le \theta^3 \end{array}\right\} \implies \frac{1}{(\theta R)^\alpha}D(u,\theta R) \le R^{\frac12-\alpha}\chi''_R + \frac12\frac{1}{R^\alpha}D(u,R). \qquad (66)$$

Since (66) implies in particular $D(u,\theta R) \le R^{\frac12}\chi''_R + \frac12 D(u,R)$, we see that in order to convert (66) into a self-propelling iteration, we need $R^{\frac12}\chi''_R \le \frac12\theta^4$, $(\theta R)^{\frac12}\chi''_{\theta R} \le \frac12\theta^4$ and so on. This prompts to consider the random variable

$$\bar\chi_R := \max_{n=0,1,\cdots}(\theta^n)^{\frac12-\alpha}\chi''_{\theta^n R} \ge \max_{n=0,1,\cdots}(\theta^n)^{\frac12}\chi''_{\theta^n R},$$

which in view of (recall $\alpha < \frac12$)

$$\begin{aligned}
\bar\chi_R &\le \sum_{n=0}^{\infty}(\theta^n)^{\frac12-\alpha}\chi''_{\theta^n R} \\
&= \frac{1}{1-\theta^{\frac12-\alpha}} \times \text{convex combination of } \{\chi''_{\theta^n R}\}_{n=0,1,\cdots}
\end{aligned}$$

has Gaussian moments $\big\langle \exp\big(\frac{1}{C}\bar\chi_R^2\big)\big\rangle \lesssim 1$ since by construction, the random variables $\{\chi''_R\}_R$ have a uniform Gaussian moment bounds. From (66) we learn

$$\begin{aligned}
&D(u,R) \le \theta^4 \quad \text{and} \quad R^{\frac12}\bar\chi_R \le \frac12\theta^4 \\
&\implies \forall\, n \in \mathbb{N} \ \ \frac{1}{(\theta^n R)^\alpha}D(u,\theta^n R) \le R^{\frac12-\alpha}\bar\chi_R + \frac12\frac{1}{(\theta^{n-1}R)^\alpha}D(u,\theta^{n-1}R),
\end{aligned}$$

where we assumed w. l. o. g. that $\bar\chi_R \geq \frac{1}{2\theta^2}$ so that $R^{\frac12}\bar\chi_R \leq \frac12\theta^4$ implies en passant $R \leq \theta^3$. Thanks to the factor $\frac12 < 1$ the last statement can be iterated to yield

$$D(u,R) \leq \theta^4 \quad \text{and} \quad R^{\frac12}\bar\chi_R \leq \frac12\theta^4$$
$$\Longrightarrow \quad \sup_{n=0,1,\cdots} \frac{1}{(\theta^n R)^\alpha} D(u,\theta^n R) \leq 2\big(R^{\frac12-\alpha}\bar\chi_R + \frac{1}{R^\alpha}D(u,R)\big),$$

which implies (using once more $D(u,r) \leq (\frac{R}{r})^{\frac32}D(u,R)$ for any scales $r \leq R$ to bridge the dyadic gaps)

$$D(u,R) \leq \theta^4 \quad \text{and} \quad R^{\frac12}\bar\chi_R \leq \frac12\theta^4 \quad \Longrightarrow \quad \sup_{r\leq R}\frac{1}{r^\alpha}D(u,r) \leq \frac{2\theta^{4-\frac32}}{R^\alpha}.$$

Summing up, we learned that for any length scale R, which we relabel by ρ, we have

$$D(u,\rho) \ll 1 \quad \text{and} \quad \rho^{\frac12}\bar\chi_\rho \ll 1 \quad \Longrightarrow \quad \sup_{r\leq\rho}(\frac{\rho}{r})^\alpha D(u,r) \lesssim 1.$$

Once more by our scaling invariance (4) & (5), this can be rephrased as

$$\frac{1}{R^{\frac12}}D(u,\rho R) \ll 1 \quad \text{and} \quad \rho^{\frac12}\bar\chi_{\rho,\rho R} \ll 1 \quad \Longrightarrow \quad \sup_{r\leq\rho}(\frac{\rho}{r})^\alpha \frac{1}{R^{\frac12}}D(u,rR) \lesssim 1$$

for a family $\{\bar\chi_{\rho,R'}\}_{\rho,R'}$ of random variables with uniformly bounded Gaussian moments, which we use for $R = \frac{1}{\rho}$ so that (with $r' = \frac{r}{\rho}$)

$$\rho^{\frac12}D(u,1) \ll 1 \quad \text{and} \quad \rho^{\frac12}\bar\chi_{\rho,1} \ll 1 \quad \Longrightarrow \quad \rho^{\frac12}\sup_{r'\leq 1}\frac{1}{r'^\alpha}D(u,r') \lesssim 1.$$

We rewrite this in terms of $M = \rho^{-\frac12}$, which we think of being large (and relabel r' with r):

$$D(u,1) \ll M \quad \text{and} \quad \bar\chi_{\frac{1}{M^2},1} \ll M \quad \Longrightarrow \quad \sup_{r\leq 1}\frac{1}{r^\alpha}D(u,r) \lesssim M. \qquad (67)$$

We now assume that $M \geq 1$ is of dyadic form. For any exponent $\epsilon > 0$, which we think of being small, we introduce the random variable

$$\bar\chi := \max_{M\geq 1 \text{dyadic}} M^{-\epsilon}\bar\chi_{\frac{1}{M^2},1}$$

and note that with the same reasoning as above, $\bar\chi$ has Gaussian moments. With help of $\bar\chi$, (67) may be rephrased as

$$D(u,1)\ll M \quad\text{and}\quad \bar\chi\ll M^{1-\epsilon} \implies \sup_{r\le 1}\frac{1}{r^\alpha}D(u,r)\lesssim M,$$

which by the arbitrariness of the dyadic $M\ge 1$ implies

$$\big(\sup_{r\le 1}\frac{1}{r^\alpha}D(u,r)\big)^{1-\epsilon}\lesssim D(u,1)+\bar\chi+1.$$

From this we learn that the Gaussian moment for $(\sup_{r\le 1}\frac{1}{r^\alpha}D(u,r))^{1-\epsilon}$, cf. (9), follow from the Gaussian moments for $\bar\chi$ and for $D(u,1)$, cf. Lemma 1. The constant in the exponential can be absorbed into the loss ϵ w. r. t. to the Gaussian moments.

Proof of Lemma 2. Since thanks to $r\ll 1$ we have $\int\eta_r(h)(e^{|h|}-1)^2dh\lesssim r^2$, it is enough to show for any shift h

$$\begin{aligned}&\Big\langle\int_{-\frac12}^0\int\tilde\eta^2(u^h-u)^2dxdt\Big\rangle_1\\ &\quad\lesssim\ |h|+(e^{|h|}-1)^2\Big\langle\int_{-1}^0\int\tilde\eta^2(u-c)^2dxdt\Big\rangle_1,\end{aligned}\tag{68}$$

where, as in Lemma 1, for the upcoming calculations we have replaced the exponential cut-off $\eta=\eta_2^2$ by its smooth version $\tilde\eta^2$ where

$$\tilde\eta(x):=\exp(-\frac12\sqrt{x^2+1})\sim\eta_2(x)\tag{69}$$

and we have set for abbreviation $c:=\int_{-1}^0\int\tilde\eta^2udxdt$. By the martingale argument based on the stochastic differential equation

$$\partial_t(u^h-u)=-(-\partial_x^2)(\pi(u^h)-\pi(u))+(\xi^h-\xi)$$

we have

$$\begin{aligned}&\frac{d}{dt}\frac12\Big\langle\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\eta(u^h-u)dx\Big\rangle_1\\ &=\ -\Big\langle\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u^h)-\pi(u))dx\Big\rangle_1\\ &\quad+\frac12\int\tilde\eta(\tilde\eta-\frac12\tilde\eta^he^{-|h|}-\frac12\tilde\eta^{-h}e^{-|h|})dx.\end{aligned}\tag{70}$$

Let us make two comments on (70): Like in Lemma 1 we use physics notation in the sense that an operator acts on all the terms to its right, e. g. in the above expression $(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u^h)-\pi(u)) = (1-\partial_x^2)^{-1}[\tilde\eta(-\partial_x^2)(\pi(u^h)-\pi(u))]$. The last term in (70), which comes from the quadratic variation of the white noise in time, cf. Lemma 1 for a heuristic discussion, assumes this form because $\frac{1}{2}\exp(-|x-y|)$ is the (translation-invariant) kernel of the operator $(1-\partial_x^2)^{-1}$, so that $\tilde\eta(x)\frac{1}{2}\exp(-|x-y|)\tilde\eta(y)$ is the kernel of the operator $\tilde\eta(1-\partial_x^2)^{-1}\tilde\eta$, so that the quadratic variation is indeed given by

$$\frac{1}{2}\int\int \tilde\eta(x)\frac{1}{2}\exp(-|x-y|)\tilde\eta(y) \\ \times\big(\delta((x+h)-(y+h)) - \delta((x+h)-y) - \delta(x-(y+h)) + \delta(x-y)\big)dxdy,$$

where the spatial Dirac distributions come from the spatial white noise ξ_{spat}, more precisely, they represent the covariance $\langle(\xi_{spat}^h-\xi_{spat})(x)(\xi_{spat}^h-\xi_{spat})(y)\rangle_1$ of the increment $\xi_{spat}^h-\xi_{spat}$.

We integrate (70) against the weight $t+1$ in time over $t\in(-1,0)$. This yields (68) once we establish the following three estimates: The following estimate on the quadratic variation

$$\int\tilde\eta(\tilde\eta-\frac{1}{2}\tilde\eta^h e^{-|h|}-\frac{1}{2}\tilde\eta^{-h}e^{-|h|})dx \lesssim |h|, \tag{71}$$

the following bound on the term under the time derivative

$$\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(u^h-u)dx \lesssim (e^{|h|}-1)^2\int\tilde\eta^2(u-c)^2dx, \tag{72}$$

and the fact that "elliptic term" controls the desired term up to the term in (72)

$$\begin{aligned}\int\tilde\eta^2(u^h-u)^2dx &\le \frac{1}{C}\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u^h)-\pi(u))dx \\ &+ C\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(u^h-u)dx. \end{aligned}\tag{73}$$

We first address the quadratic variation term (71). Writing

$$\tilde\eta-\frac{1}{2}\tilde\eta^h e^{-|h|}-\frac{1}{2}\tilde\eta^{-h}e^{-|h|} = \tilde\eta(1-e^{-|h|}) + e^{-|h|}(\tilde\eta-\frac{1}{2}\tilde\eta^h-\frac{1}{2}\tilde\eta^{-h})$$

and performing a discrete integration by parts, we see that this term takes the form of

$$(1-e^{-|h|})\int\tilde\eta^2dx + e^{-|h|}\int(\tilde\eta^h-\tilde\eta)^2dx,$$

so that the estimate follows from the elementary estimate $\int(\tilde\eta^h-\tilde\eta)^2dx \le h^2\int(\partial_x\tilde\eta)^2dx$.

We note that by duality, the estimate of the time-derivative term (72) is equivalent to

$$\int\zeta\tilde\eta(u^h-u)dx \lesssim (e^{|h|}-1)\left(\int(\zeta^2+(\partial_x\zeta)^2dx\int\tilde\eta^2(u-c)^2dx\right)^{\frac{1}{2}},$$

which follows by discrete versions of integration by parts and Leibniz' rule

$$\int\zeta\tilde\eta(u^h-u)dx = \int\big((\zeta^{-h}-\zeta)\tilde\eta+\zeta^{-h}(\tilde\eta^{-h}-\tilde\eta)\big)(u-c)dx,$$

Cauchy-Schwarz' inequality, the standard estimate

$$\int(\zeta^{-h}-\zeta)^2dx \le h^2\int(\partial_x\zeta)^2dx, \tag{74}$$

and the following property of our cut-off function with exponential tails

$$\begin{aligned}
&|\tilde\eta^{-h}(x)-\tilde\eta(x)| \\
=\;& \exp(-\frac{1}{2}\sqrt{x^2+1})|\exp(\frac{1}{2}\sqrt{x^2+1}-\frac{1}{2}\sqrt{(x-h)^2+1})-1| \\
\le\;& \tilde\eta(x)|\exp(\frac{|h|}{2})-1|.
\end{aligned}$$

Let us finally address the elliptic term (73). To this purpose we write (in our physicist's way of omitting parentheses)

$$\begin{aligned}
&\tilde\eta(-\partial_x^2)(\pi(u^h)-\pi(u)) \\
=\;& (1-\partial_x^2)(\pi(u^h)-\pi(u))\tilde\eta+2\partial_x(\pi(u^h)-\pi(u))\partial_x\tilde\eta \\
&-(\pi(u^h)-\pi(u))(1-\partial_x^2)\tilde\eta,
\end{aligned}$$

so that by the symmetry of $(1-\partial_x^2)^{-1}$ (already used for (70))

$$\begin{aligned}
&\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u^h)-\pi(u))dx \\
=\;& \int\tilde\eta^2(u^h-u)(\pi(u^h)-\pi(u))dx \\
&-2\int(\partial_x\tilde\eta)(\pi(u^h)-\pi(u))\partial_x(1-\partial_x)^{-1}\tilde\eta(u^h-u)dx \\
&-\int((1-\partial_x^2)\tilde\eta)(\pi(u^h))-\pi(u))(1-\partial_x^2)^{-1}\tilde\eta(u^h-u)dx.
\end{aligned}$$

Using that the operators $\partial_x(1-\partial_x^2)^{-\frac{1}{2}}$ and $(1-\partial_x^2)^{-\frac{1}{2}}$ have operator norm 1 w. r. t. to L^2, we deduce the inequality (where we use the abbreviation $\pi^h-\pi:=\pi(u^h)-\pi(u)$)

$$\begin{aligned}
&\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi^h-\pi)dx\\
&\quad\ge\ \int\tilde\eta^2(u^h-u)(\pi^h-\pi)dx\\
&\qquad-\Big(2\big(\int(\partial_x\tilde\eta)^2(\pi^h-\pi)^2dx\big)^{\frac{1}{2}}+\big(\int((1-\partial_x^2)\tilde\eta)^2(\pi^h-\pi)^2dx\big)^{\frac{1}{2}}\Big)\\
&\qquad\times\Big(\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(u^h-u)dx\Big)^{\frac{1}{2}}. \qquad (75)
\end{aligned}$$

By the monotonicity properties (2) of π, this yields

$$\begin{aligned}
&\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u^h)-\pi(u))dx\\
&\quad\ge\ \lambda\int\tilde\eta^2(u^h-u)^2dx\\
&\qquad-\Big(2\big(\int(\partial_x\tilde\eta)^2(u^h-u)^2dx\big)^{\frac{1}{2}}+\big(\int((1-\partial_x^2)\tilde\eta)^2(u^h-u)^2dx\big)^{\frac{1}{2}}\Big)\\
&\qquad\times\Big(\int\tilde\eta(u^h-u)(1-\partial_x^2)^{-1}\tilde\eta(u^h-u)dx\Big)^{\frac{1}{2}}. \qquad (76)
\end{aligned}$$

Our smoothing out of the exponential cut-off function ensures

$$|\partial_x\tilde\eta|+|\partial_x^2\tilde\eta|\lesssim\tilde\eta, \qquad (77)$$

which allows us to use Young's inequality in order to arrive at (73).

Proof of Lemma 3. We will establish this lemma in the strengthened version with the bulk average $\int_{-1}^0\int\eta u dxdt$ replaced by the surface average $c:=\int\eta u dx_{|t=-1}$. To this purpose we rewrite (1) in form of

$$\partial_t(u-c)=-(-\partial_x^2)(\pi(u)-\pi(c))+\xi.$$

As in Lemma 2, we replace η by $\tilde\eta^2\sim\eta$ in the statement of this lemma, with $\tilde\eta$ being the mollified version of η_2, cf. (69). By the martingale argument we have like in Lemma 1, cf. (25),

$$\begin{aligned}
&\frac{d}{dt}\frac{1}{2}\Big\langle\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(u-c)dx\Big\rangle_1\\
&\quad=\ -\Big\langle\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u)-\pi(c))dx\Big\rangle_1+\frac{1}{2}\int\frac{1}{2}\tilde\eta^2dx.
\end{aligned}$$

We smuggle in an exponential term in the time variable with a rate $T \ll 1$ to be adjusted later:

$$
\begin{aligned}
&\frac{d}{dt}\exp(-\frac{t}{T})\frac{1}{2}\Big\langle\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(u-c)dx\Big\rangle_1\\
&= \quad -\exp(-\frac{t}{T})\Big(\Big\langle\frac{1}{2T}\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(u-c)dx\\
&\qquad +\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u)-\pi(c))dx\Big\rangle_1+\frac{1}{4}\int\tilde\eta^2dx\Big).
\end{aligned}
$$

Lemma 3 will follow from integration over $t\in(0,1)$ of this identity, using the obvious estimates on the quadratic variation term

$$
\int\tilde\eta^2dx\lesssim 1,
$$

and on the term under the time derivative

$$
\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(u-c)dx\le\int\tilde\eta^2(u-c)^2,
$$

once we show that the elliptic term controls the desired term for T sufficiently small:

$$
\begin{aligned}
\frac{1}{C}\int\tilde\eta^2(u-c)^2dx \quad &\le \quad \int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u)-\pi(c))dx\\
&+ \quad \frac{1}{2T}\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(u-c)dx. \qquad (78)
\end{aligned}
$$

The argument for this estimate (78) on the elliptic term follows the lines of the one in Lemma 2: Replacing the couple (u^h,u) from there by (u,c), we arrive at

$$
\begin{aligned}
&\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(-\partial_x^2)(\pi(u)-\pi(c))dx\\
&\ge \quad \lambda\int\tilde\eta^2(u-c)^2dx\\
&\quad -\Big(2\big(\int(\partial_x\tilde\eta)^2(u-c)^2dx\big)^{\frac{1}{2}}+\big(\int((1-\partial_x^2)\tilde\eta)^2(u-c)^2dx\big)^{\frac{1}{2}}\Big)\\
&\quad \times\Big(\int\tilde\eta(u-c)(1-\partial_x^2)^{-1}\tilde\eta(u-c)dx\Big)^{\frac{1}{2}}. \qquad (79)
\end{aligned}
$$

Appealing to the estimates (77) of the smoothened exponential cut-off $\tilde\eta$ and Young's inequality, we obtain (78) for a sufficiently large $\frac{1}{T}$.

Proof of Lemma 4. We fix an $r \le 1$ and note that the statement of this lemma follows from

$$\begin{aligned} &\Big\langle\Big(\fint_{-r^2}^{0}(\int \eta_r u - \fint_{-r^2}^{0}\int \eta_r u)^2 dxdt\Big)^{\frac{1}{2}}\Big\rangle_r \\ &\lesssim\ r^{\frac{1}{2}} + \Big\langle\Big(\fint_{-r^2}^{0}\int \eta_r (u - \int \eta_r u)^2 dxdt\Big)^{\frac{1}{2}}\Big\rangle_r, \end{aligned} \tag{80}$$

where $\langle\cdot\rangle_r$ denotes the expectation w. r. t. to the white noise restricted to $(t,x) \in (-r^2, 0) \times \mathbb{R}$, just by taking the expectation w. r. t. to $\langle\cdot\rangle_1$. By the scale invariance (4) & (5), it is thus sufficient to establish the above for $r = 1$. We shall replace the exponential averaging function η by its mollified version

$$\tilde\eta(x) = \frac{1}{c_0}\exp(-\sqrt{x^2+1}) \quad \text{with} \quad c_0 := \int \exp(-\sqrt{x^2+1})dx,$$

noting that $\tilde\eta \sim \eta$ and pointing out the slight difference to Lemmas 2 and 3, cf. (69). Indeed, $\tilde\eta \sim \eta$ is enough to replace η by $\tilde\eta$ on the r. h. s. of (80); for the l. h. s. this follows the L^2-average in time of the estimate

$$\begin{aligned} &\Big|\int \eta u dx - \int \tilde\eta u dx\Big| \\ &= \ \Big|\int (\eta - \tilde\eta)(u - \int \eta u) dx\Big| \lesssim \Big(\int \eta (u - \int \eta u)^2 dx\Big)^{\frac{1}{2}}. \end{aligned}$$

Hence with the abbreviation $U(t) := \int \tilde\eta u dx$ we need to show that

$$\Big\langle\Big(\int_{-1}^{0}(U - \fint_{-1}^{0} U)^2 dxdt\Big)^{\frac{1}{2}}\Big\rangle_1 \ \lesssim\ 1 + \Big\langle\Big(\int_{-1}^{0}\int \tilde\eta (u - U)^2 dxdt\Big)^{\frac{1}{2}}\Big\rangle_1. \tag{81}$$

After these preparations, we note that we may rewrite equation (1) in form of

$$\partial_t u = \partial_x^2\big(\pi(u) - \pi(U)\big) + \xi,$$

From this we deduce the stochastic *ordinary* differential equation

$$\partial_t \int \tilde\eta u dx = \int \big(\pi(u) - \pi(U)\big)\partial_x^2 \tilde\eta dx + \sigma \partial_t W,$$

where W is a standard temporal Wiener process and the variance is given by

$$\sigma^2 := \int \tilde\eta dx \sim 1. \tag{82}$$

We use the differential equation in its time-integrated version

$$\int_{-1}^{0}(\partial_t(U-\sigma W))^2dt=\int_{-1}^{0}\Big(\int(\pi(u)-\pi(U))\partial_x^2\tilde{\eta}dx\Big)^2dt.$$

Thanks to the Lipschitz continuity of π, (2), and the fact that due to our mollification $\tilde{\eta}$ of the exponential averaging function, we have $|\partial_x^2\tilde{\eta}|\lesssim\tilde{\eta}$, this turns into the estimate

$$\int_{-1}^{0}(\partial_t(U-\sigma W))^2dt\lesssim\int_{-1}^{0}\int\tilde{\eta}(u-U)^2dxdt.$$

By Poincaré's inequality (with vanishing mean value) and the triangle inequality, and appealing to (82), this turns into

$$\int_{-1}^{0}(U-\int_{-1}^{0}U)^2dt\lesssim\int_{-1}^{0}(W-\int_{-1}^{0}W)^2dt+\int_{-1}^{0}\int\tilde{\eta}(u-U)^2dxdt.$$

By Jensen's inequality and the defining properties on the quadratic moments of the Brownian motion, this implies

$$\begin{aligned}&\Big\langle\Big(\int_{-1}^{0}(U-\int_{-1}^{0}U)^2dt\Big)^{\frac{1}{2}}\Big\rangle_1\\ \lesssim\ &\Big\langle\int_{-1}^{0}(W-\int_{-1}^{0}W)^2dt\Big\rangle^{\frac{1}{2}}+\Big\langle\Big(\int_{-1}^{0}\int\tilde{\eta}(u-U)^2dxdt\Big)^{\frac{1}{2}}\Big\rangle_1\\ \lesssim\ &1+\Big\langle\Big(\int_{-1}^{0}\int\tilde{\eta}(u-U)^2dxdt\Big)^{\frac{1}{2}}\Big\rangle_1,\end{aligned}$$

which is (81).

Proof of Lemma 5. First of all, the observable $\sup_{(-1,0)\times\mathbb{R}}\eta(g^h-g)^2$ is independent from the noise ξ outside of $(-1,0)\times\mathbb{R}$, so that we can replace the average $\langle\cdot\rangle_1$ by $\langle\cdot\rangle$. We start by arguing that

$$\langle(g(t,x)-g(s,y))^2\rangle\lesssim\sqrt{t-s}+|x-y|\quad\text{for all }(t,x),(s,y)\in\mathbb{R}\times\mathbb{R}.\tag{83}$$

Because of the initial conditions and symmetry, we may w. l. o. g. assume that $-1\le s\le t$. The heat kernel $K(t,x):=\frac{1}{\sqrt{4\pi a_0t}}\exp(-\frac{|x|}{4a_0t})$ provides us with the representation

$$g(t,x)=\int_{-1}^{t}\int K(t-t',x-x')\xi(t',x')dx'dt'.\tag{84}$$

Because of the defining property $\langle(\int\zeta\xi dxdt)^2\rangle=\int\zeta^2dxdt$ for a test function ζ of white noise (that is, $\langle\zeta(t',x')\xi(s',y')\rangle=\delta(t'-s')\delta(x'-y')$ in the rough but efficient physics language) this yields the identity

$$\begin{aligned}
\langle g(t,x)g(s,y)\rangle &= \int_{-1}^{s}\int K(t-\tau,x-z)K(s-\tau,y-z)dzd\tau\\
&= \int_{-1}^{s}K(t+s-2\tau,x-y)d\tau\\
&= \frac{1}{2}\int_{t-s}^{t+s+2}K(\sigma,x-y)d\sigma,
\end{aligned}$$

where we used the semi-group property of $t\mapsto K(t,\cdot)$ in the middle identity. We now pass from covariance to increment:

$$\begin{aligned}
&\langle(g(t,x)-g(s,y))^2\rangle\\
&\quad= \langle g^2(t,x)\rangle+\langle g^2(s,y)\rangle-2\langle g(t,x)g(s,y)\rangle\\
&\quad= \Big(\frac{1}{2}\int_0^{2(t+1)}+\frac{1}{2}\int_0^{2(s+1)}-\int_{t-s}^{(t+1)+(s+1)}\Big)K(\sigma,0)d\sigma\\
&\qquad+\int_{t-s}^{t+s+2}(K(\sigma,0)-K(\sigma,x-y))d\sigma.
\end{aligned}$$

By positivity and monotonicity of $K(\sigma,0)$ in σ and $K(\sigma,0)\geq K(\sigma,z)$, this yields the inequality

$$\begin{aligned}
&\langle(g(t,x)-g(s,y))^2\rangle\\
&\quad\leq \int_0^{t-s}K(\sigma,0)d\sigma+\int_0^{\infty}(K(\sigma,0)-K(\sigma,x-y))d\sigma\\
&\quad\leq \int_0^{t-s}K(\sigma,0)d\sigma+|x-y|\int_0^{\infty}(K(\sigma,0)-K(\sigma,1))d\sigma,
\end{aligned}$$

where we used the scale invariance of $K(\sigma,z)$ in the second step. This inequality implies (83) because of $K(\sigma,0)\lesssim\sigma^{-\frac{1}{2}}$ and $K(\sigma,0)-K(\sigma,1)\lesssim\min\{\sigma^{-\frac{1}{2}},\sigma^{-\frac{3}{2}}\}$.

We now apply Kolmogorov's continuity theorem; for the convenience of the reader and because of its similarity to the proof of the main result of the paper, we give a self-contained argument. We first appeal to Gaussianity to post-process (83), which we rewrite as

$$\langle\frac{1}{R}(g(t,x)-g(s,y))^2\rangle\lesssim1\quad\text{provided } |t-s|\leq3R^2, |x-y|\leq R$$

for a given scale R. We note that from (84) we see that the properties of being Gaussian and centered transmits from ξ to $\frac{1}{\sqrt{R}}(g(t,x)-g(s,y))$, so that by the above normalization we have

$$\Big\langle \exp\big(\frac{1}{CR}(g(t,x)-g(s,y))^2\big)\Big\rangle \lesssim 1 \quad \text{for } |t-s|\le 3R^2,\ |x-y|\le R. \tag{85}$$

Our goal is to estimate exponential moments of the local Hölder-norm

$$[g]_{\alpha,(-1,0)\times(-1,1)} \ := \ \sup_{R\in(0,1)} \frac{1}{R^\alpha} \sup_{\substack{(t,x),(s,y)\in(-1,0)\times(-1,1)\\ \sqrt{|t-s|}+|x-y|<R}} |g(t,x)-g(s,y)|,$$

which amounts to exchange the expectation and the supremum over (t,x), (s,y) in (85) at the prize of a decreased Hölder exponent $\alpha<\frac{1}{2}$ To this purpose, we now argue that for $\alpha>0$, the supremum over a continuum can be replaced by the supremum over a discrete set: For $R<1$ we define the grid

$$\Gamma_R = [-1,0]\times[-1,1]\cap(R^2\mathbb{Z}\times R\mathbb{Z})$$

and claim that

$$\begin{aligned} &[g]_{\alpha,(-1,0)\times(-1,1)} \\ &\lesssim \ \sup_R \frac{1}{R^\alpha} \sup_{\substack{(t,x),(s,y)\in\Gamma_R\\ |t-s|\le 3R^2,|x-y|\le R}} |g(t,x)-g(s,y)| =: \Lambda\ , \end{aligned} \tag{86}$$

where the first *sup* runs over all R of the form 2^{-N} for an integer $N\ge 1$. Hence we have to show for arbitrary $(t,x),(s,y)\in(-1,0)\times(-1,1)$ that

$$|g(t,x)-g(s,y)| \lesssim \Lambda\big(\sqrt{|t-s|}+|x-y|\big)^\alpha. \tag{87}$$

By density, we may assume that $(t,x),(s,y)\in r^2\mathbb{Z}\times r\mathbb{Z}$ for some dyadic $r=2^{-N}<1$ (this density argument requires the qualitative a priori information of the continuity of g, which can be circumvented by approximating ξ). By symmetry and the triangle inequality, we may assume $s\le t$ and $x\le y$. For every dyadic level $n=N,N-1,\cdots$ we now recursively construct two sequences (t_n,x_n) (s_n,y_n) of space-time points (in fact, the space and time points can be constructed separately), starting from $(t_N,x_N)=(t,x)$ and $(s_N,y_N)=(s,y)$, with the following properties

a) they are in the corresponding lattice of scale 2^{-n}, i. e. $(t_n,x_n),(s_n,x_n)\in(2^{-n})^2\mathbb{Z}\times 2^{-n}\mathbb{Z}$,

b) they are close to their predecessors in the sense of $|t_n - t_{n+1}|, |s_n - s_{n+1}| \leq 3(2^{-(n+1)})^2$ and $|x_n - x_{n+1}|, |y_n - y_{n+1}| \leq 2^{-(n+1)}$, so that by definition of Λ we have

$$|g(t_n,x_n) - g(t_{n+1},x_{n+1})|,|g(s_n,y_n) - g(s_{n+1},y_{n+1})| \leq \Lambda(2^{-(n+1)})^\alpha, \tag{88}$$

and

c) such that $|t_n - s_n|$ and $|x_n - y_n|$ are minimized among these points.

Because of the latter, we have

$$(t_M,x_M) = (s_M,y_M) \quad \text{for some } M \text{ with} \quad 2^{-M} \leq \max\{\sqrt{|t-s|}, |x-y|\},$$

so that by the triangle inequality we gather from (88)

$$|g(t,x) - g(s,y)| \leq \sum_{n=N-1}^{M} \Lambda(2^{-(n+1)})^\alpha \leq \Lambda \frac{(2^{-M})^\alpha}{2^\alpha - 1},$$

which yields (87).

Equipped with (86), we now may upgrade (85) to

$$\Big\langle \exp\big(\frac{1}{C}[g]^2_{\alpha,(-1,0)\times(-1,1)}\big)\Big\rangle \lesssim 1 \tag{89}$$

for $\alpha < \frac{1}{2}$. Indeed, (86) can be reformulated on the level of characteristic functions as

$$I\big([g]^2_{\alpha,(-1,0)\times(-1,1)} \geq M\big) \leq \sup_R \max_{(t,x),(s,y)\in\Gamma_R} I\big(\frac{1}{R}(g(t,x)-g(s,y))^2 \geq \frac{M}{CR^{1-2\alpha}}\big),$$

where as in (86) R runs over all 2^{-N} for integers $N \geq 1$. Replacing the suprema by sums in order to take the expectation, we obtain

$$\Big\langle I\big([g]^2_{\alpha,(-1,0)\times(-1,1)} \geq M\big)\Big\rangle \leq \sum_R \sum_{(t,x),(s,y)} \Big\langle I\big(\frac{1}{R}(g(t,x)-g(s,y))^2 \geq \frac{M}{CR^{1-2\alpha}}\big)\Big\rangle.$$

We now appeal to Chebyshev's inequality in order to make use of (85):

$$\begin{aligned}
&\Big\langle I\big([g]^2_{\alpha,(-1,0)\times(-1,1)} \geq M\big)\Big\rangle \\
&\quad\lesssim \quad \sum_R \sum_{(t,x),(s,y)} \exp(-\frac{M}{CR^{1-2\alpha}}) \\
&\quad\lesssim \quad \sum_R \frac{1}{R^3} \exp(-\frac{M}{CR^{1-2\alpha}}) \\
&\quad\overset{R\leq 1, M\geq 1}{\leq} \quad \exp(-\frac{M}{C}) \sum_R \frac{1}{R^3} \exp(-\frac{1}{C}(\frac{1}{R^{1-2\alpha}} - 1)) \lesssim \exp(-\frac{M}{C}),
\end{aligned}$$

where in the second step we have used that the number of pairs $(t,x),(s,y)$ of neighboring lattice points is bounded by $C\frac{1}{R^3}$ and in the last step we have used that stretched exponential decay (recall $1-2\alpha>0$) beats polynomial growth. The last estimate immediately yields (89).

It remains to post-process (89), which we only need for second moments but with the spatial origin replaced by any point x:

$$\big\langle [g]^2_{\alpha,(-1,0)\times(x-1,x+1)}\big\rangle \lesssim 1 \quad \text{for all } x\in\mathbb{R}, \tag{90}$$

to obtain the statement of the lemma in form of

$$\langle \sup_{(-1,0)\times\mathbb{R}} \eta(g^h-g)^2\rangle \lesssim \min\{|h|^{2\alpha},1\}. \tag{91}$$

To this purpose, we distinguish the cases $|h|\le\frac12$ and $|h|\ge\frac12$. In the first case we have

$$\begin{aligned}
\sup_{(-1,0)\times\mathbb{R}} \eta(g^h-g)^2 \quad &\lesssim \quad \sum_{x\in\mathbb{Z}} \exp(-|x|) \sup_{(-1,0)\times(x-\frac12,x+\frac12)} (g^h-g)^2 \\
&\overset{|h|\le\frac12}{\lesssim} \quad |h|^{2\alpha}\sum_{x\in\mathbb{Z}} \exp(-|x|)[g]^2_{\alpha,(-1,0)\times(x-1,x+1)},
\end{aligned}$$

from which (91) follows by taking the expectation and inserting (90). In case of $|h|\ge\frac12$, we proceed via

$$\begin{aligned}
&\sup_{(-1,0)\times\mathbb{R}} \eta(g^h-g)^2 \\
&\quad\lesssim \quad \sup_{(-1,0)\times\mathbb{R}} \eta g^2 + \sup_{(-1,0)\times\mathbb{R}} \eta^{-h} g^2 \\
&\quad\lesssim \quad \sum_{x\in 2\mathbb{Z}} (\exp(-|x|)+\exp(-|x-h|)) \sup_{(-1,0)\times(x-1,x+1)} g^2 \\
&\quad\overset{g(t=-1)=0}{\lesssim} \quad \sum_{x\in 2\mathbb{Z}} (\exp(-|x|)+\exp(-|x-h|)) \sup_{(-1,0)\times(x-1,x+1)} [g]^2_{\alpha,(-1,0)\times(x-1,x+1)}.
\end{aligned}$$

Proof of Corollary 1. We start by defining a modified local Hölder norm, based on the $D(u,r)$. For $R>0$ set

$$\Gamma_R = \cap[-1,1]\times[-1,1]\cap\left(R^2\mathbb{Z}\times R\mathbb{Z}\right). \tag{92}$$

Then we define the modified Hölder semi-norm

$$[\![u]\!]_\alpha = \sup_R \sup_{(\bar t,\bar x)\in\Gamma_R} \frac{1}{R^\alpha} D(u^{(\bar t,\bar x)},R)\,,$$

where for a space-time point $(\bar t,\bar x)$ we write $u^{(\bar t,\bar x)}(t,x)=u(t+\bar t,x+\bar x)$ and the first supremum is taken over all $R=2^{-N}$ for integer $N\geq 1$. We claim that

$$[u]_\alpha \lesssim [\![u]\!]_\alpha\,. \tag{93}$$

This claim is established below, but first we proceed to prove Corollary 1 assuming that (93) holds.

To this end, fix $\alpha<\alpha'<\frac{1}{2}$. From (9) we get for any $R\in(0,1)$ and $1\leq\sigma<\infty$

$$\begin{aligned}
&\Big\langle I\Big(\frac{1}{R^\alpha}D(u,R)\geq\sigma\Big)\Big\rangle\\
&=\Big\langle I\Big(\exp\Big(\big(\frac{1}{R^{\alpha'}}D(u,R)\big)^{2(1-\epsilon)}\Big)\geq\exp\Big(\big(\sigma R^{\alpha-\alpha'}\big)^{2(1-\epsilon)}\Big)\Big)\Big\rangle\\
&\leq\exp\Big(-\big(\sigma R^{\alpha-\alpha'}\big)^{2(1-\epsilon)}\Big)\Big\langle\exp\Big(\big(\frac{1}{R^{\alpha'}}D(u,R)\big)^{2(1-\epsilon)}\Big)\Big\rangle\\
&\overset{(9)}{\lesssim}\exp\Big(-\big(\sigma R^{\alpha-\alpha'}\big)^{2(1-\epsilon)}\Big)\\
&\lesssim\exp\Big(-\frac{\sigma^{2(1-\epsilon)}}{C}-\frac{1}{C}R^{2(1-\epsilon)(\alpha-\alpha')}\Big)\,,
\end{aligned}\tag{94}$$

for a suitable constant C depending only on ϵ. In the third line we have used Chebyshev's inequality. By translation invariance the same bounds holds if u is replaced by $u^{(\bar t,\bar x)}$ for any space-time point $(\bar t,\bar x)$. Therefore, we get

$$\begin{aligned}
\Big\langle I\big([\![u]\!]_\alpha\geq\sigma\big)\Big\rangle &\leq \sum_R\sum_{(\bar t,\bar x)\in\Lambda_R}\Big\langle I\Big(\frac{1}{R^\alpha}D(u^{(\bar t,\bar x)},R)\geq\sigma\Big)\Big\rangle\\
&\overset{(94),(92)}{\lesssim}\exp\Big(-\frac{\sigma^{2(1-\epsilon)}}{C}\Big)\sum_R R^{-3}\exp\Big(-\frac{1}{C}R^{2(1-\epsilon)(\alpha-\alpha')}\Big)\\
&\lesssim\exp\Big(-\frac{\sigma^{2(1-\epsilon)}}{C}\Big)\,.
\end{aligned}$$

This fast decay of the tails of the distribution of the $[\![u]\!]_\alpha$ implies the desired integrability property.

It remains to establish the bound (93). We rely on Campanato's characterization of Hölder spaces [9, Theorem 5.5] which in our current context states that $[u]_\alpha$ is controlled by

$$\sup_{r<\frac{1}{2}}\sup_{(t_0,x_0)\in[-1,1]\times[-1,1]}\frac{1}{r^\alpha}\Big(\fint_{-r^2}^0\fint_{-r}^r\Big(u^{(t_0,x_0)}-\fint_{-r^2}^0\fint_{-r}^r u^{(t_0,x_0)}\Big)^2\Big)^{\frac{1}{2}}\,. \tag{95}$$

To see that $[\![u]\!]_\alpha$ controls this norm, we observe that for $r > 0$ satisfying $2^{-N-2} < r \leq 2^{-N-1}$ any arbitrary $(t_0, x_0) \in [-1,1] \times [-1,1]$ can be well approximated in $\Lambda_{2^{-N}}$, in the sense that there exists $(\bar{t}, \bar{x}) \in \Lambda_{2^{-N}}$ satisfying $|x_0 - \bar{x}| \leq 2^{-(N+1)}$ and $|t_0 - \bar{t}| \leq 2^{-2(N+1)}$. Then we get, for $R = 2^{-N}$ using the definition of η_R

$$
\begin{aligned}
&\frac{1}{r^{2\alpha}} \fint_{-r^2}^{0} \fint_{-r}^{r} \Big(u^{(t_0,x_0)} - \fint_{-r^2}^{0} \fint_{-r}^{r} u^{(t_0,x_0)} dxdt \Big)^2 dxdt \\
&\lesssim \frac{1}{R^{2\alpha}} \fint_{-R^2}^{0} \int \Big(\eta_R u^{(\bar{t},\bar{x})} - \fint_{-r^2}^{0} \fint_{-r}^{r} u^{(t_0,x_0)} dxdt \Big)^2 dxdt \\
&\lesssim \frac{1}{R^{2\alpha}} D(u^{(\bar{t},\bar{x})}, R)^2 \\
&\qquad + \frac{1}{R^{2\alpha}} \Big(\fint_{-r^2}^{0} \fint_{-r}^{r} \Big(u^{(t_0,x_0)} - \fint_{-R^2}^{0} \int \eta_R u^{(\bar{t},\bar{x})} dxdt \Big) dxdt \Big)^2 \\
&\lesssim \frac{1}{R^{2\alpha}} D(u^{(\bar{t},\bar{x})}, R)^2 \,.
\end{aligned}
$$

Therefore, we can conclude that $[\![u]\!]_\alpha$ controls the Campanato norm defined in (95) and the proof of (93) (and therefore the proof of Corollary 1) is complete.

References

[1] S. N. Armstrong and C. K. Smart, Quantitative stochastic homogenization of convex integral functionals. ArXiv e-prints (2014).

[2] M. Avellaneda and F.-H. Lin, Compactness methods in the theory of homogenization. *Comm. Pure Appl. Math.* **40** (1987), no. 6, 803–847.

[3] A. Debussche, S. de Moor, M. Hofmanová, A regularity result for quasilinear stochastic partial differential equations of parabolic type. *SIAM J. Math. Anal.*, **47** (2015), no. 2, 1590-1614

[4] J. Conlon, private communication (2015)

[5] J. Fischer and F. Otto, A higher-order large-scale regularity theory for random elliptic operators. ArXiv e-prints (2015).

[6] B. Gess, Finite speed of propagation for stochastic porous media equations. *SIAM J. Math. Anal.* **45** (2013), no. 5, 2734-2766

[7] A. Gloria, S. Neukamm, and F. Otto, A regularity theory for random elliptic operators. ArXiv e-prints (2014).

[8] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities. Séminaire de Probabilités, XXXIII, 120216, *Lecture Notes in Math.*, **1709**, Springer, Berlin, 1999

[9] M. Giaquinta, L. Martinazzi, An introduction to the regularity theory for elliptic systems, harmonic maps and minimal graphs. Appunti. Scuola Normale Superiore di Pisa (Nuova Serie), 2. Edizioni della Normale, Pisa, 2005. xii+302 pp.

[10] D. Marahrens and F. Otto, Annealed estimates on the Green function, ArXiv e-prints (2013), online in *Probability Theory and Related Fields.*

[11] A. Naddaf and T. Spencer, On homogenization and scaling limit of some gradient perturbations of a massless free field. *Comm. Math. Phys.*, **183** (1997), no. 1, 55–84.