

# GENERALIZED COUPLINGS AND ERGODIC RATES FOR SPDES AND OTHER MARKOV MODELS

BY OLEG BUTKOVSKY<sup>1</sup>, ALEXEI KULIK<sup>2</sup> AND MICHAEL SCHEUTZOW<sup>3</sup>

<sup>1</sup>Weierstrass Institute, [oleg.butkovskiy@gmail.com](mailto:oleg.butkovskiy@gmail.com)

<sup>2</sup>Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, [kulik.alex.m@gmail.com](mailto:kulik.alex.m@gmail.com)

<sup>3</sup>Institut für Mathematik, Technische Universität Berlin, [ms@math.tu-berlin.de](mailto:ms@math.tu-berlin.de)

We establish verifiable general sufficient conditions for exponential or subexponential ergodicity of Markov processes that may lack the strong Feller property. We apply the obtained results to show exponential ergodicity of a variety of nonlinear stochastic partial differential equations with additive forcing, including 2D stochastic Navier–Stokes equations. Our main tool is a new version of the generalized coupling method.

**1. Introduction.** There are many Markov processes that are not strong Feller and whose transition probabilities are mutually singular. Their ergodic properties cannot be analyzed with classical methods from, for example, [33] and require special treatment. One of the first papers in this direction was [23]; the results obtained there have been later generalized and extended in [6], [12], [26], Chapter 4. These works provide sufficient conditions for exponential or subexponential convergence of transition probabilities of a Markov process toward its invariant measure in the Wasserstein metric.

Unfortunately, checking these conditions in practice turns out to be rather difficult. One of the main results of this paper (Theorem 2.6) provides a set of *verifiable* sufficient conditions for exponential or subexponential ergodicity. Furthermore, we develop a special framework to ease the application of these conditions to stochastic partial differential equations (SPDEs). To illustrate the applicability of our framework, we establish exponential ergodicity of five important nonlinear SPDE models, including the 2D stochastic Navier–Stokes equation. This generalizes the corresponding theorems from [18] and [27], Section 6. To obtain these results, we develop a new version of the *generalized coupling* method, extending the ideas of [31] and [19].

Let us recall that the classical approach to the analysis of ergodic properties of Markov processes ([33, 34]) combines a local mixing condition on a certain set (a *small set*) with a recurrence condition that the Markov process visits this set often enough. If these conditions are satisfied, then the Markov process has a unique invariant probability measure (IPM) and the transition probabilities converge to the IPM in total variation; the rate of convergence is determined by the recurrence rate, that is, how quickly the Markov process returns to the small set. This can be quantified in terms of a Lyapunov function or the moments of the first return times.

The classical approach works quite well for strong Feller Markov processes such as stochastic differential equations with uniformly elliptic noise, MCMC in finite dimensions, nonlinear autoregressive models, etc. If a Markov process lacks the strong Feller property, then usually the local mixing condition does not hold. Possible examples of processes with “bad” mixing properties include many infinite-dimensional Markov processes such as SPDEs, stochastic functional differential equations (SFDEs), MCMC in infinite dimensional

Hilbert spaces, Markov processes with switching, etc. Typically, the transition probabilities of these processes with different initial points are mutually singular, and thus ergodicity in total variation clearly fails. For just one example of such a system, we refer to [38], where the *intrinsic memory* effect was observed for an SFDE: once the value of the segment process is known at some time moment  $t$ , the values at all the previous time moments are uniquely and measurably defined; see also the discussions in [23] and [26], Section 4.1. Clearly, the intrinsic memory effect yields mutual singularity of transition probabilities for the Markov process; in addition, such an effect prevents the mixing property of the system, which is a prerequisite for the classical approach.

To overcome this difficulty and to analyze the ergodic behavior of Markov processes without the strong Feller property, we adopt the method developed by M. Hairer, J. Mattingly and M. Scheutzow in [23]. The method is based on the concept of a *d-small set*, which is a much more general object than a small set; see Section 2 below for the precise definition. The strategy in a sense mimics the classical one; namely, it shows that if a Markov process visits a *d*-small set often enough, then, under some additional constraint on a Markov kernel (the so-called *nonexpansion property*), there exists a unique IPM and the transition probabilities weakly converge to the IPM with a rate that is quantified by means of a properly chosen probability metric; again the convergence rate is determined by the recurrence properties of the chain. For further extensions and generalizations within this approach, we refer to [6], [12], [26], Chapter 4.

The crucial step in the above approach is to construct a *distance-like* function  $d$  such that the Markov kernel indeed has the nonexpansion property w.r.t.  $d$  and that a certain set is indeed *d*-small. This step can be quite challenging; we refer to [23], Section 5.2, where such a construction is provided for an SFDE, though in a complicated and nontransparent way, which makes further extensions and modifications very difficult. One of the main results of our paper, Theorem 2.4, provides a transparent and fairly simple algorithm for the construction of a distance-like function  $d$  with required properties. We strongly believe this to be a substantial complement to the approach initiated in [23] and developed in [6], [12], [26], which hopefully makes this method well applicable for a wide variety of Markov systems.

The latter assertion is supported by several particular Markov models, analyzed in the second part of the paper. First, as a test case for our approach we reconsider SFDEs. We obtain a simple and transparent proof of their weak ergodicity at exponential, subexponential, and polynomial rates. Second, we consider five particular nonlinear SPDE models where the analysis of the ergodic properties is substantially more difficult due to the presence of strongly singular terms, such as the nonlinear gradient term  $(\mathbf{u} \cdot \nabla)\mathbf{u}$  in the Navier–Stokes equation. Using the general algorithm developed in the first part of the paper, we manage to overcome these difficulties, and establish exponential ergodicity in each of these five models. This in particular extends the results from [18] and [27], Section 6, where unique ergodicity and weak convergence to the IPM without a specified rate are proved for the same models.

Our approach is based on the concept of a *generalized coupling*, which we briefly explain here. By a *coupling*, one usually means a pair of stochastic processes with prescribed laws of the components; that is, the marginal laws. Heuristically, by a *generalized coupling* we will mean a pair of stochastic processes, whose marginal laws are not necessarily equal to a prescribed pair of probability distributions, but are in a sense close to this pair. Under the name *asymptotic coupling*, this concept was introduced in [23] in the spirit of earlier works [19, 31, 32], and was later developed in [18]. Originally, the marginal laws of the asymptotic coupling were assumed to be absolutely continuous with respect to the prescribed probability distributions; this, in a combination with the Birkhoff theorem, appears to be a convenient tool for proving unique ergodicity. Moreover, it was shown in [27] that the same notion can be used to guarantee weak stabilization of the transition probabilities; since this notion was used in a nonasymptotic way, the name was changed there to a *generalized coupling*.

In the paper, we further develop the above ideas and introduce a new definition of a generalized coupling. We require certain nonasymptotic total variation bounds between the marginal distributions and the prescribed laws to hold true; see Section 2.2. We show that the existence of a “good” generalized coupling makes it easy to construct a nonexpanding distance-like function  $d$  with a large class of  $d$ -small sets; this, combined with recurrent-type conditions, provides ergodic rates. Thus there are two important advantages of our method compared to previous techniques. On the one hand, generalized couplings with the required properties can be constructed quite efficiently using stochastic-control-type arguments, which makes the entire approach quite flexible and easy to apply. On the other hand, unlike previous results based on asymptotic/generalized couplings, our method provides explicit bounds on convergence rates, rather than just uniqueness of the IPM or weak convergence of transition probabilities. Let us mention briefly that the ergodic rates for a Markov process have further natural applications to limit theorems for additive functionals of the process and other, more complicated systems with Markov perturbations. In this context, there is a clearly seen difference between the stabilization of transition probabilities (with nonspecified rate) on the one hand, and explicit ergodic rates on the other hand. The first one typically guarantees the law of large numbers (or more generally, *averaging principle*), while the second one gives rise to the central limit theorem and *diffusion approximation*-type theorems; see [26], Section 5.4, comment 2 and Remark 6.3.9. For further discussion, references and systematic treatment of the limit theorems for systems with weakly ergodic Markov perturbations we refer to [26], Chapters 5, 6.

The rest of the paper is organized as follows. In Section 2.1, we recall well-known results concerning the convergence of Markov processes. We present our main results in Section 2.2. In Section 3, we give an application to SFDEs. In Section 4, we develop a general framework for proving exponential ergodicity for nonlinear SPDEs and establish exponential ergodicity of a number of important SPDEs.

**2. A general setup for ergodicity and estimating convergence rates.** This section consists of two parts. In the first part, we introduce the basic notation and recall some useful definitions and important theorems. In the second part, we present our main results.

2.1. *Generalities and notation.* Let  $(E, \rho)$  be a Polish space and  $\mathcal{E} = \mathcal{B}(E)$  be the corresponding Borel  $\sigma$ -field. Consider a Markov transition function  $\{P_t(x, A), x \in E, A \in \mathcal{E}\}_{t \in \mathbb{R}_+}$ . We use the standard notation for the corresponding semigroup of integral operators

$$P_t \varphi(x) = \int_E \varphi(y) P_t(x, dy), \quad x \in E, t \in \mathbb{R}_+.$$

We also denote by  $\{P_x, x \in E\}$  the corresponding Markov family, that is,  $P_x$  denotes the law of the Markov process  $X = \{X_t, t \geq 0\}$  with  $X_0 = x$  and the given transition function. The law of  $X$  will be understood in the sense of finite-dimensional distributions; that is, we will not rely on the trajectory-wise properties of  $X$ .

A function  $d : E \times E \rightarrow \mathbb{R}_+$  is called *distance-like* if it is symmetric, lower semicontinuous, and  $d(x, y) = 0 \Leftrightarrow x = y$ ; see [23], Definition 4.3. Denote by  $\mathcal{P}(E)$  the set of all probability measures on  $(E, \mathcal{E})$ , and for  $\mu, \nu \in \mathcal{P}(E)$  denote by  $\mathcal{C}(\mu, \nu)$  the set of all *couplings* between  $\mu$  and  $\nu$ , that is, probability measures on  $(E \times E, \mathcal{E} \otimes \mathcal{E})$  with marginals  $\mu$  and  $\nu$ . For a given distance-like function  $d$ , the corresponding *coupling distance*  $W_d : \mathcal{P}(E) \times \mathcal{P}(E) \rightarrow \mathbb{R}_+ \cup \{\infty\}$  is defined by

$$W_d(\mu, \nu) := \inf_{\lambda \in \mathcal{C}(\mu, \nu)} \int_{E \times E} d(x, y) \lambda(dx, dy), \quad \mu, \nu \in \mathcal{P}(E).$$

If  $d$  is a lower semicontinuous metric on  $E$ , then  $W_d$  is the usual *Wasserstein-1* (or *Kantorovich*) distance. In particular, if  $d$  is the discrete metric, that is,  $d(x, y) = \mathbb{1}(x \neq y)$ , then  $W_d$  coincides with the *total variation* distance  $d_{\text{TV}}$ ; the latter can also be defined as follows:

$$d_{\text{TV}}(\mu, \nu) := \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

Recall that if the original metric  $\rho$  is bounded, then convergence in total variation implies convergence in  $W_\rho$ ; furthermore, in this case convergence in  $W_\rho$  is equivalent to weak convergence.

As explained in the [Introduction](#), if a Markov process lacks the strong Feller property, then the total variation metric is too strong to measure the distance between the transition probabilities of the process and its invariant measure; this is the reason for us to focus on the study of weaker distances  $W_d$ . As mentioned above, [23] and subsequent works [6], [12], [26], Chapter 4, provide sufficient conditions for convergence of the law of a Markov process to its invariant measure in  $W_d$  for properly chosen  $d$ . For the convenience of the reader, we formulate here two results of such type, which actually motivate the form in which our main results are presented below. For additional details of the theory, discussions and examples, we refer to the above mentioned references.

Define for a measurable function  $\varphi: \mathbb{R}_+ \rightarrow (0, \infty)$

$$(2.1) \quad H_\varphi(x) := \int_1^x \frac{1}{\varphi(u)} du, \quad x \geq 1.$$

**PROPOSITION 2.1** ([6], Theorem 2.4, [26], Theorem 4.5.2). *Assume that the Markov semigroup  $P$  is Feller. Suppose there exist a measurable function  $V: E \rightarrow [0, \infty)$  and a bounded distance-like function  $d$  on  $E$  such that the following conditions hold:*

1.  *$V$  satisfies the Lyapunov condition, that is, there exist a concave differentiable function  $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  increasing to infinity with  $\varphi(0) = 0$  and a constant  $K > 0$  such that for any  $t \geq 0$ ,  $x \in E$*

$$(2.2) \quad P_t V(x) \leq V(x) - \int_0^t P_s(\varphi \circ V)(x) ds + Kt.$$

2. *One has  $\rho \leq d$ , where  $\rho$  is the original metric on the Polish space  $E$ .*

3. *There exist  $0 < t_1 < t_2 < \infty$  such that for all  $t \in [t_1, t_2]$ ,*

$$(2.3) \quad W_d(P_t(x, \cdot), P_t(y, \cdot)) \leq d(x, y), \quad x, y \in E.$$

4. *There exists  $t > 0$  such that for any  $M > 0$  there exists  $\varepsilon = \varepsilon(M, t) > 0$  such that*

$$(2.4) \quad W_d(P_t(x, \cdot), P_t(y, \cdot)) \leq (1 - \varepsilon) d(x, y), \quad x, y \in \{V \leq M\}.$$

*Then the Markov semigroup  $(P_t)$  has a unique invariant measure  $\pi$ . Moreover, for any  $\delta \in (0, 1)$  there exist constants  $C_1, C_2 > 0$  such that for any  $x \in E$ ,*

$$W_d(P_t(x, \cdot), \pi) \leq \frac{C_1(1 + \varphi(V(x)))^\delta}{\varphi(H_\varphi^{-1}(C_2 t))^\delta}, \quad t \geq 0.$$

If the function  $V$  satisfies a stronger version of the Lyapunov inequality, then condition 4 of the above proposition can be relaxed. More precisely, the following statement holds.

**PROPOSITION 2.2** ([23], Theorem 4.8). *Assume that condition 1 of Proposition 2.1 holds with  $\varphi(x) = \gamma x$ ,  $\gamma > 0$ . Then condition 4 of Proposition 2.1 can be replaced with the following weaker condition:*

4\*. *There exist  $\varepsilon > 0$ ,  $t > 0$  such that inequality (2.4) is satisfied for all  $x, y \in \{V \leq 4K/\gamma\}$ .*

Thus, the difference between the assumptions in the exponential and the subexponential cases is the following. In the exponential case, it is enough to show that for any  $M > 0$  there exists  $t > 0$  such that the level set  $\{V \leq M\}$  satisfies inequality (2.4). On the other hand, in the subexponential case one cannot pick  $t$  depending on  $M$ ; one has to check inequality (2.4) for some fixed time  $t$  simultaneously for *all*  $M$ . We will see later in Section 2.2 that this difference will become crucial and we will present different sets of assumptions for the subexponential and exponential cases.

REMARK 2.3. If it is already known that the semigroup  $P$  has an invariant measure, then condition 2 of Proposition 2.1 can be omitted, see [23], Theorem 4.8.

A convenient tool to verify (2.3) and (2.4) is provided by the following pair of notions.

DEFINITION 2.1 ([23], Definition 4.6). A distance-like function  $d$  bounded by 1 is called *contracting* for  $P_t$  if there exists  $\alpha < 1$  such that for any  $x, y \in E$  with  $d(x, y) < 1$  we have

$$W_d(P_t(x, \cdot), P_t(y, \cdot)) \leq \alpha d(x, y).$$

DEFINITION 2.2 ([23], Definition 4.4). A set  $B \subset E$  is called *d-small* for  $P_t$  if for some  $\varepsilon > 0$

$$\sup_{x, y \in B} W_d(P_t(x, \cdot), P_t(y, \cdot)) \leq 1 - \varepsilon.$$

Clearly, if  $d$  is contracting for  $P_t$  and the set  $\{V \leq M\}$  is  $d$ -small for  $P_t$ , then inequalities (2.3) and (2.4) are satisfied.

Let us briefly mention a closely related concept of a *coarse Ricci curvature* [35], Definition 3, introduced recently in somewhat different framework, mainly focused at functional inequalities. In the terms of [35], conditions 3 and 4 of Proposition 2.1 can be reformulated as follows: the Ricci curvature of the whole space should be nonnegative and the Ricci curvature of any level set of  $V$  should be positive.

2.2. *Main results.* Propositions 2.1 and 2.2 allow one to bound the rate of convergence of a Markov process to its invariant measure in the Wasserstein metric. The crucial question within this context is how to choose the distance-like function  $d$  in a way that conditions 3 and 4 (or 4\*) are satisfied. The main goal of this section is to provide an algorithm of such construction, which relies on explicit and easy-to-verify set of assumptions.

We will call a function  $\theta : E \times E \rightarrow \mathbb{R}_+$  a *premetric* if it is lower semicontinuous and  $\theta(x, y) = 0 \Leftrightarrow x = y$ ; we fix a premetric  $\theta$  on  $E$  until the end of this section. Our first assumption will serve as a replacement for the contractivity condition.

ASSUMPTION A. There exist a nonincreasing function  $r : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\lim_{t \rightarrow \infty} r(t) = 0$  and a locally bounded function  $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that for any  $x, y \in E$ , there exist random processes  $X^{x,y} = (X_t^{x,y})_{t \geq 0}$ ,  $Y^{x,y} = (Y_t^{x,y})_{t \geq 0}$  with the following properties:

1.  $\text{Law}(X^{x,y}) = P_x$ , and

$$d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) \leq L(t)\theta(x, y), \quad t \geq 0;$$

2.  $E\theta(X_t^{x,y}, Y_t^{x,y}) \leq r(t)\theta(x, y)$ ,  $t \geq 0$ .

Assumption **A.1** means that the pair  $(X^{x,y}, Y^{x,y})$  has the sense of a generalized coupling: its first component is distributed as  $P_x$  and the defect between the law of the second component at time  $t$  and the “true law”  $P_t(y, \cdot)$  can be effectively controlled. The function  $L$  here plays the role of a Lipschitz coefficient. We do not suppose that  $L$  is decreasing to 0 at infinity; moreover, we allow the situation when  $L(t) \rightarrow \infty$  as  $t \rightarrow \infty$ , that is, the defect between the generalized coupling and the “true law” might be quite large. Assumption **A.2** controls the decay rate of the premetric  $\theta$  between the components of this generalized coupling. Note also that we are not imposing any additional conditions on the relation between  $P_y$  and  $\text{Law}(Y^{x,y})$ ; in particular we are not assuming any absolute continuity.

The second assumption will replace the  $d$ -small property.

**ASSUMPTION B1.** There exist a set  $B \subset E$  and  $t_0 > 0$  such that for any  $\varepsilon > 0$  there exists a set  $D \in \mathcal{E}$  such that:

1.  $\inf_{x \in B} P_{t_0}(x, D) > 0$ ;
2.  $\sup_{x, y \in D} \theta(x, y) \leq \varepsilon$ .

We note that a similar in spirit assumption can be found in the literature; see, for example, [22], Corollary 2.2. Assumption **B1** is not of a “generalized-coupling-type”; it is essentially a support condition. However, in the exponential case due to the reasons discussed after Proposition 2.2, it can be replaced by a “generalized-coupling-type” assumption that may be easier to check in some setups.

**ASSUMPTION B2.** There exist a set  $B \subset E$ , a nonincreasing function  $R : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\lim_{t \rightarrow \infty} R(t) = 0$ , and  $\varepsilon > 0$  such that for any  $x, y \in B$ , there exist random processes  $X^{x,y} = (X_t^{x,y})_{t \geq 0}$ ,  $Y^{x,y} = (Y_t^{x,y})_{t \geq 0}$  with the following properties:

1.  $\text{Law}(X^{x,y}) = P_x$ , and
- $$(2.5) \quad d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) \leq 1 - \varepsilon, \quad t \geq 0, x, y \in B;$$
2.  $\mathbb{E}\theta(X_t^{x,y}, Y_t^{x,y}) \leq R(t)$ ,  $t \geq 0$ ,  $x, y \in B$ .

Assumption **B2** means that there exists a “good” set  $B \subset E$ , where we have a family of generalized couplings with especially nice properties. Namely, the first component of the generalized coupling coincides with the “true law,” and the distance between the second component at time  $t$  and the corresponding “true law” at time  $t$  is uniformly bounded on  $B$ .

Now for any fixed  $N > 0$  we consider the following distance-like function:

$$(2.6) \quad d_N(x, y) := N\theta(x, y) \wedge N\theta(y, x) \wedge 1, \quad x, y \in E.$$

Our first main result shows how the  $d$ -smallness and contractivity conditions can be verified under the assumptions **A**, **B1** and **B2**.

#### THEOREM 2.4.

(i) Assume that **A** holds for some functions  $r, L$ . Then for any  $N > 0$ ,  $t > 0$  such that  $r(t) \leq 1/3$  and  $N \geq 2L(t)$  the distance-like function  $d_N$  is contracting for  $P_t$  (in the sense of Definition 2.1).

(ii) Assume that **B1** holds for some  $t_0 > 0$  and a set  $B \subset E$ . Then for any  $N > 0$  the set  $B$  is  $d_N$ -small for  $P_{t_0}$  (in the sense of Definition 2.2).

(iii) Assume that **B2** holds for some function  $R$ , a set  $B \subset E$  and  $\varepsilon > 0$ . Then for any  $N > 0$ ,  $t > 0$  such that  $NR(t) \leq \varepsilon/2$  the set  $B$  is  $d_N$ -small for  $P_t$ .

PROOF. (i). Fix  $t > 0$ ,  $N > 0$  that satisfy the assumptions of the theorem. Take any  $x, y \in E$  such that  $d_N(x, y) < 1$ . Without loss of generality, suppose that  $\theta(x, y) \leq \theta(y, x)$ .

We begin by observing that by Assumption **A.1** and the coupling lemma (see, e.g., [42], Theorem 4.1), there exists a pair of random variables  $\widehat{Y}$ ,  $Z$  such that  $\text{Law}(\widehat{Y}) = \text{Law}(Y_t^{x,y})$ ,  $\text{Law}(Z) = P_t(y, \cdot)$  and

$$\mathbb{P}(\widehat{Y} \neq Z) = d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) \leq L(t)\theta(x, y).$$

By the gluing lemma (see, e.g., [26], Lemma 4.3.2, and [42], p. 23), we can assume that  $\widehat{Y}$ ,  $Z$  are defined on the same probability space with  $X_t^{x,y}$ ,  $Y_t^{x,y}$ , and  $\widehat{Y} = Y_t^{x,y}$ . Then the joint law of  $X_t^{x,y}$ ,  $Z$  is a coupling for  $P_t(x, \cdot)$ ,  $P_t(y, \cdot)$ , and thus

$$\begin{aligned} W_{d_N}(P_t(x, \cdot), P_t(y, \cdot)) &\leq \mathbb{E}d_N(X_t^{x,y}, Z) \\ &= \mathbb{E}d_N(X_t^{x,y}, Z)\mathbb{1}(Y_t^{x,y} = Z) + \mathbb{E}d_N(X_t^{x,y}, Z)\mathbb{1}(Y_t^{x,y} \neq Z) \\ &\leq \mathbb{E}d_N(X_t^{x,y}, Y_t^{x,y}) + \mathbb{P}(Y_t^{x,y} \neq Z) \\ &\leq N\mathbb{E}\theta(X_t^{x,y}, Y_t^{x,y}) + L(t)\theta(x, y) \\ &\leq N\theta(x, y)(r(t) + L(t)N^{-1}) \\ &\leq \frac{5}{6}N\theta(x, y) = \frac{5}{6}d_N(x, y), \end{aligned}$$

where the last equality follows from the assumptions  $d_N(x, y) < 1$  and  $\theta(x, y) \leq \theta(y, x)$ . This proves the contraction property.

(ii). Fix  $N > 0$ . In this case, we show the  $d_N$ -smallness of  $P_{t_0}$  using the independent coupling, similarly to [23], Proposition 5.3. Take  $\varepsilon := 1/(2N)$ . By Assumption **B1**, there exists a set  $D \in \mathcal{E}$  such that  $\delta := \inf_{x \in B} P_{t_0}(x, D) > 0$  and  $\sup_{x, y \in D} \theta(x, y) \leq \varepsilon$ .

Now we fix  $x, y \in B$  and construct independent random variables  $X, Y$  such that

$$\text{Law}(X) = P_{t_0}(x, \cdot); \quad \text{Law}(Y) = P_{t_0}(y, \cdot).$$

Then

$$\begin{aligned} W_{d_N}(P_{t_0}(x, \cdot), P_{t_0}(y, \cdot)) &\leq \mathbb{E}d_N(X, Y) \\ &\leq \mathbb{P}(\{X \notin D\} \cup \{Y \notin D\}) + N\varepsilon\mathbb{P}(X \in D, Y \in D) \\ &\leq 1 - \delta^2(1 - N\varepsilon) \leq 1 - \frac{1}{2}\delta^2, \end{aligned}$$

showing that  $B$  is  $d_N$ -small.

(iii). We establish the  $d_N$ -smallness of  $P_t$  by an argument, similar to the one used in the proof of part (i) of the lemma. We fix  $t > 0$ ,  $N > 0$ ,  $x, y \in B$ . Using the same notation and the same construction as in the first part of the proof, we obtain a triple of random variables  $X_t^{x,y}$ ,  $Y_t^{x,y}$ ,  $Z$  such that

$$\begin{aligned} \text{Law}(X_t^{x,y}) &= P_t(x, \cdot); & \text{Law}(Z) &= P_t(y, \cdot); \\ \mathbb{E}\theta(X_t^{x,y}, Y_t^{x,y}) &\leq R(t); & \mathbb{P}(Y_t^{x,y} \neq Z) &\leq 1 - \varepsilon. \end{aligned}$$

Hence, using the fact that  $d_N \leq N\theta$ , we deduce

$$\begin{aligned} W_{d_N}(P_t(x, \cdot), P_t(y, \cdot)) &\leq \mathbb{E}d_N(X_t^{x,y}, Z) \leq \mathbb{E}d_N(X_t^{x,y}, Y_t^{x,y}) + \mathbb{P}(Y_t^{x,y} \neq Z) \\ &\leq NR(t) + 1 - \varepsilon \\ &\leq \varepsilon/2 + 1 - \varepsilon = 1 - \varepsilon/2, \end{aligned}$$

where the fourth inequality follows from the fact that  $NR(t) \leq \varepsilon/2$ . This proves the  $d_N$ -smallness of  $B$ .  $\square$

**REMARK 2.5.** We will see in Sections 3 and 4 that in various cases generalized couplings with the required properties can be constructed using a stochastic control approach. Then our entire argument will consist of two principal steps:

- (i) using a stochastic control technique, we construct a generalized coupling  $X^{x,y}, Y^{x,y}$ , which exhibits certain contraction properties (Assumptions **A.2**, **B2.2**)
- (ii) using the error-in-law bounds for this generalized coupling (Assumptions **A.1**, **B2.1**), we construct then a true coupling  $X_t^{x,y}, Z$  for  $P_t(x, \cdot), P_t(y, \cdot)$ , which confirms the required contraction property of  $d_N$ .

We will call this type of argument a *control-and-reimburse* strategy. The same general idea—to apply the stochastic control in order to improve the system, and then to take into account the impact of the control—was actually used, in more implicit form, in the previously mentioned construction [23], Section 5.2, of the distance-like function  $d(x, y)$  for an SFDE, and in the approach to the study of weak ergodicity of SPDEs developed in [19]. A similar argument was used to prove ergodicity in total variation in [1] for degenerate diffusions, and in [4] for solutions to Lévy driven SDEs. Related ideas were used to establish the Harnack inequality for SDEs and SFDEs [13, 44].

Our next result establishes unique ergodicity under Assumptions **A** and **B1**. It is a combination of Theorem 2.4 and Proposition 2.1. Recall the definition of  $H_\varphi$  in (2.1).

**THEOREM 2.6.** *Assume that the Markov semigroup  $P$  is Feller. Suppose that:*

1. *There exists a measurable function  $V$  that satisfies condition 1 of Proposition 2.1.*
2.  *$\rho \wedge 1 \leq \theta$ , where  $\rho$  is the original metric on  $E$ .*
3. *Assumption **A** holds for functions  $r, L$ .*
4. *There exists  $t_0 > 0$  such that for any  $M > 0$  Assumption **B1** holds for  $B = \{V \leq M\}$  and  $t_0$ .*

*Then the Markov semigroup  $(P_t)$  has a unique invariant measure  $\pi$ . Moreover, for any  $\delta \in (0, 1)$  there exist constants  $C_1, C_2 > 0$  such that for any  $x \in E$ ,*

$$(2.7) \quad W_{\rho \wedge 1}(P_t(x, \cdot), \pi) \leq \frac{C_1(1 + \varphi(V(x))^\delta)}{\varphi(H_\varphi^{-1}(C_2 t))^\delta}, \quad t \geq 0.$$

**PROOF OF THEOREM 2.6.** Let us check that all the conditions of Proposition 2.1 are satisfied for the function  $V$  and a distance-like function  $d_N$  introduced in (2.6), where  $N \geq 1$  will be chosen later. The first condition of Proposition 2.1 is satisfied by the first assumption of the theorem. Since  $\rho \wedge 1 \leq \theta$ , we have  $\rho \wedge 1 \leq d_N$ , and thus the second condition of Proposition 2.1 is satisfied.

Now we pick  $t_* \geq 0$  such that  $r(t_*) \leq 1/3$ . This is possible thanks to Assumption **A**. Set  $N := (2 \sup_{t \in [t_*, t_* + t_0]} L(t)) \vee 1$ . Theorem 2.4(i) implies that for some  $\alpha < 1$  we have

$$(2.8) \quad W_{d_N}(P_{t_* + t_0}(x, \cdot), P_{t_* + t_0}(y, \cdot)) \leq \alpha d_N(x, y) \quad \text{whenever } d_N(x, y) < 1$$

and that the distance-like function  $d_N$  is contracting for  $P_t$  for any  $t \in [t_*, t_* + t_0]$ . This immediately implies that inequality (2.3) holds for any  $t \in [t_*, t_* + t_0]$ , and thus condition 3 of Proposition 2.1 is satisfied. In particular, we have

$$(2.9) \quad W_{d_N}(P_{t_*}(x, \cdot), P_{t_*}(y, \cdot)) \leq d_N(x, y), \quad x, y \in E.$$

To check condition 4 of Proposition 2.1, we take  $t := t_* + t_0$  and fix arbitrary  $M > 0$ . It follows from Theorem 2.4(ii) and Assumption 4 of the theorem that the set  $\{V \leq M\}$  is  $d_N$ -small for  $P_{t_0}$  with some  $\varepsilon = \varepsilon(t, M)$ . Now take any  $x, y \in \{V \leq M\}$ . If  $d_N(x, y) < 1$  then using (2.8), we get

$$W_{d_N}(P_t(x, \cdot), P_t(y, \cdot)) \leq \alpha d_N(x, y).$$

If  $d_N(x, y) = 1$ , then using (2.9) and  $d_N$ -small property, we derive

$$\begin{aligned} W_{d_N}(P_{t_0+t_*}(x, \cdot), P_{t_0+t_*}(y, \cdot)) &\leq W_{d_N}(P_{t_0}(x, \cdot), P_{t_0}(y, \cdot)) \\ &\leq 1 - \varepsilon = (1 - \varepsilon)d_N(x, y). \end{aligned}$$

Thus, condition 4 of Proposition 2.1 is met. Therefore, all conditions of Proposition 2.1 are satisfied, and hence the Markov semigroup  $(P_t)$  has a unique invariant measure  $\pi$  and (2.7) holds.  $\square$

**THEOREM 2.7.** *In the exponential case (i.e., when condition 1 of Proposition 2.1 holds for a linear function  $\varphi(x) = \gamma x$ ,  $\gamma > 0$ ) condition 4 of Theorem 2.6 can be replaced by the following assumption:*

4\*. *Assumption B2 holds for some function  $R$ , the level set of the Lyapunov function  $\{V \leq 4K/\gamma\}$ , and some  $\varepsilon > 0$ .*

**PROOF OF THEOREM 2.7.** We use essentially the same line of argument as in the proof of Theorem 2.6 with some modifications. Pick  $t_* \geq 0$  such that  $r(t_*) \leq 1/3$  and set  $N := (2 \sup_{t \in [t_*, 2t_*]} L(t)) \vee 1$ . We see again that conditions 1–3 of Proposition 2.1 hold for  $V$  and  $d_N$ . In particular, we have for some  $\alpha < 1$ ,

$$(2.10) \quad W_{d_N}(P_{t_*}(x, \cdot), P_{t_*}(y, \cdot)) \leq \alpha d_N(x, y) \quad \text{whenever } d_N(x, y) < 1.$$

Furthermore, it follows from Theorem 2.4(i) that for any  $t \in [t_*, 2t_*]$  we have

$$W_{d_N}(P_t(x, \cdot), P_t(y, \cdot)) \leq d_N(x, y), \quad x, y \in E.$$

This implies that for any  $t \geq t_*$

$$(2.11) \quad W_{d_N}(P_t(x, \cdot), P_t(y, \cdot)) \leq d_N(x, y), \quad x, y \in E.$$

Pick now  $t_0 \geq t_*$  such that  $NR(t_0) \leq \varepsilon/2$ . Then condition 4\* of the theorem, the definition of  $N$ , and Theorem 2.4(iii) imply now that the set  $B := \{V \leq 4K/\gamma\}$  is  $d_N$ -small for  $P_{t_0}$ .

Let us check now that condition 4\* of Proposition 2.2 holds for  $t = t_* + t_0$ . Take any  $x, y \in \{V \leq 4K/\gamma\}$ . If  $d_N(x, y) < 1$  then taking into account (2.10), (2.11), and the fact that  $t_0 \geq t_*$ , we get

$$W_{d_N}(P_{t_0+t_*}(x, \cdot), P_{t_0+t_*}(y, \cdot)) \leq W_{d_N}(P_{t_*}(x, \cdot), P_{t_*}(y, \cdot)) \leq \alpha d_N(x, y).$$

If  $d_N(x, y) = 1$ , then using (2.11) and  $d_N$ -small property, we derive

$$\begin{aligned} W_{d_N}(P_{t_0+t_*}(x, \cdot), P_{t_0+t_*}(y, \cdot)) &\leq W_{d_N}(P_{t_0}(x, \cdot), P_{t_0}(y, \cdot)) \\ &\leq 1 - \varepsilon = (1 - \varepsilon)d_N(x, y). \end{aligned}$$

Thus, condition 4\* of Proposition 2.2 is met.

Therefore, all conditions of Proposition 2.2 are satisfied, and thus the Markov semigroup  $(P_t)$  has a unique invariant measure  $\pi$  and (2.7) holds.  $\square$

**REMARK 2.8.** Taking into account Remark 2.3, we deduce that if it is already known that the semigroup  $P$  has an invariant measure  $\pi$ , then condition 2 of Theorem 2.6 can be dropped. In this case, the left-hand side of formula (2.7) is replaced by  $W_{\tilde{\theta}}(P_t(x, \cdot), \pi)$ , where  $\tilde{\theta}(x, y) := \theta(x, y) \wedge \theta(y, x) \wedge 1$ .

**3. Easy application: SFDEs.** In this section, we illustrate our method of generalized coupling by establishing a rate of convergence to the invariant measure for solutions to SFDEs. We show how Assumptions **A** and **B1** can be immediately verified for these processes. This can be regarded as a “warm-up” before the next section, which deals with SPDEs. Related ideas will be applied there in a more complicated setup and additional challenges will arise.

Let us introduce the model. Fix  $n, m \in \mathbb{N}$  and  $r > 0$ . Denote by  $\mathcal{C} = \mathcal{C}([-r, 0], \mathbb{R}^n)$  the space of continuous functions endowed with the supremum norm  $\|\cdot\|$ . For a matrix  $M \in \mathbb{R}^{d \times m}$ , we denote by  $\|M\|$  its Frobenius norm, that is,  $\|M\| := \sqrt{\sum M_{ij}^2}$ . For a real number  $a$ , let  $a_+ := \max(a, 0)$ .

Consider the following SFDE:

$$(3.1) \quad \begin{aligned} dX^x(t) &= f(X_t^x) dt + g(X_t^x) dW(t), \quad t \geq 0, \\ X_0^x &= x, \end{aligned}$$

where  $f : \mathcal{C} \rightarrow \mathbb{R}^n$  and  $g : \mathcal{C} \rightarrow \mathbb{R}^{n \times m}$  are measurable functions,  $W$  is a  $m$ -dimensional Brownian motion, the initial condition  $x \in \mathcal{C}$ , and we use the standard notation  $X_t(s) := X(t+s)$ ,  $s \in [-r, 0]$ .

Suppose that the function  $f$  is continuous and bounded on bounded subsets of  $\mathcal{C}$ . Further, suppose that  $f$  is one-sided Lipschitz and  $g$  is Lipschitz, that is, there exists  $C > 0$  such that for any  $x, y \in \mathcal{C}$ ,

$$\langle f(x) - f(y), x(0) - y(0) \rangle_+ + \|g(x) - g(y)\|^2 \leq C \|x - y\|^2.$$

It is known that under these assumptions equation (3.1) has a unique strong solution [43]. Moreover, this solution  $X = (X_t)_{t \geq 0}$  is a Feller Markov process with the state space  $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$ . Denote by  $(P_t)_{t \geq 0}$  the corresponding semigroup. Assume also the uniform nondegeneracy condition:

$$(3.2) \quad \sup_{x \in \mathcal{C}} \|g^{-1}(x)\| < \infty,$$

where  $g^{-1}(x)$  denotes a right inverse of the matrix  $g(x)$ .

The next theorem describes the convergence rate of  $\text{Law}(X_t)$  to its invariant measure in the Wasserstein metric.

**THEOREM 3.1.** *Assume that there exists a Lyapunov function  $V : \mathcal{C} \rightarrow \mathbb{R}_+$  that satisfies condition 1 of Proposition 2.1. Suppose that either (i) or (ii) holds:*

$$(i) \quad \lim_{\|x\| \rightarrow \infty} V(x) = +\infty.$$

(ii)  $\lim_{|x(0)| \rightarrow \infty} V(x) = +\infty$ . Assume additionally that the drift and diffusion of SFDE (3.1) satisfy the following growth condition: there exists  $C > 0$  such that for any  $x \in \mathcal{C}$ ,

$$(3.3) \quad \langle f(x), x(0) \rangle_+ + \|g(x)\|^2 \leq C + C|x(0)|^2.$$

Then SFDE (3.1) has a unique invariant measure  $\pi$ . Further,  $\text{Law}(X_t)$  converges to  $\pi$  in the Wasserstein metric  $W_{\|\cdot\| \wedge 1}$  and the rate of convergence is given by (2.7).

This theorem is essentially not new; it is only a mild generalization of [23], Section 5, and [6], Theorem 3.2. However, using the generalized coupling method, we managed to drastically simplify the key ingredient of the proof, namely, the verification of the contraction property for the solution of an SFDE. Since this method is model insensitive, similar arguments allow us to establish contraction properties of solutions of SPDEs; see Section 4. On the other hand, transferring the ideas used in [23], Section 5, from SFDEs to SPDEs or other

Markov models seems to be rather difficult (we mention though a recent preprint [2], where SFDEs with infinite memory are analyzed in a way quite similar to [23], Section 5).

Construction of a suitable Lyapunov function  $V$  that satisfies condition (2.2) for a specific SFDE is a completely independent task and is out of the scope of the paper; we refer here to [37], [23], Remark 5.2, [7], [6], Section 3.2, [26], Section 4.6.1, for possible methods of building  $V$ . Let us just briefly explain the difference between conditions (i) and (ii) of Theorem 3.1. Condition (ii) allows to consider a larger and more natural class of Lyapunov functions. For example, the function  $V(x) := |x(0)|^2$ ,  $x \in \mathcal{C}$  satisfies condition (ii) but not condition (i) of the theorem. The price to pay is that the drift and the diffusion of the SFDE have to satisfy a certain growth condition.

The proof of Theorem 3.1 is based on the following key lemma.

LEMMA 3.2.

(i) *There exist  $N_0 > 0$ ,  $t_0 > 0$  such that for any  $N \geq N_0$ ,  $t \geq t_0$  the distance*

$$d_N(x, y) := N \|x - y\| \wedge 1$$

*is contracting for  $P_t$ .*

(ii) *For any  $N > 0$ ,  $t \geq 2r$ ,  $M > 0$  the level set*

$$B_M := \{x \in \mathcal{C} : \|x\| \leq M\},$$

*is  $d_N$ -small for  $P_t$ .*

(iii) *Assume additionally that the drift and diffusion satisfy the growth condition (3.3). Then for any  $N > 0$ ,  $t \geq 3r$ ,  $M > 0$  the level set*

$$H_M := \{x \in \mathcal{C} : |x(0)| \leq M\}$$

*is  $d_N$ -small for  $P_t$ .*

PROOF OF LEMMA 3.2. The proofs of all three parts of the lemma are based on the verification of Assumptions **A** and **B1** for the Markov semigroup  $(P_t)_{t \geq 0}$  and applying then Theorem 2.4. In all the cases, we take  $E = \mathcal{C}$ ,  $\rho(x, y) = \theta(x, y) = \|x - y\|$ . It is clear that the space  $(\mathcal{C}, \rho)$  is Polish and  $\theta$  is a premetric.

(i). Let us check that  $(P_t)_{t \geq 0}$  satisfies Assumption **A**. Let  $\lambda > 0$  be a parameter to be chosen later. For each  $x, y \in \mathcal{C}$ , we consider the following generalized coupling. We put  $X^{x,y} = X^x$  and let  $Y^{x,y,\lambda}$  be the strong solution of the following equation:

$$\begin{aligned} dY^{x,y,\lambda}(t) &= f(Y_t^{x,y,\lambda}) dt + g(Y_t^{x,y,\lambda}) dW^{x,y,\lambda}(t), \quad t \geq 0, \\ Y_0^{x,y,\lambda} &= y, \end{aligned}$$

where

$$dW^{x,y,\lambda}(t) := \beta^{x,y,\lambda}(t) dt + dW(t),$$

and

$$\beta^{x,y,\lambda}(t) := \lambda g(Y_t^{x,y,\lambda})^{-1} (X^x(t) - Y^{x,y,\lambda}(t)).$$

The existence and uniqueness of  $Y^{x,y,\lambda}$  follows again from [43]. By [23], Lemma 3.6, there exists some  $\lambda_0 > 0$  that does not depend on  $x, y$  such that

$$(3.4) \quad \mathbb{E} \|X_t^x - Y_t^{x,y,\lambda_0}\|^2 \leq C e^{-2t} \|x - y\|^2, \quad t \geq 0.$$

From now on, we take  $\lambda = \lambda_0$  and denote  $Y^{x,y} := Y^{x,y,\lambda_0}$ ,  $\beta := \beta^{x,y,\lambda_0}$ . By construction,  $\text{Law}(X^{x,y}) = P_x$ . By Theorem A.2 and inequality (A.1), we have for any fixed  $t \geq 0$

$$\begin{aligned} d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) &\leq d_{\text{TV}}(\text{Law}(W^{x,y}(s), s \in [0, t]), \text{Law}(W(s), s \in [0, t])) \\ &\leq \left( \mathbf{E} \int_0^t |\beta(s)|^2 ds \right)^{1/2} \leq C_1 \lambda_0 \|x - y\|, \end{aligned}$$

for some  $C_1 > 0$ , where in the last step we have used (3.2) and (3.4). Thus condition A.1 is satisfied with  $L(t) = C_1 \lambda_0$ . Using again (3.4), we derive

$$\mathbf{E} \|X_t^{x,y} - Y_t^{x,y}\| \leq C e^{-t} \|x - y\|.$$

Thus, condition A.2 is satisfied with  $r(t) = C_2 e^{-t}$ , for some  $C_2 > 0$ .

Take now  $N_0 := 2C_1 \lambda_0$  and choose  $t_0$  such that  $C_2 e^{-t_0} \leq 1/3$ . Then by above the strong solution of SFDE (3.1) satisfies all the conditions of Theorem 2.4(i) for any  $N \geq N_0$  and  $t \geq t_0$ . Thus the distance  $d_N$  is contracting for  $P_t$  for any  $N \geq N_0$  and  $t \geq t_0$ .

(ii). Fix  $t \geq 2r$  and  $M > 0$ . Let us check that Assumption B1 holds for  $B_M$ . Given  $\varepsilon > 0$ , we take  $D := \{x \in \mathcal{C} : \|x\| \leq \varepsilon/2\}$ . Then, clearly, B1.2 holds. By [23], Lemma 3.8, B1.1 is also satisfied. Therefore, the statement of the lemma follows from Theorem 2.4(ii).

(iii). Fix  $t_0 \geq 3r$  and  $M > 0$ . As in the proof of part (ii) of the lemma, let us verify that Assumption B1 holds for  $H_M$ . Given  $\varepsilon > 0$ , let us set again  $D := \{x \in \mathcal{C} : \|x\| \leq \varepsilon/2\}$ . Clearly, B1.2 holds.

To check B1.1, we note first that by the Itô formula we have for any  $t \geq 0$ ,

$$d|X^x(t)|^2 = 2\langle X^x(t), f(X_t^x) \rangle dt + \|g(X_t^x)\|^2 dt + dM(t),$$

where  $M$  is a local martingale with  $M(0) = 0$  and

$$dM(t) = 2\langle X^x(t), g(X_t^x) dW(t) \rangle.$$

For arbitrary  $\delta > 0$ , let  $\tau_\delta := \inf\{t \geq 0 : |X^x(t)| \geq \delta^{-1}\}$ . Applying the Burkholder–Davis–Gundy inequality and Assumption (3.3), we derive for any  $t \in [0, r]$

$$\begin{aligned} \mathbf{E} \sup_{s \in [0, t \wedge \tau_\delta]} |X^x(s)|^4 &\leq C|x(0)|^4 + C\mathbf{E} \int_0^{t \wedge \tau_\delta} (1 + |X^x(s)|^4) ds \\ &\leq C|x(0)|^4 + C\mathbf{E} \int_0^{t \wedge \tau_\delta} \left(1 + \sup_{u \in [0, s \wedge \tau_\delta]} |X^x(u)|^4\right) ds \\ &\leq C|x(0)|^4 + C\mathbf{E} \int_0^t \left(1 + \sup_{u \in [0, s \wedge \tau_\delta]} |X^x(u)|^4\right) ds, \end{aligned}$$

where the constant  $C > 0$  does not depend on  $\delta$ . Clearly, for any  $t \in [0, 1]$  we have  $\mathbf{E} \sup_{s \in [0, t \wedge \tau_\delta]} |X^x(s)|^4 \leq \delta^{-4} < \infty$ . Therefore, the Gronwall inequality yields

$$\mathbf{E} \sup_{s \in [0, r \wedge \tau_\delta]} |X^x(s)|^4 \leq C|x(0)|^4 + C,$$

where the constant  $C > 0$  is again independent of  $\delta$ . By Fatou's lemma, we finally obtain

$$\mathbf{E} \sup_{s \in [0, r]} |X^x(s)|^4 \leq C|x(0)|^4 + C.$$

Thus, for some large  $L = L(M) > 0$ , by the Chebyshev inequality we have

$$\inf_{x \in H_M} \mathbf{P}(\|X_r^x\| \leq L) > 1/2.$$

This combined with [23], Lemma 3.8, implies

$$\inf_{x \in H_M} \mathbb{P}(\|X_{t_0}^x\| \leq \varepsilon/2) \geq \frac{1}{2} \inf_{x: \|x\| \leq L} \mathbb{P}(\|X_{t_0-r}^x\| \leq \varepsilon/2) > 0.$$

Therefore, condition **B1.1** holds and the statement of the lemma follows from Theorem 2.4(ii).  $\square$

Now we can present the proof of Theorem 3.1.

**PROOF OF THEOREM 3.1.** Let us check that all the conditions of Propositions 2.1 are satisfied. Recall the definition of  $N_0$  from Lemma 3.2 and set  $N := N_0 \vee 1$ . We take  $E := \mathcal{C}$ ,  $\rho(x, y) = \|x - y\| \wedge 1$ ,  $d(x, y) = (N\|x - y\|) \wedge 1$ ,  $x, y \in E$ .

The first condition of Propositions 2.1 is satisfied by the assumptions of the theorem. The second condition obviously holds. By Lemma 3.2(i), there exists  $t_0 > 0$  such that the distance  $d$  is contracting for  $P_t$  for all  $t \geq t_0$ . If condition (i) (resp., condition (ii)) of the theorem holds, then by Lemma 3.2(ii) (resp., Lemma 3.2(iii)) for any  $M > 0$  the level set  $\{V \leq M\}$  is  $d$ -small for  $P_{3r}$ . This implies that the third and the fourth condition of Propositions 2.1 are satisfied.

Thus, all the conditions of Propositions 2.1 are satisfied. The proof of the theorem is completed by an application of this proposition.  $\square$

**4. Exponential ergodicity for SPDEs.** In this section, we develop a general framework for establishing exponential ergodicity in the SPDE setting. For the convenience of the reader, first we outline the argument and indicate the main difficulty. Consider, in the spirit of [10], an SDE in a Hilbert space  $H$  of the form

$$(4.1) \quad dX(t) = AX(t)dt + B(X(t))dt + \Sigma(X(t))dW(t), \quad t \geq 0,$$

where  $W$  is a cylindrical Wiener process taking values in a Hilbert space  $G$ ;  $A$  is a nonpositive self-adjoint linear operator  $H \rightarrow H$  with compact inverse;  $B, \Sigma$  are measurable mappings  $H \rightarrow H$  and  $H \rightarrow L_2(G, H)$ , respectively. Here,  $L_2(G, H)$  denotes the space of all Hilbert–Schmidt operators  $G \rightarrow H$ . It follows that  $A$  has negative eigenvalues  $-\infty < \dots \leq -\lambda_2 \leq -\lambda_1 < 0$  and that the corresponding eigenvectors form a complete orthonormal basis of  $H$ . We refer to [10], Chapters 4.2 and 7, for the precise definitions. Assume that  $A, B$  and  $\Sigma$  are such that equation (4.1) has a unique strong solution.

We see that the principal linear part of the drift coefficient satisfies

$$(Ax, x)_H \leq -\lambda_1 \|x\|_H^2, \quad x \in H.$$

The simplest case, for example, [10], Chapter 11.6, is the one where the nonlinear part  $B$  of the drift coefficient as well as the diffusion coefficient  $\Sigma$  are Lipschitz continuous and their Lipschitz constants are sufficiently small compared with  $\lambda_1$ . For such a *dissipative* system, one easily gets the following  $L_2$ -contraction property: for two solutions  $X, Y$  of (4.1) with the same noise and initial conditions  $X(0) = x, Y(0) = y$ ,

$$(4.2) \quad \mathbb{E} \|X(t) - Y(t)\|_H^2 \leq e^{-ct} \|x - y\|_H^2, \quad x, y \in H,$$

where  $c > 0$  is some constant that depends only on  $\lambda_1$  and Lipschitz constants of  $B$  and  $\Sigma$ . Clearly, (4.2) yields exponential ergodicity of the model.

If the Lipschitz constants of  $B$  and  $\Sigma$  are large, then the entire system is not dissipative and the  $L_2$ -contraction property (4.2) for the *true* coupling  $(X, Y)$  has no chance to be satisfied. Nevertheless, one can provide an analogue of (4.2) for a certain *generalized* coupling using a stochastic control-type argument, similar to the one which we have used for SFDEs before.

Namely, assume for a while that  $\Sigma$  is uniformly nondegenerate; that is, the norm of the linear operator  $\Sigma(x)^{-1}$  is uniformly bounded over all  $x \in H$ . Let  $X$  be the same as above, and  $Y$  be defined by

$$(4.3) \quad \begin{aligned} dY(t) &= AY(t) dt + B(Y(t)) dt \\ &+ \Sigma(Y(t)) dW(t) - \lambda(Y(t) - X(t)) dt, \quad Y(0) = y. \end{aligned}$$

If  $\lambda > 0$  is large enough compared with the Lipschitz constants for  $B$ ,  $\Sigma$ , then the pair  $(X, Y)$  satisfies (4.2). On the other hand, the SDE for  $Y$  can be written as

$$dY(t) = AY(t) dt + B(Y(t)) dt + \Sigma(Y(t)) d\widetilde{W}(t),$$

where

$$d\widetilde{W}(t) = dW(t) - \lambda \Sigma(Y(t))^{-1} (Y(t) - X(t)) dt.$$

Then the law of  $\widetilde{W}$  is absolutely continuous with respect to the law of  $W$ , and moreover, it is possible to bound the total variation distance between these laws, and thus to verify Assumptions **A** and **B2** for this generalized coupling. The argument here is essentially the same as in Section 3 above. It is based on Theorem A.2 and Lemma A.1 with a minor difference that now we have to use the analogues of these results for  $H$ -valued processes. This difference is inessential; see Remark A.4.

If  $\Sigma$  fails to be nondegenerate, this argument still applies, but with a certain modification. Namely, let  $H_N$  be the span of the first  $N$  eigenvectors of  $A$  (which correspond to eigenvalues  $-\lambda_1, -\lambda_2, \dots, -\lambda_N$ ) and  $P_N$  be the projector on this span. Assume that for all  $x \in H$  we have  $\text{Range } \Sigma(x) \supset H_N$ , and the corresponding pseudo-inverse operator  $\Sigma(x)^{-1} : H_N \rightarrow G$  is uniformly bounded over  $x \in H$ . Instead of (4.3), consider the following equation:

$$(4.4) \quad \begin{aligned} dY(t) &= AY(t) dt + B(Y(t)) dt \\ &+ \Sigma(Y(t)) dW(t) - \frac{\lambda_N}{2} P_N(Y(t) - X(t)) dt, \quad Y(0) = y. \end{aligned}$$

Then the previous argument remains applicable under the assumption that the Lipschitz constants of  $B$  and  $\Sigma$  are small compared with  $\lambda_N$ . This is essentially the argument developed in [19], which combines the dissipativity property of  $A$  for high modes of the model with the ‘‘stabilization by noise’’ effect for lower modes, and works well in SPDE models with Lipschitz nonlinearities, such as, for example, the stochastic reaction–diffusion equation; see [19], Section 6.1.

This approach is still applicable in SPDE models which contain strongly singular terms, such as the nonlinear gradient term  $(\mathbf{u} \cdot \nabla)\mathbf{u}$  in the Navier–Stokes equation. The technique here dates back (in the deterministic setting) to the exceptional paper [14] and now is used in the theory of finite dimensional attractors [8, 39] and meteorology [3, 11, 24]. It was shown in [30] that for the stochastic Navier–Stokes equation a more sophisticated version of (4.2) is available (see [30], formula (18), and (4.13) below), which leads to exponential ergodicity under a certain balance between energy dissipation and energy influx. In [18], this principal calculation was combined with a stochastic control argument; this led to the construction of an asymptotic coupling under a milder balance condition which involves only the higher modes of the system. This construction appears to be not model specific; in [18], using this construction, five SPDE models were shown to be uniquely ergodic.

Here, we continue (and in a sense finalize) this argument, and show that essentially the same construction can be used to verify Assumptions **A** and **B2** (and thus to prove exponential ergodicity) for SPDEs with non-Lipschitz nonlinearities.

Note that, comparing the proof of Theorem 3.1 and the results of the current section, we can clearly see the difference between the Assumptions **B1** and **B2**. Assumption **B1** for

SFDEs was verified using the support theorem. For SPDEs the support theorem is not easily available, which makes it difficult to use the argument based on Assumption **B1**. Fortunately, we can use instead Assumption **B2**, which can be verified using the generalized coupling method.

The structure of the rest of this section is as follows. To make the argument visible, we first perform basic calculation for the two-dimensional stochastic Navier–Stokes equation, which provides a dissipativity-type bound for this model. Then we present a general toolkit which makes it possible to combine this bound with an energy-type estimate and verify Assumptions **A** and **B2**. Finally, we apply this general toolkit to three other SPDE models from [18]: the hydrostatic Navier–Stokes model, the fractionally dissipative Euler model and the damped Euler–Voigt model. Additionally, we also treat the Boussinesq equations.

Note that the remaining SPDE from [18], the damped nonlinear wave equation, does not have any non-Lipschitz nonlinearities; therefore, the exponential ergodicity of this SPDE can be shown directly by an argument similar to the one used in [19]. Thus we do not treat it here.

*4.1. 2D Navier–Stokes equation, I: Basic calculations.* In this and the next sections, we will frequently use the following notation: for a function  $f: \mathbb{R}_+ \rightarrow \mathbb{R}$  we denote the part of the trajectory

$$(4.5) \quad f_{[0,t]} := \{f(s), s \in [0, t]\}, \quad t \geq 0.$$

Recall the standard notation. Denote by  $V$  the subspace of  $H^1(D)^2$ , which contains  $\mathbf{u}$  such that  $\nabla \cdot \mathbf{u} = 0$  and  $\mathbf{u}|_{\partial D} = 0$ . Denote by  $H$  the completion of  $V$  w.r.t. the  $L_2(D)^2$ -norm, by  $P_H$  the projector in  $L_2(D)^2$  on  $H$ , and by  $A := -P_H \Delta$  the Stokes operator. The eigenvectors  $e_1, e_2, \dots$  of the Stokes operator (with the corresponding eigenvalues  $0 < \lambda_1 \leq \lambda_2 \leq \dots$ ) form a complete orthonormal basis of  $H$ . We denote by  $\|\cdot\|_H$  the standard  $L_2(D)^2$ -norm and for  $\mathbf{w} = (w_1, w_2) \in V$  put  $\|\mathbf{w}\|_V^2 := \|\nabla w_1\|_H^2 + \|\nabla w_2\|_H^2$ .

Consider the 2D stochastic Navier–Stokes equation evolving on a bounded domain  $D \subset \mathbb{R}^2$  with a smooth boundary  $\partial D$ :

$$(4.6) \quad \begin{aligned} d\mathbf{u}(z, t) + (\mathbf{u}(z, t) \cdot \nabla)\mathbf{u}(z, t) dt &= (v \Delta \mathbf{u}(z, t) - \nabla p(z, t) + \mathbf{f}(z)) dt \\ &+ \sum_{k=1}^m \sigma_k(z) dW^k(t), \quad z \in D, t \geq 0; \\ \mathbf{u}(\cdot, 0) &= \mathbf{x}, \quad \nabla \cdot \mathbf{u} = 0, \quad \mathbf{u}|_{\partial D} = 0, \end{aligned}$$

where  $\mathbf{u} = (u_1, u_2)$  is the unknown velocity field,  $p$  is the unknown pressure,  $m \in \mathbb{N}$ ,  $W = (W^1, W^2, \dots, W^m)$  is a standard  $m$ -dimensional Brownian motion,  $\mathbf{f}, \sigma_1, \dots, \sigma_m \in L_2(D)^2$ ,  $v > 0$ . As usual (see, e.g., [18], Remark 3.1), without loss of generality, we can assume that  $\mathbf{f}$  and all  $\sigma_i$  are divergence free, and thus, are in  $H$ .

It is known ([18], Section 3.1.1) that for any initial condition  $\mathbf{x} \in H$  this equation has a unique strong solution, which in the case of ambiguity will be denoted later by  $\mathbf{u}^{\mathbf{x}}$ . Further,  $\mathbf{u}$  is a Feller Markov process with state space  $H$ .

The generalized coupling construction which we will use is the same as in [27], Section 6.2.1, and is a slight variation of the construction from [18], Section 3.1.2. Namely, we consider  $\sigma$  as a linear operator  $\mathbb{R}^m \rightarrow L_2(D)^2$  and fix the maximal possible  $N$  such that

$$(4.7) \quad H_N := P_N H \subset \text{Range}(\sigma) = \text{Span}(\sigma_k, k = 1, \dots, m);$$

here,  $P_N$  stands for the projection to the span of the first  $N$  eigenvectors  $e_1, e_2, \dots, e_N$  of the Stokes operator.

For given  $\mathbf{x}, \mathbf{y} \in H$ , take  $\mathbf{X}^{\mathbf{x}, \mathbf{y}} := \mathbf{u}^{\mathbf{x}}$  and define  $\mathbf{Y}^{\mathbf{x}, \mathbf{y}}$  as the solution to the following stochastic 2D Navier–Stokes equation (4.6) with the initial condition  $\mathbf{y}$  and the additional control term:

$$\begin{aligned}
(4.8) \quad & d\mathbf{Y}^{\mathbf{x}, \mathbf{y}}(z, t) + (\mathbf{Y}^{\mathbf{x}, \mathbf{y}}(z, t) \cdot \nabla) \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(z, t) dt \\
& = (\nu \Delta \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(z, t) - \nabla \tilde{p}(z, t) + \mathbf{f}(z)) dt \\
& \quad + \frac{\nu \lambda_{N+1}}{2} P_N (\mathbf{X}^{\mathbf{x}, \mathbf{y}}(\cdot, t) - \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(\cdot, t)) dt \\
& \quad + \sum_{k=1}^m \sigma_k(z) dW^k(t), \\
& \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(z, 0) = \mathbf{y}(z), \quad \nabla \cdot \mathbf{Y}^{\mathbf{x}, \mathbf{y}} = 0, \quad \mathbf{Y}^{\mathbf{x}, \mathbf{y}}|_{\partial D} = 0.
\end{aligned}$$

Note that (4.8) is just (4.4) up to a proper change of notation. There are two different ways to understand this equation as a modification of (4.6), both of them being useful for particular purposes. First, one can interpret (4.8) as an analogue of (4.6) with the operator  $\Delta$  changed to  $\hat{\Delta} := \Delta - (\nu \lambda_{N+1}/2) P_N$ , and with an additional forcing term  $(\nu \lambda_{N+1}/2) P_N \mathbf{u}^{\mathbf{x}}$ . This allows one to apply the Girsanov theorem to show that the strong solution to (4.8) is uniquely defined; see [27], Remark 8, for a detailed exposition. Second, we can write (4.8) in the form (4.6) with  $\mathbf{x}$  changed to  $\mathbf{y}$ , and with  $dW(t)$  replaced by

$$\begin{aligned}
(4.9) \quad & dW^{\mathbf{x}, \mathbf{y}}(t) := dW(t) + \beta^{\mathbf{x}, \mathbf{y}}(t) dt, \\
& \beta^{\mathbf{x}, \mathbf{y}}(t) := \frac{\nu \lambda_{N+1}}{2} \sigma^{-1} P_N (\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t)).
\end{aligned}$$

By (4.7), the pseudo-inverse operator  $\sigma^{-1}: H_N \rightarrow \mathbb{R}^m$  is well-defined and bounded; thus there exists a constant  $C > 0$  such that for all  $t \geq 0$ ,

$$(4.10) \quad \|\beta^{\mathbf{x}, \mathbf{y}}(t)\|_{\mathbb{R}^m} \leq C \|P_N (\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t))\|_H \leq C \|\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t)\|_H.$$

To check the first condition of Assumption **A**, we note that by construction we have  $\text{Law}(\mathbf{X}^{\mathbf{x}, \mathbf{y}}) = P_x$ . Further, recall that for any  $t \geq 0$  the strong solution to equation (4.6) with the initial value  $\mathbf{y}$ ,  $\mathbf{u}^{\mathbf{y}}(t)$ , is an image of the driving noise under a measurable mapping

$$\Phi_t^{\mathbf{y}}: \mathcal{C}([0, t], \mathbb{R}^m) \rightarrow H.$$

In other words,  $\mathbf{u}^{\mathbf{y}}(t) = \Phi_t^{\mathbf{y}}(W_{[0, t]})$ , recall the convention (4.5). It follows from the Girsanov theorem ([29], Theorem 7.4) that  $\text{Law}(W_{[0, t]}^{\mathbf{x}, \mathbf{y}})$  is absolutely continuous with respect to  $\text{Law}(W_{[0, t]})$ . Therefore, by the uniqueness of the solution, we have  $\mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t) = \Phi_t^{\mathbf{y}}(W_{[0, t]}^{\mathbf{x}, \mathbf{y}})$ .

By the mere definition of the total variation distance,

$$\begin{aligned}
d_{\text{TV}}(P_t(\mathbf{y}, \cdot), \text{Law}(\mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t))) & = d_{\text{TV}}(\text{Law}(\mathbf{u}^{\mathbf{y}}(t)), \text{Law}(\mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t))) \\
& = d_{\text{TV}}(\text{Law}(\Phi_t^{\mathbf{y}}(W_{[0, t]})), \text{Law}(\Phi_t^{\mathbf{y}}(W_{[0, t]}^{\mathbf{x}, \mathbf{y}}))) \\
& \leq d_{\text{TV}}(\text{Law}(W_{[0, t]}), \text{Law}(W_{[0, t]}^{\mathbf{x}, \mathbf{y}})).
\end{aligned}$$

Thus, for any  $\delta \in (0, 1]$  we derive that there exists  $C_\delta > 0$  such that for any  $t \geq 0$  we have by (4.10) and Theorem A.5,

$$\begin{aligned}
(4.11) \quad & d_{\text{TV}}(P_t(\mathbf{y}, \cdot), \text{Law}(\mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t))) \\
& \leq C_\delta \left( \mathbb{E} \left( \int_0^t \|\mathbf{X}^{\mathbf{x}, \mathbf{y}}(s) - \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(s)\|_H^2 ds \right)^\delta \right)^{1/(1+\delta)}.
\end{aligned}$$

As we have already explained, the crucial difficulty in the estimation of the  $\|\cdot\|_H$ -difference between  $\mathbf{X}^{x,y}$  and  $\mathbf{Y}^{x,y}$  appears because of the non-Lipschitz structure of the term  $(\mathbf{u} \cdot \nabla)\mathbf{u}$  in the original equation. To overcome this difficulty, we use the idea from [14]; see also [18], Section 3.1.2. The Ladyzhenskaia trick [28], formula (6), yields the following generic bound:

$$(4.12) \quad \left| \int_D (\mathbf{v} \cdot \nabla)\mathbf{u} \cdot \mathbf{v} dz \right| \leq 2\|\mathbf{v}\|_H \|\mathbf{v}\|_V \|\mathbf{u}\|_V, \quad \mathbf{u}, \mathbf{v} \in V.$$

Combining the Itô formula, the Poincaré inequality, bound (4.12) and the Gronwall inequality, one gets for any  $t \geq 0$ ,

$$(4.13) \quad \|\mathbf{X}^{x,y}(t) - \mathbf{Y}^{x,y}(t)\|_H^2 \leq \|\mathbf{x} - \mathbf{y}\|_H^2 \exp\left(-\nu\lambda_{N+1}t + \frac{4}{\nu} \int_0^t \|\mathbf{X}^{x,y}(s)\|_V^2 ds\right);$$

see [27], formula (6.10), and [18], formula (3.9). This inequality can be understood as a proper substitute for the dissipativity bound (4.2). However, due to the extra term, which involves the stronger  $\|\cdot\|_V$ -norm of the solution  $\mathbf{X}^{x,y}$ , this bound cannot be directly used to verify Assumption A.2. Fortunately, the extra term can be efficiently bounded using the energy estimates. We have (see [18], p. 627, line 16)

$$d\|\mathbf{u}\|_H^2 + 2\nu\|\mathbf{u}\|_V^2 dt = 2(\mathbf{f}, \mathbf{u})_H dt + \|\sigma\|_H^2 dt + 2(\sigma, \mathbf{u})_H dW.$$

Using the Cauchy inequality,

$$2(\mathbf{f}, \mathbf{u})_H \leq \nu\|\mathbf{u}\|_V^2 + \frac{1}{\nu}\|A^{-1/2}\mathbf{f}\|_H^2,$$

we obtain the integral estimate

$$(4.14) \quad \begin{aligned} & \|\mathbf{u}(t)\|_H^2 + \nu \int_0^t \|\mathbf{u}(s)\|_V^2 ds \\ & \leq \|\mathbf{u}(0)\|_H^2 + \left(\frac{1}{\nu}\|A^{-1/2}\mathbf{f}\|_H^2 + \|\sigma\|_H^2\right)t + M(t), \end{aligned}$$

where  $M$  is a continuous local martingale with

$$(4.15) \quad d\langle M \rangle_t = 4(\sigma, \mathbf{u})_H^2 dt \leq 4\|\sigma\|_H^2 \|\mathbf{u}\|_H^2 dt \leq C\|\sigma\|_H^2 \|\mathbf{u}\|_V^2 dt,$$

where  $C > 0$  and the last inequality follows from the Poincaré inequality.

With the estimates (4.11) and (4.13)–(4.15) in hand, we are able to construct a premetric  $\theta$  (which is going to be nonsymmetric due to the presence of the extra factor in (4.13)) such that Assumptions A and B2 are verified. This allows to prove exponential ergodicity of  $\mathbf{u}^x$ . Such a construction is quite generic and can be used in various SPDE models. Thus, for the convenience of further applications, we introduce it separately and in an abstract form.

**4.2. A general toolkit for SPDEs.** In this subsection, motivated by the above analysis, we introduce a general framework for establishing exponential ergodicity of the solutions of SPDEs. We use the general setting from Section 2.1; that is,  $(E, \rho)$  is a Polish space,  $\{P_t(x, A), x \in E, A \in \mathcal{E}\}_{t \in \mathbb{R}_+}$  is a Markov transition function and  $\{P_x, x \in E\}$  is the corresponding Markov family. Recall the notion of a premetric that was defined in the beginning of Section 2.2. Assume the following.

**ASSUMPTION H.** There exist a lower semicontinuous function  $U: E \rightarrow [0, \infty)$ , a measurable function  $S: E \rightarrow [0, \infty]$  and a premetric  $q$  on  $E$  such that for any given  $x, y \in E$  there exists a couple of progressively measurable random processes  $X^{x,y} = (X_t^{x,y})_{t \geq 0}$ ,  $Y^{x,y} = (Y_t^{x,y})_{t \geq 0}$ , that satisfies the following set of conditions:

**H1** (dissipativity bound):

$$(4.16) \quad q(X_t^{x,y}, Y_t^{x,y}) \leq q(x, y) \exp\left(-\zeta t + \kappa \int_0^t S(X_s^{x,y}) ds\right), \quad t \geq 0,$$

where  $\zeta > 0, \kappa \geq 0$ ;

**H2** (energy estimate):

$$(4.17) \quad U(X_t^{x,y}) + \mu \int_0^t S(X_s^{x,y}) ds \leq U(X_0^{x,y}) + bt + M_t, \quad t \geq 0,$$

where  $\mu > 0, b \geq 0$  are some constants such that

$$(4.18) \quad \zeta > \frac{\kappa b}{\mu};$$

and  $M$  is a continuous local martingale with  $M_0 = 0$  and

$$(4.19) \quad d\langle M \rangle_t \leq b_1 S(X_t^{x,y}) dt + b_2 dt, \quad t \geq 0,$$

where  $b_1 \geq 0, b_2 \geq 0$ .

**H3** (error-in-law bounds):  $\text{Law}(X^{x,y}) = P_x$  and for every  $\delta \in (0, 1]$  there exists a constant  $C_\delta > 0$  such that for any  $t > 0$ ,

$$(4.20) \quad d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) \leq C_\delta \left( \mathbb{E} \left( \int_0^t q(X_s^{x,y}, Y_s^{x,y}) ds \right)^\delta \right)^{1/2}.$$

Additionally, for any  $\delta \in (0, 1], R > 0$  there exists  $\varepsilon = \varepsilon(R, \delta) > 0$  such that  $\mathbb{E}(\int_0^t q(X_s^{x,y}, Y_s^{x,y}) ds)^\delta < R$  implies

$$(4.21) \quad d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) \leq 1 - \varepsilon.$$

The functions  $U, S, q$  as well as the constants  $\zeta, \kappa, \mu, b, b_1, b_2, C_\delta, \varepsilon$  should not depend on the pair  $x, y$ .

Typically, the process  $Y^{x,y}$  is constructed as the solution to the original SPDE where the Brownian motion is replaced by the Brownian motion with a suitable drift that pushes  $Y^{x,y}$  toward  $X^{x,y}$  (recall (4.9) and the corresponding construction for the stochastic Navier–Stokes equation). In this case, condition **H3** can be replaced by the following simple condition on this drift. Recall convention (4.5).

**LEMMA 4.1.** *Let  $W$  be an  $m$ -dimensional Brownian motion,  $m \geq 1$ . Assume that there exists a constant  $c > 0$  such that for each  $t \geq 0, x, y \in E$  there exists a measurable function  $\Phi = \Phi^{t,x,y}: C[0, t] \rightarrow E$  and progressively measurable processes  $\beta^{x,y}, \xi^{x,y}: \Omega \times [0, t] \rightarrow \mathbb{R}^m$  such that:*

1. We have  $d\xi_s^{x,y} = dW_s + \beta_s^{x,y} ds, s \in [0, t]$ .
2.  $\text{Law}(\Phi(W_{[0,t]})) = P_t(y, \cdot)$  and  $\Phi(\xi_{[0,t]}) = Y_t^{x,y}$ .
3. For each  $s \in [0, t]$  we have  $|\beta_s|^2 \leq cq(X_s^{x,y}, Y_s^{x,y})$ .

*Then conditions (4.20) and (4.21) hold.*

**PROOF.** Fix  $\delta \in (0, 1], t > 0$ . Denote

$$M_\delta := \mathbb{E} \left( \int_0^t |\beta_s|^2 ds \right)^\delta.$$

It follows from condition 3 of the lemma that

$$(4.22) \quad M_\delta \leq c^\delta \mathbb{E} \left( \int_0^t q(X_s^{x,y}, Y_s^{x,y}) ds \right)^\delta.$$

Further, inequality (A.13) from Theorem (A.5) and condition 2 of the lemma imply

$$\begin{aligned} d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) &= d_{\text{TV}}(\text{Law}(\Phi(\xi_{[0,t]})), \text{Law}(\Phi(W_{[0,t]}))) \\ &\leq d_{\text{TV}}(\text{Law}(\xi_{[0,t]}), \text{Law}(W_{[0,t]})) \\ &\leq 2^{(1-\delta)/(1+\delta)} (M_\delta)^{1/(1+\delta)}. \end{aligned}$$

Combining this with (4.22), we obtain (4.20).

Similarly, inequality (A.14) from Theorem (A.5), condition 2 of the lemma and (4.22) yield (4.21).  $\square$

Now let us present the main result of this section. It shows that Assumptions **H1–H3** together with the existence of a suitable Lyapunov function imply exponential ergodicity.

**THEOREM 4.2.** *Suppose that the Markov kernel  $(P_t)$  is Feller and satisfies Assumptions **H1–H3** for some functions  $U, S, q$ . Assume further that there exists a measurable function  $V: E \rightarrow \mathbb{R}_+$  such that for some  $\gamma > 0, K > 0$  we have*

$$(4.23) \quad \mathbb{E}_x V(X_t) \leq V(x) - \gamma \mathbb{E}_x \int_0^t V(X_s) ds + Kt, \quad t \geq 0, x \in E$$

and for any  $M > 0$  the functions  $U(\cdot), q(\cdot, \cdot)$  are bounded on the level sets  $\{V \leq M\}$  and  $\{V \leq M\} \times \{V \leq M\}$ , respectively.

If additionally  $\rho \leq q^\delta$  for some  $\delta > 0$ , then  $(P_t)$  has a unique invariant measure  $\pi$ . Moreover, there exist constants  $C, r > 0$  such that

$$(4.24) \quad W_{\rho \wedge 1}(P_t(x, \cdot), \pi) \leq C(1 + V(x))e^{-rt}, \quad t \geq 0, x \in E.$$

**REMARK 4.3.** If we assume additionally that a Markov semigroup  $P$  has an invariant measure  $\pi$ , then condition  $\rho \leq q^\delta$  of Theorem 4.2 can be dropped. In this case, the left-hand side of bound (4.24) is replaced by  $W_{\tilde{q}}(P_t(x, \cdot), \pi)$ , where  $\tilde{q}(x, y) := q(x, y)^\delta \wedge q(y, x)^\delta \wedge 1$ , where  $\delta$  is arbitrary positive real number.

The proof of Theorem 4.2 is based on the following key lemma. It shows that Assumptions **A** and **B2** follow from Assumptions **H1–H3**.

**LEMMA 4.4.** *Suppose that Assumptions **H1–H3** hold. Then there exists  $\alpha_0 > 0$  that depends only on  $\zeta, \kappa, \mu, b, b_1, b_2, C_\delta, \varepsilon$ , such that for any  $\alpha \in (0, \alpha_0]$  there exist constants  $C, Q, \lambda > 0$  such that:*

(i) Assumption **A** holds for the premetric

$$\theta(x, y) = e^{QU(x)} q(x, y)^\alpha,$$

and the rate functions  $L(t) := C, r(t) = C \exp(-\lambda t)$ .

(ii) For any set  $B$  such that  $U(\cdot)$  is bounded on  $B$ , and  $q(\cdot, \cdot)$  is bounded on  $B \times B$  there exists  $C > 0$  such that Assumption **B2** holds for the same premetric  $\theta$ , the set  $B$  and the rate function  $R(t) := C \exp(-\lambda t)$ .

**PROOF OF LEMMA 4.4.** (i). Take any  $x, y \in E$ . First of all, let us derive a good bound on  $q(X_t^{x,y}, Y_t^{x,y}), t \geq 0$ . Let  $\gamma > 0$  be a sufficiently small parameter to be chosen later. Define

$$\mathbb{E}_\gamma := \sup_{t \geq 0} (M_t - \gamma \langle M \rangle_t).$$

By the Dambis–Dubins–Schwarz theorem (see, e.g., [36], Theorem 5.1.6), there exists (maybe on an extended probability space) a Brownian motion  $\widehat{W}$  such that  $M_t = \widehat{W}_{(M)_t}$ ,  $t \geq 0$ . Therefore,

$$\Xi_\gamma \leq \sup_{t \geq 0} (\widehat{W}_t - \gamma t)$$

and

$$(4.25) \quad \mathbb{P}(\Xi_\gamma \geq R) \leq \mathbb{P}\left(\sup_{t \geq 0} (\widehat{W}_t - \gamma t) \geq R\right) = e^{-2\gamma R}, \quad R \geq 0;$$

see [5], Part II, formula 2.0.2.(1). By (4.17) and (4.19), we have

$$(4.26) \quad U(X_t^{x,y}) + (\mu - \gamma b_1) \int_0^t S(X_s^{x,y}) ds \leq U(x) + (b + \gamma b_2)t + \Xi_\gamma.$$

From now on, we assume that  $0 < \gamma < \mu/b_1$ . Then, combining (4.26) with (4.16), we obtain

$$(4.27) \quad q(X_t^{x,y}, Y_t^{x,y}) \leq q(x, y) \exp(-\chi t + \nu(U(x) - U(X_t^{x,y}) + \Xi_\gamma)),$$

where we denoted

$$\nu = \frac{\kappa}{\mu - \gamma b_1}, \quad \chi := \zeta - \frac{\kappa(b + \gamma b_2)}{\mu - \gamma b_1}.$$

Recall that thanks to (4.18) we have  $\zeta - \kappa b/\mu > 0$ . Therefore, we can take  $\gamma$  small enough so that  $\chi > 0$ .

We begin with verifying **A.1**. By **H3**, we have  $\text{Law}(X^{x,y}) = \mathbb{P}_x$ . To estimate the distance between  $\text{Law}(Y_t^{x,y})$  and  $P_t(y, \cdot)$  we also use **H3** together with (4.27). We get for any  $\delta \in (0, 1)$ ,

$$(4.28) \quad \begin{aligned} d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) &\leq C_\delta \left( \mathbb{E} \left( \int_0^t q(X_s^{x,y}, Y_s^{x,y}) ds \right)^\delta \right)^{1/2} \\ &\leq C q(x, y)^{\delta/2} e^{\nu \delta U(x)/2} \mathbb{E} \exp\left(\frac{\nu \delta}{2} \Xi_\gamma\right). \end{aligned}$$

Define now for  $\alpha \in (0, 1)$

$$(4.29) \quad \theta_\alpha(z_1, z_2) := q(z_1, z_2)^\alpha e^{\alpha \nu U(z_1)}, \quad z_1, z_2 \in E.$$

Recall that (4.25) implies that  $\mathbb{E} \exp(K \Xi_\gamma) < \infty$  for  $K < 2\gamma$ . Thus, by taking

$$(4.30) \quad \alpha_0 := (\gamma/\nu) \wedge (1/2),$$

we see that for any  $\alpha \in (0, \alpha_0]$  inequality (4.28) implies

$$d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) \leq C \theta_\alpha(x, y)$$

and, therefore, Assumption **A.1** holds for the premetric  $\theta_\alpha$ ,  $\alpha \in (0, \alpha_0]$ .

To check **A.2**, we employ again (4.27). Using the definition of the premetric  $\theta_\alpha$  in (4.29), we derive

$$(4.31) \quad \begin{aligned} \mathbb{E} \theta_\alpha(X_t^{x,y}, Y_t^{x,y}) &= \mathbb{E} q(X_t^{x,y}, Y_t^{x,y})^\alpha e^{\alpha \nu U(X_t^{x,y})} \\ &\leq \mathbb{E} e^{-\lambda t} q(x, y)^\alpha e^{\alpha \nu U(x)} e^{\alpha \nu \Xi_\gamma} \\ &\leq C e^{-\lambda t} \theta_\alpha(x, y), \end{aligned}$$

where we denoted  $\lambda = \alpha \chi$ . Thus Assumption **A.2** also holds for the premetric  $\theta_\alpha$ ,  $\alpha \in [0, \alpha_0]$ .

(ii). We will use the same notation as in the part (i) of the proof. Fix  $\alpha \in (0, \alpha_0]$  and take any  $x, y \in B$ . First, let us verify **B2.1**. Clearly,  $\text{Law}(X^{x,y}) = P_x$  thanks to **H3**. Fix now  $t > 0$ . To estimate the total variation distance between  $\text{Law}(Y_t^{x,y})$  and  $P_t(y, \cdot)$  denote

$$M_\alpha := \mathbb{E} \left( \int_0^t q(X_s^{x,y}, Y_s^{x,y}) ds \right)^\alpha.$$

Inequality (4.27) and the boundedness of  $q$  and  $U$  on the set  $B$  imply

$$M_\alpha \leq C \mathbb{E} \exp(\nu \alpha \mathbb{E}_\gamma) \leq C_1,$$

where we also used (4.30) and the fact that  $\alpha \leq \alpha_0$ . Now condition (4.21) from **H3** implies

$$d_{\text{TV}}(\text{Law}(Y_t^{x,y}), P_t(y, \cdot)) \leq 1 - \varepsilon$$

for some  $\varepsilon > 0$ . This yields (2.5).

To verify **B2.2**, we use (4.31) and the definition of the metric  $\theta_\alpha$ , which is (4.29). Using again the fact that  $q$  and  $U$  are bounded on  $B$ , we get

$$\mathbb{E} \theta_\alpha(X_t^{x,y}, Y_t^{x,y}) \leq C e^{-\lambda t},$$

for some  $C = C(B) > 0$ . Thus, **B2.2** holds for the metric  $\theta_\alpha$ .  $\square$

Now we can present a proof of Theorem 4.2.

**PROOF OF THEOREM 4.2.** The theorem follows from Lemma 4.4 and Theorem 2.7.

Recall the definition of  $\alpha_0$  from Lemma 4.4. Take  $\alpha := \alpha_0 \wedge \delta$ . Then by Lemma 4.4 there exists  $Q > 0$  such that Assumptions **A** and **B2** holds for the premetric  $\theta(x, y) := e^{QU(x)} q(x, y)^\alpha$ .

Let us now check that all the conditions of Theorem 2.7 are satisfied for this premetric  $\theta$ . The first condition of Theorem 2.7 is satisfied due to (4.23). Since, by assumption,  $\rho \leq q^\delta$  and  $U(x) \geq 0$  for any  $x \in E$ , we have

$$\rho(x, y) \wedge 1 \leq q(x, y)^\delta \wedge 1 \leq q(x, y)^\alpha \wedge 1 \leq \theta(x, y), \quad x, y \in E,$$

and thus the second condition of Theorem 2.7 holds.

The third condition of Theorem 2.7 holds by Lemma 4.4(i).

Finally, since  $U$  and  $q$  are bounded on the level sets of  $V$ , we see that Lemma 4.4(ii) implies that the fourth condition of Theorem 2.7 is also satisfied. Now the statement of the theorem follows immediately from Theorem 2.7. The statement of Remark 4.3 follows now from Remark 2.8.  $\square$

**4.3. 2D Navier–Stokes equation, II: Exponential ergodicity.** Now with the general toolkit in hand we can proceed directly and establish exponential ergodicity of the 2D stochastic Navier–Stokes equation. Note that exponential ergodic bounds had been obtained previously for the Navier–Stokes equation only on a 2-dimensional torus; see [31] and [21]. Unlike these results, Theorem 4.5 does not rely on the geometry of the domain  $D$ , or on the fine structure of the forcing. This illustrates well, we believe, that the generalized coupling method is quite insensitive with respect to the local structure of the model.

For ergodicity of 2D Navier–Stokes equations driven by different types of noise we refer also to [25], Section 3.

Recall that  $\lambda_i$  is the  $i$ th largest eigenvalue of the Stokes operator  $A = -P_H \Delta$ .

THEOREM 4.5. *Assume that there exists  $N \in \mathbb{Z}_+$  such that*

$$P_N H \subset \text{Span}(\sigma_k, k = 1, \dots, m)$$

and

$$(4.32) \quad \lambda_{N+1} > 4\nu^{-4} \|A^{-1/2}\mathbf{f}\|_H^2 + 4\nu^{-3} \|\sigma\|_H^2.$$

Then the stochastic Navier–Stokes equation (4.6) has a unique invariant measure  $\pi$ . Further, there exist constants  $C > 0$ ,  $r > 0$  such that

$$W_{\|\cdot\|_H \wedge 1}(\text{Law}(\mathbf{u}^{\mathbf{x}}(t)), \pi) \leq C(1 + \|\mathbf{x}\|_H^2)e^{-rt}, \quad t \geq 0, \mathbf{x} \in H.$$

PROOF. Let us check that all the conditions of Theorem 4.2 are satisfied. Motivated by the preliminary analysis in Section 4.1, let us take

$$U(\mathbf{x}) := \|\mathbf{x}\|_H^2, \quad S(\mathbf{x}) := \|\mathbf{x}\|_V^2, \quad q(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|_H^2, \quad \mathbf{x}, \mathbf{y} \in H.$$

Then it follows from (4.13) that Assumption **H1** is satisfied with  $\zeta := \nu\lambda_{N+1}$  and  $\kappa := 4/\nu$ .

Further, we see from bounds (4.14) and (4.15) that **H2** also holds with  $\mu := \nu$  and  $b := \|A^{-1/2}\mathbf{f}\|_H^2/\nu + \|\sigma\|_H^2$ . Constraint (4.18) follows from (4.32).

To verify **H3**, we employ Lemma 4.1. As discussed in Section 4.1, for any  $t \geq 0$  there exists some measurable mapping  $\Phi$  such that  $\mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t) = \Phi_t^{\mathbf{y}}(W_{[0,t]}^{\mathbf{x}, \mathbf{y}})$  and  $\mathbf{u}^{\mathbf{y}} = \Phi_t^{\mathbf{y}}(W_{[0,t]}^{\mathbf{y}})$ . Inequality (4.10) implies that for some  $C > 0$

$$|\beta_s|^2 \leq Cq(\mathbf{X}^{\mathbf{x}, \mathbf{y}}(s), \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(s)), \quad s \in [0, t].$$

Hence, by Lemma 4.1 condition **H3** also holds.

Finally, we introduce the following Lyapunov function:  $V(\mathbf{x}) := \|\mathbf{x}\|_H^2$ . By definition, for any  $M > 0$  the functions  $U$  and  $q$  are bounded on the level sets  $\{V \leq M\}$  and  $\{V \leq M\} \times \{V \leq M\}$ , respectively. Inequality (4.23) follows from the energy estimate (4.14) and the Poincaré inequality  $\|u\|_V \geq C\|u\|_H$ .

Since  $\rho := \|\mathbf{x} - \mathbf{y}\|_H \leq q(\mathbf{x}, \mathbf{y})^{1/2}$ , we see that all conditions of Theorem 4.2 hold. This immediately implies the statement of the theorem.  $\square$

4.4. *2D hydrostatic Navier–Stokes equation.* In this section, we treat the stochastic hydrostatic Navier–Stokes equations. Fix  $L, h > 0$  and consider a domain  $D \subset \mathbb{R}^2$  defined by  $D := \{(z_1, z_2) : z_1 \in (0, L), z_2 \in (-h, 0)\}$ . Introduce the lateral side of the boundary  $\Gamma_l := \{0, L\} \times [-h, 0]$  and the horizontal side  $\Gamma_h := [0, L] \times \{-h, 0\}$ . Consider the following equations for an unknown velocity field  $(u, w)$  and pressure  $p$  on  $D$ :

$$(4.33) \quad du + (u\partial_{z_1}u + w\partial_{z_2}u)dt = (\nu\Delta u - \partial_{z_1}p)dt + \sum_{k=1}^m \sigma_k dW^k,$$

$$(4.34) \quad \begin{aligned} \partial_{z_2}p &= 0, & \partial_{z_1}u + \partial_{z_2}w &= 0, & u(0) &= x, \\ u|_{\Gamma_l} &= 0, & \partial_{z_2}u|_{\Gamma_h} &= w|_{\Gamma_h} = 0. \end{aligned}$$

where  $W = (W^1, W^2, \dots, W^m)$  is a standard  $m$ -dimensional Brownian motion,  $\nu > 0$ . We assume that for any  $k = 1, \dots, m$  we have  $\sigma_k \in H^2(D)$ ,  $\sigma_k|_{\Gamma_l} = 0$ ,  $\partial_{z_2}\sigma_k|_{\Gamma_h} = 0$ ,  $\int_{-h}^0 \sigma_k dz_2 \equiv 0$ . Denote

$$\begin{aligned} H &:= \left\{ \varphi \in L_2(D) : \int_{-h}^0 \varphi(z_1, z_2) dz_2 = 0 \text{ for all } z_1 \in (0, L) \right\}, \\ V &:= \left\{ \varphi \in H^1(D) : \int_{-h}^0 \varphi(z_1, z_2) dz_2 = 0 \text{ for all } z_1 \in (0, L) \text{ and } \varphi|_{\Gamma_l} = 0 \right\}. \end{aligned}$$

Let  $\|\cdot\|_H$  be the standard  $L_2(D)$ -norm and put  $\|\varphi\|_V^2 := \|\partial_{z_1}\varphi\|_H^2 + \|\partial_{z_2}\varphi\|_H^2$ , where  $\varphi \in V$ . We consider the following complete orthonormal basis of  $H$ :

$$\left\{ e_{i,j} := \frac{2}{\sqrt{hL}} \sin\left(\frac{i\pi z_1}{L}\right) \cos\left(\frac{j\pi z_2}{h}\right) \right\}_{i,j \geq 1},$$

which corresponds to eigenvectors of the negative Laplacian operator with boundary conditions given by (4.34). Their associated eigenvalues are

$$\{\lambda_{i,j} := \pi^2(i^2/L^2 + j^2/h^2)\}_{i,j \geq 1}.$$

It is known ([18], Section 3.2.1, see also [17], Theorems 1.3 and 1.5) that under the above assumptions on  $\sigma$ , for any  $x \in V$  stochastic hydrostatic Navier–Stokes equation (4.33) has a unique strong solution, which in the case of ambiguity will be further denoted by  $u^x$ . Moreover,  $u$  is a Feller Markov process with the state space  $V$  and for Lebesgue-almost all  $t > 0$  the random element  $u^x(t) \in \{\varphi \in H^2(D) : \partial_{z_2}\varphi|_{\Gamma_h} = 0\}$ . The existence and uniqueness of the invariant measure of  $u$  was shown in [18], Section 3.2. The weak convergence of transition probabilities to the invariant measure was established in [27], Section 6.2.2. We strengthen these results and prove exponential ergodicity of  $u$ .

**THEOREM 4.6.** *Assume that*

$$(4.35) \quad \text{Span}(e_{i,j} : \lambda_{i,j} \leq K) \subset \text{Span}(\sigma_k, k = 1, \dots, m),$$

where

$$(4.36) \quad K := 4(1+h)v^{-3} \left(1 + \frac{\sqrt{hL}}{\sqrt{2\pi}}\right) (\|\sigma\|_H^2 + \|\partial_{z_2}\sigma\|_H^2).$$

Then the stochastic hydrostatic Navier–Stokes equation (4.33) has a unique invariant measure  $\pi$ . Further, there exist constants  $C > 0, r > 0$  such that

$$W_{\|\cdot\|_H \wedge 1}(\text{Law}(u^x(t)), \pi) \leq C(1 + \|x\|_H^2 + \|\partial_{z_2}x\|_H^2)e^{-rt}, \quad t \geq 0, x \in V.$$

**PROOF.** As in the proof of Theorem 4.5, we will utilize Theorem 4.2. Let us check that all the conditions of Theorem 4.2 are satisfied. Put

$$\begin{aligned} U(x) &:= \|x\|_H^2 + \|\partial_{z_2}x\|_H^2, & S(x) &= \|x\|_V^2 + \|\partial_{z_2}x\|_V^2, \\ q(x, y) &= \|x - y\|_H^2, & x, y &\in V. \end{aligned}$$

We use the same generalized coupling construction as in [27], Section 6.2.2 (which in turn is a minor variation of the construction from [18], Section 3.2.4).

Fix  $x, y \in V$ . Take  $X^{x,y} := u^x$  and let  $Y^{x,y}$  be the solution to the same equation with initial condition  $Y_0^{x,y} = y$  and with the additional control term in the right-hand side

$$\frac{1}{2}vK P^K(X_t^{x,y} - Y_t^{x,y}) dt;$$

here  $P^K$  is the projection to the space  $\text{Span}(e_{i,j} : \lambda_{i,j} \leq K)$ .

Then an application of the Itô formula, Poincaré inequality and Ladyzhenskaia bound [28], formula (6), yields (see the inequality at [18], line 1, p. 633)

$$(4.37) \quad \begin{aligned} &\frac{1}{2}d\|X^{x,y}(t) - Y^{x,y}(t)\|_H^2 + \frac{1}{2}vK\|X^{x,y}(t) - Y^{x,y}(t)\|_H^2 dt \\ &\leq 4(1 \vee h)v^{-1}\|X^{x,y}(t) - Y^{x,y}(t)\|_H^2 \\ &\quad \times (\|X^{x,y}(t)\|_V^2 + \|\partial_{z_2}X^{x,y}(t)\|_H\|\partial_{z_2}X^{x,y}(t)\|_V). \end{aligned}$$

By the Poincaré inequality, for any  $\varphi \in H^1(D)$  such that  $\varphi|_{\partial D} = 0$  we have

$$\pi^2(h^{-2} + L^{-2})\|\varphi\|_{L^2} \leq \|\nabla\varphi\|_{L^2}.$$

This combined with (4.37) and the Gronwall inequality implies now

$$\begin{aligned} & q(X^{x,y}(t), Y^{x,y}(t)) \\ & \leq q(x, y) \exp\left(-\nu Kt + \kappa \int_0^t (\|X^{x,y}(s)\|_V^2 + \|\partial_{z_2} X^{x,y}(s)\|_V^2) ds\right), \end{aligned}$$

where

$$\kappa := 8(1 \vee h)v^{-1} \left(1 + \frac{\sqrt{hL}}{\sqrt{2\pi}}\right).$$

Thus Assumption **H1** is satisfied.

To check **H2**, we use the standard energy estimates [18], formulae (3.15) and (3.17). We get

$$\begin{aligned} (4.38) \quad & d(\|X^{x,y}(t)\|_H^2 + \|\partial_{z_2} X^{x,y}(t)\|_H^2) + 2\nu(\|X^{x,y}(t)\|_V^2 + \|\partial_{z_2} X^{x,y}(t)\|_V^2) dt \\ & = (\|\sigma\|_H^2 + \|\partial_{z_2}\sigma\|_H^2) dt + M(t), \end{aligned}$$

where  $M$  is a continuous local martingale with the quadratic variation

$$\begin{aligned} d\langle M \rangle_t & \leq (4\|X^{x,y}(t)\|_H^2 \|\sigma\|_H^2 + 4\|\partial_{z_2} X^{x,y}(t)\|_H^2 \|\partial_{z_2}\sigma\|_H^2) dt \\ & \leq C\|X^{x,y}(t)\|_V^2 dt \leq CS(X^{x,y}(t)), \end{aligned}$$

for some  $C > 0$ ; here, in the second inequality, we used the Poincaré inequality. Thus **H2** is satisfied with  $\mu := 2\nu$  and  $b := \|\sigma\|_H^2 + \|\partial_{z_2}\sigma\|_H^2$ . Inequality (4.18) holds due to the definition of  $K$  in (4.36).

Denote

$$W^{x,y}(t) := W(t) + \int_0^t \frac{1}{2} \nu K \sigma^{-1} P^K (X_t^{x,y} - Y_t^{x,y}) dt, \quad t \geq 0.$$

By (4.35), we see that this process is well-defined. Arguing as in the proof of theorem Theorem 4.5, we see that the uniqueness of the solution to (4.33) implies that for each  $t \geq 0$  there exists a measurable mapping  $\Phi_t^y: \mathcal{C}([0, t], \mathbb{R}^m) \rightarrow V$  such that  $Y^{x,y}(t) = \Phi_t^y(W_{[0,t]}^{x,y})$  and  $u^y(t) = \Phi_t^y(W_{[0,t]}^y)$ . Further, thanks again to condition (4.35), we see that (similar to (4.10)) there exists a constant  $C > 0$  such that for any  $t \geq 0$ ,

$$\left\| \frac{1}{2} \nu K \sigma^{-1} P^K (X_t^{x,y} - Y_t^{x,y}) \right\|_{\mathbb{R}^m}^2 \leq C \|X_t^{x,y} - Y_t^{x,y}\|_H^2 = Cq(X_t^{x,y}, Y_t^{x,y}).$$

Thus, by Lemma 4.1, condition **H3** is also satisfied.

Finally, we consider the following Lyapunov function:

$$V(x) = \|x\|_H^2 + \|\partial_{z_2} x\|_H^2, \quad x \in V.$$

Condition (4.23) follows from the energy estimate (4.38). By definition, for any  $M > 0$  the functions  $q$  and  $U$  are bounded on the level sets  $\{V \leq M\} \times \{V \leq M\}$  and  $\{V \leq M\}$ , respectively.

We recall that as shown in [18], Section 3.2, the process  $u$  has an invariant measure. Therefore, all the conditions of Theorem 4.2 and Remark 4.3 hold. This immediately implies the statement of the theorem.  $\square$

4.5. *The fractionally dissipative Euler model.* In this and the next sections, we will denote the fractional Laplacian by  $\Lambda^\gamma := -(-\Delta)^{\gamma/2}$ ,  $\gamma \in (0, 2)$ . For a vector  $\mathbf{f}(z) = (f_1(z), f_2(z))$ , where  $z = (z_1, z_2) \in \mathbb{R}^2$  we denote as usual  $\text{curl } \mathbf{f} := \partial_{z_1} f_2 - \partial_{z_2} f_1$ .

Fix  $r > 2$  and consider the stochastic fractionally dissipative Euler model on the periodic box  $D := [-\pi, \pi]^2$ . In the velocity formulation, the model is given by

$$(4.39) \quad \begin{aligned} d\mathbf{u}(z, t) + (\mathbf{u}(z, t) \cdot \nabla) \mathbf{u}(z, t) dt &= (\Lambda^\gamma \mathbf{u}(z, t) - \nabla p(z, t)) dt \\ &+ \sum_{k=1}^m \boldsymbol{\sigma}_k(z) dW^k(t), \quad z \in D, t \geq 0; \\ \mathbf{u}(\cdot, 0) &= \mathbf{x}, \quad \nabla \cdot \mathbf{u} = 0, \end{aligned}$$

where  $\mathbf{u} = (u_1, u_2)$  is the unknown velocity field,  $p$  is the unknown pressure,  $m \in \mathbb{N}$ ,  $W = (W^1, W^2, \dots, W^m)$  is a standard  $m$ -dimensional Brownian motion. We assume that for any  $k = 1, \dots, m$  we have  $\boldsymbol{\sigma}_k \in H^{r+3}(D)^2$ ,  $\nabla \cdot \boldsymbol{\sigma}_k = 0$ ,  $\int_D \boldsymbol{\sigma}_k(z) dz = 0$ .

Introduce the following space:

$$V := \left\{ \boldsymbol{\varphi} \in H^r(D)^2 : \nabla \cdot \boldsymbol{\varphi} = 0, \int_D \boldsymbol{\varphi}(z) dz = 0 \right\}.$$

The eigenvectors  $e_1, e_2, \dots$  of the operator  $-\Lambda^\gamma|_V$  (with the corresponding eigenvalues  $0 < \lambda_1 = 1 \leq \lambda_2 \leq \dots$ ) form a complete orthonormal basis of  $V$ .

It is known ([18], Section 3.3.1, see also [9], footnote on p. 821, Proposition 1.1 and Theorem 1.4) that for any  $\mathbf{x} \in V$  equation (4.39) has a unique strong solution, which in the case of ambiguity will be further denoted by  $\mathbf{u}^{\mathbf{x}}$ . Moreover,  $\mathbf{u}$  is a Feller Markov process with the state space  $V$ . The existence and uniqueness of the invariant measure of  $\mathbf{u}$  was shown in [9], Theorem 1.5, and [18], Section 3.3. The weak convergence of transition probabilities to the invariant measure was established in [27], Section 6.2.3. It was conjectured in [18], Section 3.3, that to establish the rate of convergence to the invariant measure one might need to exploit high-order polynomial moment bounds of the solution in high-order Sobolev spaces. Actually, applying our method, we are able to show exponential rate of convergence without invoking these complicated bounds; we use just the calculations from [18], Section 3.3.2.

**THEOREM 4.7.** *There exists a universal  $C > 0$  such that if for some  $N \in \mathbb{Z}_+$  one has*

$$\text{Span}(e_k, k = 1, \dots, N) \subset \text{Span}(\boldsymbol{\sigma}_k, k = 1, \dots, m)$$

and

$$(4.40) \quad \lambda_{N+1} > C \|\boldsymbol{\sigma}\|_{L_{6/\gamma}}^2,$$

then the stochastic fractionally dissipative Euler equation (4.39) has a unique invariant measure  $\pi$ . Further, there exist constants  $C_1 > 0$ ,  $r > 0$  such that

$$W_{\|\cdot\|_{L_2} \wedge 1}(\text{Law}(\mathbf{u}^{\mathbf{x}}(t)), \pi) \leq C_1 (1 + \|\text{curl } \mathbf{x}\|_{L_{6/\gamma}}^2 + \|\mathbf{x}\|_{L_2}^2) e^{-rt}, \quad t \geq 0, \mathbf{x} \in V.$$

**PROOF.** Let us again check that all the conditions of Theorem 4.2 are satisfied. For brevity, we denote  $p := 6/\gamma$  and put

$$\begin{aligned} U(\mathbf{x}) &:= (1 + \|\text{curl } \mathbf{x}\|_{L_p}^p)^{2/p}, \quad S(\mathbf{x}) := \|\text{curl } \mathbf{x}\|_{L_p}^2, \\ q(\mathbf{x}, \mathbf{y}) &:= \|\mathbf{x} - \mathbf{y}\|_{L_2}^2, \quad \mathbf{x}, \mathbf{y} \in V. \end{aligned}$$

Since  $\mathbf{x} \in H^r(D)^2$ , we see that  $\text{curl } \mathbf{x} \in H^{r-1}(D)$ , and thus by the Sobolev embedding theorem  $\text{curl } \mathbf{x} \in L_\infty(D)$  thanks to the assumption  $r > 2$ . Therefore,  $U$  is finite on  $V$ .

We again use the same coupling construction as in [27], Section 6.2.3 (which in turn is a small modification of the construction from [18], Section 3.3.2). Fix  $\mathbf{x}, \mathbf{y} \in V$ . We take  $\mathbf{X}^{\mathbf{x}, \mathbf{y}} := \mathbf{u}^{\mathbf{x}}$  and let  $\mathbf{Y}^{\mathbf{x}, \mathbf{y}}$  be the solution to the same equation with the extra term in the right-hand side

$$\frac{1}{2} \lambda_{N+1} P_N (\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t)) dt$$

and started from the initial condition  $\mathbf{Y}_0^{\mathbf{x}, \mathbf{y}} = \mathbf{y}$ . Then by [18], formula (3.28) (see also [9], formulae (4.15)–(4.20)) there exists a universal  $C_1 > 0$  such that

$$q(\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t), \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t)) \leq q(\mathbf{x}, \mathbf{y}) \exp\left(-\lambda_{N+1} t + C_1 \int_0^t S(\mathbf{X}^{\mathbf{x}, \mathbf{y}}(s)) ds\right).$$

Therefore, Assumption **H1** is satisfied.

Assumption **H2** follows from the corresponding energy estimate [18], formula (3.29). Indeed, this estimate immediately implies that for some universal constants  $C_2, C_3 > 0$  we have

$$dU(\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t)) + C_2 S(\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t)) dt \leq C_3 \|\sigma\|_{L^p}^2 dt + M(t),$$

where  $M$  is a continuous local martingale with quadratic variation bounded by

$$d\langle M \rangle_t \leq 4\|\sigma\|_{L^p}^2 U(X^{\mathbf{x}, \mathbf{y}}(t)) dt \leq 4\|\sigma\|_{L^p}^2 (1 + S(X^{\mathbf{x}, \mathbf{y}}(t))) dt,$$

see [18], formula (3.30). This implies **H2**. Inequality (4.18) follows from (4.40), where we took  $C := C_1 C_3 / C_2$ .

Assumption **H3** follows from the uniqueness of a strong solution to (4.39) by exactly the same argument as in the proofs of Theorem 4.5 and 4.6.

Finally, we consider the Lyapunov function

$$V(\mathbf{x}) := U(\mathbf{x}) + \|\mathbf{x}\|_{L^2}^2, \quad \mathbf{x} \in V.$$

It is clear that for any  $M > 0$  the functions  $U$  and  $q$  are bounded on the level sets  $\{V \leq M\}$  and  $\{V \leq M\} \times \{V \leq M\}$ , respectively. Further, we have a simple energy estimate

$$(4.41) \quad d\|\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t)\|_{L^2}^2 + 2\|\Lambda^{\gamma/2} \mathbf{X}^{\mathbf{x}, \mathbf{y}}(t)\|_{L^2}^2 dt = \|\sigma\|_{L^2}^2 dt + d\tilde{M}(t),$$

for some martingale  $\tilde{M}$ . Since  $\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t)$  is a mean zero function, the Poincaré inequality implies

$$\|\Lambda^{\gamma/2} \mathbf{X}^{\mathbf{x}, \mathbf{y}}(t)\|_{L^2} \geq \|\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t)\|_{L^2}.$$

Combining this with (4.41) and with the energy estimate for  $\|\operatorname{curl} \mathbf{x}\|_{L^p}^p$  [18], formula (3.29), we obtain (4.23).

We recall that as shown in [18], Section 3.3, the process  $\mathbf{u}$  has an invariant measure. Therefore, all the conditions of Theorem 4.2 and Remark 4.3 hold. This immediately implies the statement of the theorem.  $\square$

**4.6. The 2D damped stochastically forced Euler–Voigt model.** Our next example is the 2D damped stochastically forced Euler–Voigt Model. Fix  $\gamma > 2/3$  and consider the following equation on the periodic box  $D := [-\pi, \pi]^2$ :

$$(4.42) \quad \begin{aligned} & d\mathbf{u}(z, t) + (\nu \mathbf{u}(z, t) + (\mathbf{u}_\gamma(z, t) \cdot \nabla) \mathbf{u}_\gamma(z, t) dt + \nabla p(z, t)) dt \\ &= \sum_{k=1}^m \sigma_k(z) dW^k(t), \end{aligned}$$

$$\mathbf{u}(\cdot, 0) = \mathbf{x}, \quad \nabla \cdot \mathbf{u} = 0, \quad \Lambda^\gamma \mathbf{u}_\gamma = \mathbf{u},$$

where  $\mathbf{u} = (u_1, u_2)$  is the unknown velocity field,  $p$  is the unknown pressure,  $\nu > 0$ ,  $m \in \mathbb{N}$ ,  $W = (W^1, W^2, \dots, W^m)$  is a standard  $m$ -dimensional Brownian motion. We assume that for any  $k = 1, \dots, m$  we have  $\sigma_k \in H^{1-\gamma/2}(D)^2$ ,  $\nabla \cdot \sigma_k = 0$ ,  $\int_D \sigma_k(z) dz = 0$ .

Introduce the following space:

$$V := \left\{ \boldsymbol{\varphi} \in H^{1-\gamma/2}(D)^2 : \nabla \cdot \boldsymbol{\varphi} = 0, \int_D \boldsymbol{\varphi}(z) dz = 0 \right\}.$$

The eigenvectors  $e_1, e_2, \dots$  of the operator  $-\Delta|_V$  (with the corresponding eigenvalues  $0 < \lambda_1 = 1 \leq \lambda_2 \leq \dots$ ) form a complete orthonormal basis of  $V$ .

It is known ([18], Proposition 3.4) that for any  $\mathbf{x} \in V$  equation (4.39) has a unique strong solution, which in the case of ambiguity will be further denoted by  $\mathbf{u}^{\mathbf{x}}$ . Moreover,  $\mathbf{u}$  is a Markov process with the state space  $V$ . Note however that the Feller property of  $\mathbf{u}$  is known only with respect to a weaker  $H^{-\gamma/2}$  norm rather than  $H^{1-\gamma/2}$  norm. This causes several technical difficulties, which however are not very crucial and can be successfully resolved.

First, we note that the process  $\mathbf{u}^{\mathbf{x}}$  is progressively measurable. This follows from the facts that  $\mathbf{u}^{\mathbf{x}}$  is stochastically continuous in the weaker  $H^{-\gamma/2}$  norm and the Borel  $\sigma$ -algebras generated by the  $H^{-\gamma/2}$ - and the  $H^{1-\gamma/2}$ -norms are the same.

Next, we cannot apply here Theorem 4.2, which requires the Feller property of the Markov semigroup. Therefore, to show the exponential convergence of transition probabilities we apply [23], Theorem 4.8, which does not require the Feller property, together with Lemma 4.4 and Theorem 2.4. Note that the existence of the invariant measure of  $\mathbf{u}$  does not follow from these considerations; fortunately, it was already established in [18], Proposition 3.5.

In this section, we also use the notation  $\|\varphi\|_{H^s} := \|\Lambda^s \varphi\|_{L^2}$ ,  $s \in \mathbb{R}$ .

**THEOREM 4.8.** *There exists a universal  $C > 0$  such that if for some  $N \in \mathbb{Z}_+$  one has*

$$\text{Span}(e_k, k = 1, \dots, N) \subset \text{Span}(\sigma_k, k = 1, \dots, m)$$

and

$$(4.43) \quad \lambda_{N+1}^{\gamma/2-1/3} > C \frac{\|\text{curl } \boldsymbol{\sigma}\|_{H^{-\gamma/2}}^2}{\nu^3},$$

then equation (4.42) has a unique invariant measure  $\pi$ . Further, there exist constants  $C_1 > 0$ ,  $r > 0$  such that

$$(4.44) \quad W_{\|\cdot\|_{H^{-\gamma/2}} \wedge 1}(\text{Law}(\mathbf{u}^{\mathbf{x}}(t)), \pi) \leq C_1 (1 + \|\mathbf{x}\|_{H^{1-\gamma/2}}^2) e^{-rt}, \quad t \geq 0, \mathbf{x} \in V.$$

**PROOF.** First, let us check that Assumptions **H1–H3** hold.

Put

$$U(\mathbf{x}) = S(\mathbf{x}) := \|\Lambda^{-\gamma/2} \text{curl } \mathbf{x}\|_{L^2}^2, \quad q(\mathbf{x}, \mathbf{y}) = \|\Lambda^{-\gamma/2}(\mathbf{x} - \mathbf{y})\|_{L^2}^2, \quad \mathbf{x}, \mathbf{y} \in V.$$

We use the same coupling construction as in [27], Section 6.2.4 (which again is a minor modification of the corresponding construction from [18], Section 3.4.4). Fix  $\mathbf{x}, \mathbf{y} \in V$ . Take  $\mathbf{X}^{\mathbf{x}, \mathbf{y}} := \mathbf{u}^{\mathbf{x}}$  and let  $\mathbf{Y}^{\mathbf{x}, \mathbf{y}}$  be the solution to the same equation with the extra term in the right-hand side

$$\frac{1}{2} \nu \lambda_{N+1}^{\gamma/2-1/3} P_N(\mathbf{X}^{\mathbf{x}, \mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x}, \mathbf{y}}(t)) dt$$

and started from the initial condition  $\mathbf{Y}_0^{\mathbf{x},\mathbf{y}} = \mathbf{y}$  (recall that  $\gamma > 2/3$ ). Then [18], formula (3.45), and the Sobolev embedding  $H^{1/3} \subset L_3$  imply that for some universal constant  $C_1 > 0$ ,

$$(4.45) \quad \begin{aligned} & \frac{1}{2} \frac{d}{dt} q(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t), \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)) + \nu q(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t), \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)) \\ & + \frac{1}{2} \nu \lambda_{N+1}^{\gamma/2-1/3} \|P_N \Lambda^{-\gamma/2}(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t))\|_{L_2}^2 \\ & \leq C_1 \|\Lambda^{-\gamma}(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t))\|_{H^{1/3}}^2 \|\Lambda^{-\gamma} \operatorname{curl}(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t))\|_{H^{1/3}}. \end{aligned}$$

By the Poincaré inequality,

$$\begin{aligned} & C_1 \|\mathbf{X}^{\mathbf{x},\mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)\|_{H^{-\gamma+1/3}}^2 \|\operatorname{curl}(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t))\|_{H^{-\gamma+1/3}} \\ & \leq \frac{1}{2} \|\mathbf{X}^{\mathbf{x},\mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)\|_{H^{-\gamma+1/3}}^2 \\ & \quad \times (\nu \lambda_{N+1}^{\gamma/2-1/3} + \nu^{-1} \lambda_{N+1}^{-\gamma/2+1/3} C_1^2 \|\operatorname{curl}(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t))\|_{H^{-\gamma+1/3}}^2) \\ & \leq \frac{1}{2} \nu \lambda_{N+1}^{\gamma/2-1/3} \|P_N \Lambda^{-\gamma/2}(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t))\|_{L_2}^2 + \frac{1}{2} \nu q(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t), \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)) \\ & \quad + \frac{1}{2} \nu^{-1} \lambda_{N+1}^{-\gamma/2+1/3} C_1^2 q(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t), \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)) S(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t)). \end{aligned}$$

Combining this with (4.45) and applying the Gronwall inequality, we obtain

$$q(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t), \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)) \leq q(\mathbf{x}, \mathbf{y}) \exp\left(-\nu t + C_1^2 \nu^{-1} \lambda_{N+1}^{-\gamma/2+1/3} \int_0^t S(\mathbf{X}^{\mathbf{x},\mathbf{y}}(s)) ds\right).$$

Therefore, assumption **H1** is satisfied.

Assumption **H2** follows immediately from [18], formula (3.35). Indeed, rewriting this formula in our notation, we get

$$dU(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t)) + 2\nu S(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t)) dt = \|\operatorname{curl} \boldsymbol{\sigma}\|_{H^{-\gamma/2}}^2 dt + M(t),$$

where  $M$  is a continuous local martingale with quadratic variation bounded by

$$d\langle M \rangle_t \leq 4 \|\operatorname{curl} \boldsymbol{\sigma}\|_{H^{-\gamma/2}}^2 S(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t)) dt.$$

This yields **H2**. Inequality (4.18) follows from (4.43), where we took  $C := C_1^2/2$ .

To verify **H3**, we use the same argument as in the proofs of Theorems 4.5 and 4.6. Note that condition 3 of Lemma 4.1 holds since by the Poincaré inequality for any  $t \geq 0$ ,

$$\begin{aligned} \|P_N(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t))\|_{L_2}^2 & \leq \lambda_N^{\gamma/2} \|P_N(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t) - \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t))\|_{H^{-\gamma/2}}^2 \\ & \leq \lambda_N^{\gamma/2} q(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t), \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)). \end{aligned}$$

Thus all the assumptions **H1–H3** hold.

Finally, we consider the Lyapunov function

$$V(\mathbf{x}) := \|\operatorname{curl} \mathbf{x}\|_{H^{-\gamma/2}}^2 + \|\mathbf{x}\|_{H^{-\gamma/2}}^2, \quad \mathbf{x} \in V.$$

Obviously, for any  $M > 0$  the functions  $U$  and  $q$  are bounded on the level sets  $\{V \leq M\}$  and  $\{V \leq M\} \times \{V \leq M\}$ , respectively. Further, by adding energy estimates [18], formulae (3.34) and (3.35), we obtain

$$EV(\mathbf{u}^{\mathbf{x}}(t)) \leq V(\mathbf{x})e^{-2\nu t} + K_V, \quad t \geq 0$$

for some universal constant  $K_V > 0$ .

Now we are ready to complete the proof of the theorem. The existence of an invariant measure  $\pi$  was established in [18], Proposition 3.5. Since  $V$  is a Lyapunov function for  $P_t$ ,  $V$  is integrable with respect to any invariant measure (see, e.g., [20], Proposition 4.24). Applying Lemma 4.4 and Theorem 2.4(i), (iii), we see that for some  $N > 0$ ,  $t > 0$ ,  $\alpha > 0$ ,  $C > 0$  the distance-like function

$$d_N(\mathbf{x}_1, \mathbf{x}_2) := (Ne^{C(U(\mathbf{x}_1) \wedge U(\mathbf{x}_2))} q(\mathbf{x}_1, \mathbf{x}_2)^\alpha) \wedge 1$$

is contracting for  $P_t$  and the set  $\{V \leq 4K_V\}$  is  $d_N$ -small for  $P_t$ . Thus all the conditions of [23], Theorem 4.8, are satisfied. Hence the invariant measure is unique and the exponential bound on the convergence rate (4.44) follows from [23], formula (4.6), and the integrability of  $V$  with respect to  $\pi$ . Note that to use the integrability, one first has to check that  $W_d(P_t(x, \cdot), P_t(y, \cdot))$  is measurable. This is the point, where the lack of the Feller property causes minor technical difficulty; in particular we can not refer here neither to [23], Lemma 4.13, nor to [26], Theorem 4.4.3. Note however, that the mapping

$$V \times V \ni (x, y) \mapsto (P_t(x, \cdot), P_t(y, \cdot)) \in \mathcal{P}(V) \times \mathcal{P}(V)$$

is measurable (one can get this easily using the  $H^{-\gamma/2}$ -Feller property), and the function  $d_N$  is continuous on  $V \times V$ . Then the required measurability follows by [42], Corollary 5.22.  $\square$

*4.7. Boussinesq approximation for Rayleigh–Bénard convection.* Finally, we consider Boussinesq approximation for the Rayleigh–Bénard convection perturbed by additive noise. The physical motivation behind the model as well as the relevance of the model for fluid dynamics are explained in detail in [16].

Consider the following system of stochastic equations evolving on a domain  $D := [0, L] \times [0, 1]$ :

$$(4.46) \quad \begin{aligned} & \frac{1}{Pr} (d\mathbf{u}(z, t) + (\mathbf{u}(z, t) \cdot \nabla) \mathbf{u}(z, t) dt) \\ &= \Delta \mathbf{u}(z, t) dt + Ra \mathbf{i}_2 T(z, t) dt - \nabla p(z, t) dt \\ &+ \sum_{k=1}^{m_1} \sigma_k(z) dW^k(t), \quad z \in D, t \geq 0; \end{aligned}$$

$$(4.47) \quad \begin{aligned} & dT(z, t) + (\mathbf{u}(z, t) \cdot \nabla) T(z, t) dt \\ &= \Delta T(z, t) dt + \sum_{k=1}^{m_2} \rho_k(z) dB^k(t), \quad z \in D, t \geq 0; \end{aligned}$$

$$(\mathbf{u}(0), T(0)) = \mathbf{x}, \quad \nabla \cdot \mathbf{u} = 0,$$

$$\mathbf{u}|_{z_2=0} = \mathbf{u}|_{z_2=1} = 0, \quad T|_{z_2=0} = \tilde{Ra}, \quad T|_{z_2=1} = 0$$

$$\mathbf{u}, T \text{ are periodic in } z_1,$$

where  $\mathbf{u} = (u_1, u_2)$  is the unknown velocity field;  $p$  is the unknown pressure;  $T$  is the unknown temperature;  $m_1, m_2 \in \mathbb{Z}_+$ ;  $W = (W^1, \dots, W^{m_1})$  and  $B = (B^1, \dots, B^{m_2})$  are standard independent  $m_1$ - and  $m_2$ -dimensional Brownian motions, respectively.  $Pr, Ra, \tilde{Ra}$  denote positive constants and correspond to some physical parameters:  $Pr$  is the Prandtl number,  $Ra$  and  $\tilde{Ra}$  are Rayleigh numbers. We also used the notation  $\mathbf{i}_2 := (0, 1)$ .

Introduce the following spaces:

$$V_1 := \{\varphi \in H^1(D)^2 : \nabla \cdot \varphi = 0, \varphi|_{z_2=0,1} = 0, \varphi \text{ is periodic in } z_1\},$$

$$V_2 := \{\varphi \in H^1(D) : \varphi|_{z_2=0,1} = 0, \varphi \text{ is periodic in } z_1\},$$

$$\tilde{V}_2 := \{\varphi \in H^1(D) : \varphi|_{z_2=0} = \tilde{R}a, \varphi|_{z_2=1} = 0, \varphi \text{ is periodic in } z_1\}.$$

We denote by  $H_1$  and  $H_2$  the completions of  $V_1$  and  $V_2$  with respect to  $L^2(D)^2$ - and  $L^2(D)$ -norms, correspondingly. Put  $V := V_1 \times V_2$ ,  $H := H_1 \times H_2$ . The eigenvectors  $e_1, e_2, \dots$  of the Stokes operator (with the corresponding eigenvalues  $0 < \lambda_1 \leq \lambda_2 \leq \dots$ ) form a complete orthonormal basis of  $H_1$ , and the eigenvectors  $j_1, j_2, \dots$  of the Laplace operator (with the corresponding eigenvalues  $0 < \mu_1 \leq \mu_2 \leq \dots$ ) form a complete orthonormal basis of  $H_2$ . For  $N \in \mathbb{Z}_+$ , we denote by  $P_N^{H_i}$  the projection onto the span of the first  $N$  eigenvectors of  $H_i$ ,  $i = 1, 2$ . By  $\|\cdot\|_{H_1}$ ,  $\|\cdot\|_{H_2}$ ,  $\|\cdot\|_H$ , we denote the  $L_2$  norms in the spaces  $H_1, H_2, H$ , correspondingly.

We assume that for each  $k = 1, \dots, m_1$  we have  $\sigma_k \in H^2(D)^2 \cap V_1$  and for each  $k = 1, \dots, m_2$  we have  $\rho_k \in H^2(D) \cap V_2$ .

It is known ([16], Proposition 2.1) that for any  $\mathbf{x} \in H$  a system of equations (4.46)–(4.47) has a unique strong solution, which in the case of ambiguity will be further denoted by  $(\mathbf{u}^{\mathbf{x}}, T^{\mathbf{x}})$ . Moreover,  $(\mathbf{u}, T)$  is a Feller Markov process with the state space  $H$  and for almost all  $t > 0$  we have  $(\mathbf{u}^{\mathbf{x}}(t), T^{\mathbf{x}}(t)) \in V_1 \times \tilde{V}_2$ .

It was shown in [15], Theorem 1.1, that if the velocity  $u$  and temperature  $T$  satisfies periodic boundary conditions, then the process  $(\mathbf{u}, T)$  is exponentially ergodic. However, as was noted in [16], “using periodic boundary conditions in the vertical directions is not appropriate from the physical point of view.” Therefore, following [16], we equip the system (4.46)–(4.47) with mixed periodic and nonhomogeneous Dirichlet boundary conditions that are physically relevant.

Note that the methods of [15], Theorem 1.1, are not applicable to treat the system with mixed boundary conditions; see also a related discussion in [18], p. 619, lines 36–42. However, using Theorem 4.2, we are able to establish exponential ergodicity of  $(\mathbf{u}^{\mathbf{x}}, T^{\mathbf{x}})$  with mixed boundary conditions in the case of small Rayleigh numbers.

**THEOREM 4.9.** *Suppose that*

$$\tilde{R}aRa < \pi^2\sqrt{2} - 1.$$

*Assume further that there exists  $N_1, N_2 \in \mathbb{Z}_+$  such that*

$$\mathbf{P}_{N_1}^{H_1} H_1 \subset \text{Span}(\sigma_k, k = 1, \dots, m_1); \quad \mathbf{P}_{N_2}^{H_2} H_2 \subset \text{Span}(\rho_k, k = 1, \dots, m_2);$$

*and*

$$(4.48) \quad \lambda_{N_1+1} > C_1(Pr)^{-3/2}(Pr\|\sigma\|_{H_1}^2 + Ra^2\|\rho\|_{H_2}^2),$$

$$(4.49) \quad \mu_{N_2+1} > C_1(Pr)^{-1/2}(Pr\|\sigma\|_{H_1}^2 + Ra^2\|\rho\|_{H_2}^2) + C_2Pr,$$

*where  $C_1 = \frac{(1+1/\pi)(\sqrt{2}+16Pr^{-1/2})}{1 \wedge (Ra^2(2-\pi^{-4}(1+RaRa)^2))}$ ,  $C_2 = 2\pi^{-2}(Ra + \tilde{R}a/Pr)^2$ .*

*Then the system (4.46)–(4.47) has a unique invariant measure  $\pi$ . Further, there exist constants  $C_3 > 0, r > 0$  such that*

$$W_{\|\cdot\|_{H \wedge 1}}(\text{Law}(\mathbf{u}^{\mathbf{x}}(t), T^{\mathbf{x}}(t)), \pi) \leq C_3(1 + \|\mathbf{x}\|_H^2)e^{-rt}, \quad t \geq 0, \mathbf{x} \in H.$$

REMARK 4.10. Note that conditions (4.48)–(4.49) are satisfied in the following two different scenarios. The first case is when the noise acts both on the velocity  $\mathbf{u}$  and temperature  $T$  (i.e., both  $N_1$  and  $N_2$  are large enough). The second case is when the noise acts only on temperature (i.e.,  $N_2$  is large enough). Then the system is ergodic provided that the Prandtl number is large enough (and  $N_1$  can be small, in particular equal to 0).

PROOF OF THEOREM 4.9. First, as in [16], we introduce a shifted temperature process

$$\Theta(t, z) := T(t, z) - \widetilde{R}a(1 - z_2), \quad t \geq 0, z \in D.$$

It is clear from the definition that  $\Theta$  is periodic in  $z_1$  and satisfies homogeneous boundary conditions  $\Theta|_{z_2=0} = \Theta|_{z_2=1} = 0$ .

Let us check that all the conditions of Theorem 4.2 are satisfied for the Markov process  $(\mathbf{u}, \Theta)$ . Since  $\Theta$  is just  $T$  shifted by some nonrandom function, which is constant in time, exponential ergodicity of the Markov process  $(\mathbf{u}, \Theta)$  would imply exponential ergodicity of  $(\mathbf{u}, T)$ .

Take

$$\begin{aligned} U(\mathbf{x}) &:= \frac{1}{Pr} \|(x_1, x_2)\|_{H_1}^2 + Ra^2 \|x_3\|_{H_2}^2, & S(\mathbf{x}) &:= \|\nabla \mathbf{x}\|, \\ q(\mathbf{x}, \mathbf{y}) &:= \|\mathbf{x} - \mathbf{y}\|_H, & \mathbf{x}, \mathbf{y} &\in H. \end{aligned}$$

We use the coupling construction, which is a small modification of the construction from [16], Section 5.2. Fix  $\mathbf{x}, \mathbf{y} \in H$ . We take  $\mathbf{X}^{\mathbf{x}, \mathbf{y}} := (\mathbf{u}^{\mathbf{x}}, \Theta^{\mathbf{x}})$  and let  $\mathbf{Y}^{\mathbf{x}, \mathbf{y}} := (\tilde{\mathbf{u}}, \tilde{\Theta})$  be the solution to the same system of equations with the initial condition  $\mathbf{y}$  and the additional control terms:

$$\begin{aligned} \frac{1}{Pr} (d\tilde{\mathbf{u}} + (\tilde{\mathbf{u}} \cdot \nabla) \tilde{\mathbf{u}} dt) &= \Delta \tilde{\mathbf{u}} dt + Ra i_2 \tilde{\Theta} dt - \nabla \tilde{p} dt \\ &\quad + \frac{1}{2} \lambda_{N_1+1} P_{N_1}^{H_1} (\mathbf{u}^{\mathbf{x}} - \tilde{\mathbf{u}}) dt + \sum_{k=1}^{m_1} \sigma_k dW^k, \\ d\tilde{\Theta} + (\tilde{\mathbf{u}} \cdot \nabla) \tilde{\Theta} dt &= \widetilde{R}a \tilde{u}_2 dt + \Delta \tilde{\Theta} dt \\ &\quad + \frac{1}{2} \mu_{N_2+1} P_{N_2}^{H_2} (\Theta^{\mathbf{x}} - \tilde{\Theta}) dt + \sum_{k=1}^{m_2} \rho_k dB^k, \\ (\tilde{\mathbf{u}}(0), \tilde{\Theta}(0)) &= \mathbf{y}, \quad \nabla \cdot \tilde{\mathbf{u}} = 0, \quad \tilde{\mathbf{u}}|_{z_2=0} = \tilde{\mathbf{u}}|_{z_2=1} = 0, \\ \tilde{\Theta}|_{z_2=0} &= \tilde{\Theta}|_{z_2=1} = 0, \quad \tilde{\mathbf{u}}, \tilde{\Theta} \text{ are periodic in } z_1. \end{aligned}$$

Then it follows from [16], formulae (5.9)–(5.12), that

$$\begin{aligned} &\frac{1}{2} \frac{d}{dt} (\|\mathbf{u}^{\mathbf{x}} - \tilde{\mathbf{u}}\|_{H_1}^2 + \|\Theta^{\mathbf{x}} - \tilde{\Theta}\|_{H_2}^2) \\ &\quad + \frac{Pr}{2} \lambda_{N_1+1} \|\mathbf{u}^{\mathbf{x}} - \tilde{\mathbf{u}}\|_{H_1}^2 + \frac{1}{2} \mu_{N_2+1} \|\Theta^{\mathbf{x}} - \tilde{\Theta}\|_{H_2}^2 \\ (4.50) \quad &\leq \frac{(\widetilde{R}a + PrRa)^2}{\pi^2 Pr} \|\Theta^{\mathbf{x}} - \tilde{\Theta}\|_{H_2}^2 \\ &\quad + \frac{C_4}{\sqrt{Pr}} (\|\mathbf{u}^{\mathbf{x}} - \tilde{\mathbf{u}}\|_{H_1}^2 + \|\Theta^{\mathbf{x}} - \tilde{\Theta}\|_{H_2}^2) (\|\nabla \mathbf{u}^{\mathbf{x}}\|_{H_1}^2 + \|\nabla \Theta^{\mathbf{x}}\|_{H_2}^2), \end{aligned}$$

where  $C_4 := (1 + 1/\pi)(1/\sqrt{2} + 8/\sqrt{Pr})$ . Here, we have also used the Poincaré inequality  $\|\nabla\varphi\|_{L_2} \geq \pi\|\varphi\|_{L_2}$  and the Sobolev embedding  $\|\varphi\|_{L_4}^4 \leq 2(1 + 1/\pi)\|\varphi\|_{L_2}^2\|\nabla\varphi\|_{L_2}^2$ ; both are valid for any  $\varphi \in V_2$ .

Inequality (4.50) and the Gronwall inequality imply for any  $t \geq 0$ ,

$$\begin{aligned} & q(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t), \mathbf{Y}^{\mathbf{x},\mathbf{y}}(t)) \\ & \leq q(\mathbf{x}, \mathbf{y}) \exp\left(-\zeta t + \frac{2C_4}{\sqrt{Pr}} \int_0^t \|\nabla\mathbf{u}^{\mathbf{x}}(s)\|_{H_1}^2 + \|\nabla\Theta^{\mathbf{x}}(s)\|_{H_2}^2 ds\right), \end{aligned}$$

where we put  $\zeta := (Pr\lambda_{N_1+1}) \wedge (\mu_{N_2+1} - \frac{2(\widetilde{Ra}+PrRa)^2}{\pi^2 Pr})$ . Therefore, assumption **H1** is satisfied.

To verify Assumption **H2**, we use the corresponding energy estimates for  $\mathbf{u}^{\mathbf{x}}$  and  $\Theta^{\mathbf{x}}$  (see [16], formulae (3.32) and (3.35)). We multiply the first of these estimates by  $1/Pr$ , the second one by  $Ra^2$  and get

$$\begin{aligned} (4.51) \quad & \frac{1}{Pr} d\|\mathbf{u}^{\mathbf{x}}(t)\|_{H_1}^2 + Ra^2 d\|\Theta^{\mathbf{x}}(t)\|_{H_2}^2 + \|\nabla\mathbf{u}^{\mathbf{x}}(t)\|_{H_1}^2 dt \\ & + Ra^2 \left(2 - \frac{(1 + Ra\widetilde{Ra})^2}{\pi^4}\right) \|\nabla\Theta^{\mathbf{x}}(t)\|_{H_2}^2 dt \\ & \leq (Pr\|\sigma\|_{H_1}^2 + Ra^2\|\rho\|_{H_2}^2) dt + dM(t), \end{aligned}$$

where  $M$  is a continuous local martingale with  $M(0) = 0$ ; its quadratic variation is bounded by

$$d\langle M \rangle_t \leq 4(\|\sigma\|_{H_1}^2 + Ra^4\|\rho\|_{H_2}^2)(\|\mathbf{u}^{\mathbf{x}}(t)\|_{H_1}^2 + \|\Theta^{\mathbf{x}}(t)\|_{H_2}^2) dt.$$

Inequality (4.51) yields

$$dU(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t)) + \mu S(\mathbf{X}^{\mathbf{x},\mathbf{y}}(t)) dt \leq (Pr\|\sigma\|_{H_1}^2 + Ra^2\|\rho\|_{H_2}^2) dt + dM(t),$$

where  $\mu := 1 \wedge (Ra^2(2 - \frac{(1+Ra\widetilde{Ra})^2}{\pi^4}))$ ; this implies **H2**. Inequality (4.18) follows from (4.48) and (4.49).

Assumption **H3** follows from the uniqueness of a strong solution to (4.46)–(4.47) by the same argument as in the proofs of Theorem 4.5 and 4.6.

Finally, we consider the Lyapunov function

$$V(\mathbf{x}) := U(\mathbf{x}), \quad \mathbf{x} \in H.$$

We immediately see by definition, that for any  $M > 0$  the functions  $U$  and  $q$  are bounded on the level sets  $\{V \leq M\}$  and  $\{V \leq M\} \times \{V \leq M\}$ , respectively. Inequality (4.23) follows directly from the energy estimate (4.51) and the Poincaré inequality.

Thus all the conditions of Theorem 4.2 hold. This immediately implies the statement of the theorem.  $\square$

## APPENDIX. DISTANCE BETWEEN THE LAW OF AN ITÔ PROCESS AND THE WIENER MEASURE

In the Appendix, we provide useful bounds on the various distances between the law of an Itô process and the Wiener measure under different sets of conditions. We use these bounds throughout the paper; however, we believe that they are also of independent interest.

We begin by recalling that for a pair of probability measures  $\mu \ll \nu$  over a measurable space  $(X, \mathcal{X})$  the *Kullback–Leibler (KL-) divergence of  $\mu$  from  $\nu$*  is defined by

$$D_{\text{KL}}(\mu \parallel \nu) := \int_X \log \frac{d\mu}{d\nu} d\mu = \int_X \frac{d\mu}{d\nu} \log\left(\frac{d\mu}{d\nu}\right) d\nu.$$

If a measure  $\mu$  is not absolutely continuous with respect to  $\nu$ , then for convenience we put  $D_{\text{KL}}(\mu \parallel \nu) := +\infty$ . KL-divergence is a stronger measure of difference between probability distributions than the total variation distance; they are connected in the following way.

LEMMA A.1. *Let  $\mu$  and  $\nu$  be probability measures over  $(X, \mathcal{X})$ . Then*

$$(A.1) \quad d_{\text{TV}}(\mu, \nu) \leq \sqrt{\frac{1}{2} D_{\text{KL}}(\mu \parallel \nu)};$$

$$(A.2) \quad d_{\text{TV}}(\mu, \nu) \leq 1 - \frac{1}{2} e^{-D_{\text{KL}}(\mu \parallel \nu)};$$

$$(A.3) \quad \int_X \left( \log \frac{d\mu}{d\nu} \right)_+ d\mu \leq D_{\text{KL}}(\mu \parallel \nu) + \log 2.$$

Further, for any  $N > 1$  and any set  $A \in \mathcal{X}$ ,

$$(A.4) \quad \nu(A) \geq \frac{1}{N} \mu(A) - \frac{D_{\text{KL}}(\mu \parallel \nu) + \log 2}{N \log N}.$$

PROOF. We can consider only the case when  $\mu \ll \nu$ ; otherwise,  $D_{\text{KL}}(\mu \parallel \nu) := +\infty$  and all the bounds trivially hold. (A.1) is the classical Pinsker inequality (see, e.g., [40], Lemma 2.5.(i)). Inequality (A.2) is [40], formula (2.25). To establish (A.3), denote  $f := \frac{d\mu}{d\nu}$ . Then

$$\int_X (\log f)_+ d\mu = D_{\text{KL}}(\mu \parallel \nu) + \int_X (\log f)_- d\mu$$

and

$$\begin{aligned} \int_X (\log f)_- d\mu &\leq \log \left( \int_X e^{(\log f)_-} d\mu \right) = \log \left( \int_X \max(f^{-1}, 1) f d\nu \right) \\ &\leq \log \left( \int_X (f + 1) d\nu \right) = \log 2. \end{aligned}$$

This implies (A.3).

To prove (A.4), we fix  $N > 1$  and a measurable set  $A$ . We have

$$\begin{aligned} \nu(A) &= \int_A \frac{d\nu}{d\mu} d\mu \\ &\geq \frac{1}{N} \int_A \mathbb{1} \left( \frac{d\nu}{d\mu} \geq \frac{1}{N} \right) d\mu \\ &\geq \frac{1}{N} \mu(A) - \frac{1}{N} \int_X \mathbb{1} \left( \frac{d\mu}{d\nu} \geq N \right) d\mu \\ &\geq \frac{1}{N} \mu(A) - \frac{1}{N \log N} \int_X \log \left( \frac{d\mu}{d\nu} \right)_+ d\mu. \end{aligned}$$

This combined with (A.3) yields (A.4).  $\square$

Consider now a  $d$ -dimensional ( $d \in \mathbb{N}$ ) Itô process  $(\xi_t)_{t \geq 0}$  with  $\xi_0 = 0$  and

$$(A.5) \quad d\xi_t = \beta_t dt + dW_t, \quad t \geq 0,$$

where  $W$  is a Wiener process in  $\mathbb{R}^d$ , and  $(\beta_t)_{t \geq 0}$  is a progressively measurable process. Denote by  $\mu_\xi$  the law of the process  $\xi$  in  $\mathcal{C}([0, \infty), \mathbb{R}^d)$ , and by  $\mu_W$  the law of  $W$ ; the latter will be also called the *Wiener measure* on  $\mathcal{C}([0, \infty), \mathbb{R}^d)$ .

THEOREM A.2.

$$D_{\text{KL}}(\mu_\xi \parallel \mu_W) \leq \frac{1}{2} \mathbb{E} \int_0^\infty |\beta_t|^2 dt.$$

This result is not new, it is essentially [41], Theorem 8, up to a minor difference which will be discussed later in Remark A.3. However, apparently it is not widely known, hence for the convenience of the reader we give its proof together with a short discussion. Recall that the Girsanov theorem (see, e.g., [29], Theorem 6.2) states that  $\xi$  is a Wiener process w.r.t. the new probability measure  $d\mathbb{Q} := \mathcal{E} d\mathbb{P}$  with

$$\mathcal{E} := \exp\left(-\int_0^\infty \beta_t dW_t - \frac{1}{2} \int_0^\infty |\beta_t|^2 dt\right)$$

under the assumption that

$$(A.6) \quad \mathbb{E}\mathcal{E} = 1.$$

To verify this assumption, the *Novikov condition* is most commonly used:

$$(A.7) \quad \mathbb{E} \exp\left(\frac{1}{2} \int_0^\infty |\beta_t|^2 dt\right) < \infty.$$

This condition is nonimprovable in the sense that the similar condition

$$\mathbb{E} \exp\left(\left(\frac{1}{2} - \varepsilon\right) \int_0^\infty |\beta_t|^2 dt\right) < \infty$$

with any  $\varepsilon > 0$  is not sufficient for (A.6); see [29], Example 6, Chapter 6.2. On the other hand, under a much weaker condition

$$\mathbb{P}\left(\int_0^\infty |\beta_t|^2 dt < \infty\right) = 1$$

the absolute continuity  $\mu_\xi \ll \mu_W$  holds (see, e.g., [29], Theorem 7.4), though not much information about the Radon–Nikodym density  $d\mu_\xi/d\mu_W$  is available. Theorem A.2 has the same virtue with the latter result: under a weak condition

$$(A.8) \quad \mathbb{E} \int_0^\infty |\beta_t|^2 dt < \infty$$

a bound for the KL-divergence of  $\mu_\xi$  from  $\mu_W$  is given without specifying  $d\mu_\xi/d\mu_W$ . The proof is based on a similar localization argument.

**PROOF OF THEOREM A.2.** We begin by observing that if condition (A.8) is not satisfied, then the statement of the theorem trivially holds. Thus from now on we can assume (A.8).

Assume first that additionally (A.7) holds true. Then (A.6) is satisfied; see, for example, [29], Theorem 6.1. Hence  $\mu_\xi \sim \mu_W$ , and the function

$$p(x) := \frac{d\mu_W}{d\mu_\xi}(x), \quad x \in \mathcal{C}([0, \infty), \mathbb{R}^d)$$

satisfies

$$p(\xi) = \mathbb{E}[\mathcal{E} | \sigma(\xi)];$$

see [29], Theorem 7.3. Therefore,

$$(A.9) \quad \begin{aligned} D_{\text{KL}}(\mu_\xi \parallel \mu_W) &= - \int_{\mathcal{C}([0, \infty))} \log \frac{d\mu_W}{d\mu_\xi} d\mu_\xi = - \int_{\mathcal{C}([0, \infty))} \log p d\mu_\xi \\ &= -\mathbb{E} \log p(\xi) = -\mathbb{E} \log(\mathbb{E}[\mathcal{E} | \sigma(\xi)]) \leq -\mathbb{E} \log \mathcal{E}, \end{aligned}$$

where in the last inequality we have used the conditional Jensen inequality. Assumption (A.7) yields (A.8), and thus

$$\mathbb{E} \int_0^\infty \beta_t dW_t = 0.$$

Therefore,

$$-\mathbb{E} \log \mathcal{E} = \frac{1}{2} \mathbb{E} \int_0^\infty |\beta_t|^2 dt,$$

which gives the required bound.

Now we can treat the general case. Define a sequence of stopping times

$$\tau_n := \inf \left\{ t : \int_0^t |\beta_s|^2 ds \geq n \right\}, \quad n \in \mathbb{Z}_+.$$

Let  $\beta_t^n := \beta_t \mathbb{1}_{t \leq \tau_n}$ . Consider the corresponding Itô process:

$$d\xi_t^n = \beta_t^n dt + dW_t, \quad \xi_0^n = 0$$

and denote its law by  $\mu_{\xi^n}$ . By definition, the process  $\beta^n$  satisfies the Novikov condition (A.7). Thus, by the first part of the proof

$$(A.10) \quad D_{\text{KL}}(\mu_{\xi^n} \parallel \mu_W) \leq \frac{1}{2} \mathbb{E} \int_0^{\tau_n} |\beta_t|^2 dt \leq \frac{1}{2} \mathbb{E} \int_0^\infty |\beta_t|^2 dt.$$

Note that the processes  $\xi^n$  and  $\xi$  coincide on the set  $\{\tau_n = \infty\}$ . On the other hand, condition (A.8) implies

$$\mathbb{P}(\tau_n < \infty) \leq \mathbb{P}\left(\int_0^\infty |\beta_t|^2 dt \geq n\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Hence

$$(A.11) \quad d_{\text{TV}}(\mu_{\xi^n}, \mu_\xi) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Denote  $f_n := \frac{d\mu_{\xi^n}}{d\mu_W}$ ,  $f := \frac{d\mu_\xi}{d\mu_W}$ . It follows from (A.11) that  $f_n \rightarrow f$  in  $L_1(\mu_W)$  as  $n \rightarrow \infty$ . Therefore,  $f_n \log f_n$  converges in measure  $\mu_W$  to  $f \log f$  as  $n \rightarrow \infty$ . Since the sequence  $(f_n \log f_n)_{n \in \mathbb{Z}_+}$  is uniformly bounded from below by a constant, an application of Fatou's lemma and (A.10) give

$$\begin{aligned} D_{\text{KL}}(\mu_\xi \parallel \mu_W) &= \int_{\mathcal{C}([0, \infty))} f \log f d\mu_W \leq \liminf_{n \rightarrow \infty} \int_{\mathcal{C}([0, \infty))} f_n \log f_n d\mu_W \\ &\leq \frac{1}{2} \mathbb{E} \int_0^\infty |\beta_t|^2 dt. \end{aligned}$$

This completes the proof of the theorem.  $\square$

**REMARK A.3.** A minor difference between Theorem A.2 and [41], Theorem 8, is that the latter one, prior to the localization argument, uses invertibility of Wiener maps, and thus requires  $\beta$  to be progressively measurable w.r.t. the natural filtration of  $W$ . In the above proof this step is made using the Jensen inequality (A.9), hence this hidden limitation is not involved.

**REMARK A.4.** The proof of Theorem A.2 without any substantial changes can be transferred to the infinite-dimensional setting. Namely, using [10], Theorem 10.14, instead of the classical Girsanov theorem, one gets essentially the same bound for  $D_{\text{KL}}(\mu_\xi \parallel \mu_W)$  in the setting where  $\xi$ ,  $W$  in (A.5) are cylindrical processes in a Hilbert space  $H$ . In this case,  $\mu_\xi$ ,

$\mu_W$  are *cylindrical* measures; this requires some additional technical steps, for example, the specification of the Radon–Nikodym derivative for cylindrical measures. In order to keep the exposition relatively simple and clear, we do not develop this possibility in further details here. Note also that Theorem A.2 holds true for a finite time interval  $[0, T]$  as well: one easily gets that by extending a process  $\beta_t, t \in [0, T]$  to  $[0, \infty)$  taking  $\beta_t = 0, t > T$ .

As mentioned before, Theorem A.2 provides a nontrivial bound only if condition (A.8) holds. Now let us further relax this condition and assume that only for some  $\delta \in (0, 1)$

$$(A.12) \quad M_\delta := \mathbb{E} \left( \int_0^\infty |\beta_t|^2 dt \right)^\delta < \infty.$$

In this case, it is also possible to measure the difference between  $\mu_\xi$  and  $\mu_W$ ; the price to pay is that this difference will be expressed in terms of total variation distance rather than KL-divergence.

**THEOREM A.5.** *For any  $\delta \in (0, 1)$ , we have*

$$(A.13) \quad d_{\text{TV}}(\mu_\xi, \mu_W) \leq 2^{(1-\delta)/(1+\delta)} (M_\delta)^{1/(1+\delta)};$$

$$(A.14) \quad d_{\text{TV}}(\mu_\xi, \mu_W) \leq 1 - \frac{1}{6} \min \left( \frac{1}{8}, e^{-(2^{2-\delta} M_\delta)^{1/\delta}} \right).$$

Further, for any measurable set  $A \subset \mathcal{C}([0, \infty), \mathbb{R}^d)$  and any  $N > 1$

$$(A.15) \quad \mu_W(A) \geq \frac{1}{N} \left( \mu_\xi(A) - \frac{2^{1-\delta} M_\delta}{(\log N)^\delta} - \frac{\log 2}{\log N} \right).$$

**PROOF.** Fix  $\delta \in (0, 1)$ . As before, we can assume that condition (A.12) is satisfied; otherwise, all the bounds of the theorem hold trivially.

We begin with the proof of (A.13). Let  $\tau_n, \xi^n$  and  $\mu_{\xi^n}, n \in \mathbb{Z}_+$  be the same as in the proof of Theorem A.2. By Theorem A.2, we have for  $n \in \mathbb{Z}_+$

$$(A.16) \quad \begin{aligned} D_{\text{KL}}(\mu_{\xi^n} \parallel \mu_W) &\leq \frac{1}{2} \mathbb{E} \int_0^{\tau_n} |\beta_t|^2 dt = \frac{1}{2} \mathbb{E} \min \left( \int_0^\infty |\beta_t|^2 dt, n \right) \\ &\leq \frac{1}{2} n^{1-\delta} M_\delta. \end{aligned}$$

In addition,

$$(A.17) \quad d_{\text{TV}}(\mu_\xi, \mu_{\xi^n}) \leq \mathbb{P}(\tau_n < \infty) = \mathbb{P} \left( \int_0^\infty |\beta_t|^2 dt \geq n \right) \leq n^{-\delta} M_\delta.$$

Then it follows by (A.1) that

$$d_{\text{TV}}(\mu_\xi, \mu_W) \leq d_{\text{TV}}(\mu_\xi, \mu_{\xi^n}) + d_{\text{TV}}(\mu_{\xi^n}, \mu_W) \leq n^{-\delta} M_\delta + \frac{1}{2} n^{1/2-\delta/2} \sqrt{M_\delta}.$$

Taking  $n = (4M_\delta)^{1/(1+\delta)}$ , we obtain (A.13).

To establish (A.15), we apply inequality (A.4) to the measures  $\mu_W$  and  $\mu_{\xi^n}, n > 1$ . For any  $N > 1$  and any measurable set  $A$ , we get

$$(A.18) \quad \begin{aligned} \mu_W(A) &\geq \frac{1}{N} \mu_{\xi^n}(A) - \frac{D_{\text{KL}}(\mu_{\xi^n} \parallel \mu_W) + \log 2}{N \log N} \\ &\geq \frac{1}{N} \mu_{\xi^n}(A) - \frac{n^{1-\delta} M_\delta + 2 \log 2}{2N \log N}, \end{aligned}$$

where we have also used (A.16). Further, inequality (A.17) implies

$$\mu_{\xi^n}(A) \geq \mu_{\xi}(A) - d_{\text{TV}}(\mu_{\xi}, \mu_{\xi^n}) \geq \mu_{\xi}(A) - n^{-\delta} M_{\delta}.$$

Combining this with (A.18), we obtain

$$\mu_W(A) \geq \frac{1}{N} \left( \mu_{\xi}(A) - \frac{n^{1-\delta} M_{\delta}}{2 \log N} - n^{-\delta} M_{\delta} - \frac{\log 2}{\log N} \right).$$

By choosing  $n := 2 \log N$ , we obtain (A.15).

Finally, to derive (A.14) we apply (A.15) with  $\log N := \max(3 \log 2, (2^{2-\delta} M_{\delta})^{1/\delta})$ . We get

$$\mu_W(A) \geq \frac{1}{N} \left( \mu_{\xi}(A) - \frac{5}{6} \right).$$

This yields

$$\begin{aligned} d_{\text{TV}}(\mu_{\xi}, \mu_W) &= \sup_A (\mu_{\xi}(A) - \mu_W(A)) \leq \frac{5}{6N} + \sup_A \left( 1 - \frac{1}{N} \right) \mu_{\xi}(A) \\ &\leq 1 - \frac{1}{6N} \leq 1 - \frac{1}{6} \min \left( \frac{1}{8}, e^{-(2^{2-\delta} M_{\delta})^{1/\delta}} \right). \end{aligned}$$

This completes the proof of (A.14).  $\square$

**Acknowledgments.** The authors are grateful to Jonathan Mattingly for helpful discussions, to Mikhail Lifshits for fruitful comments and to Alexander Shaposhnikov for bringing the article [41] to our attention. The authors also would like to thank the anonymous referee for the valuable remarks and suggestions which helped to improve the quality of the paper. Part of the work on the project has been done during the visit of OB to the Fields Institute (Toronto, Canada) in October 2017. OB is very grateful to the Fields institute and especially to Bryan Eelhart for their support, hospitality and incredible coffee breaks. The work on the project has been finished during the visit of AK to the Technical University of Dresden (Germany); AK is very grateful to the Technical University of Dresden and especially to René Schilling for their support and hospitality.

The first author was supported in part by DFG Research Unit FOR 2402 and ISF-UGC grant No. 1131/14.

The second author was supported in part by DFG Research Unit FOR 2402.

## REFERENCES

- [1] ABOURASHCHI, N. and VERETENNIKOV, A. YU. (2010). On stochastic averaging and mixing. *Theory Stoch. Process.* **16** 111–129. MR2779833
- [2] BAO, J., WANG, F.-Y. and CHENGGUI, Y. (2018). Ergodicity for neutral type SDEs with infinite length of memory. arXiv:1805.03431.
- [3] BLÖMKER, D., LAW, K., STUART, A. M. and ZYGALAKIS, K. C. (2013). Accuracy and stability of the continuous-time 3DVAR filter for the Navier–Stokes equation. *Nonlinearity* **26** 2193–2219. MR3078113 <https://doi.org/10.1088/0951-7715/26/8/2193>
- [4] BODNARCHUK, S. V. and KULIK, A. M. (2009). Conditions for existence and smoothness of the distribution density of an Ornstein–Uhlenbeck process with Lévy noise. *Theory Probab. Math. Statist.* **79** 23–38. MR2494533 <https://doi.org/10.1090/S0094-9000-09-00778-9>
- [5] BORODIN, A. N. and SALMINEN, P. (2015). *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. Springer, Basel. MR1477407 <https://doi.org/10.1007/978-3-0348-7652-0>
- [6] BUTKOVSKY, O. (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.* **24** 526–552. MR3178490 <https://doi.org/10.1214/13-AAP922>
- [7] BUTKOVSKY, O. and SCHEUTZOW, M. (2017). Invariant measures for stochastic functional differential equations. *Electron. J. Probab.* **22** Paper No. 98, 1–23. MR3724566 <https://doi.org/10.1214/17-EJP122>

- [8] CONSTANTIN, P. and FOIAS, C. (1988). *Navier–Stokes Equations. Chicago Lectures in Mathematics*. Univ. Chicago Press, Chicago, IL. [MR0972259](#)
- [9] CONSTANTIN, P., GLATT-HOLTZ, N. and VICOL, V. (2014). Unique ergodicity for fractionally dissipated, stochastically forced 2D Euler equations. *Comm. Math. Phys.* **330** 819–857. [MR3223489](#) <https://doi.org/10.1007/s00220-014-2003-3>
- [10] DA PRATO, G. and ZABCZYK, J. (2014). *Stochastic Equations in Infinite Dimensions*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR3236753](#) <https://doi.org/10.1017/CBO9781107295513>
- [11] DALEY, R. (1991). *Atmospheric Data Analysis. Cambridge Atmospheric and Space Science Series*. Cambridge Univ. Press.
- [12] DURMUS, A., FORT, G. and MOULINES, É. (2016). Subgeometric rates of convergence in Wasserstein distance for Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1799–1822. [MR3573296](#) <https://doi.org/10.1214/15-AIHP699>
- [13] ES-SARHIR, A., VON RENESSE, M.-K. and SCHEUTZOW, M. (2009). Harnack inequality for functional SDEs with bounded memory. *Electron. Commun. Probab.* **14** 560–565. [MR2570679](#) <https://doi.org/10.1214/ECP.v14-1513>
- [14] FOIAŞ, C. and PRODI, G. (1967). Sur le comportement global des solutions non-stationnaires des équations de Navier–Stokes en dimension 2. *Rend. Semin. Mat. Univ. Padova* **39** 1–34. [MR0223716](#)
- [15] FÖLDES, J., GLATT-HOLTZ, N., RICHARDS, G. and THOMANN, E. (2015). Ergodic and mixing properties of the Boussinesq equations with a degenerate random forcing. *J. Funct. Anal.* **269** 2427–2504. [MR3390008](#) <https://doi.org/10.1016/j.jfa.2015.05.014>
- [16] FÖLDES, J., GLATT-HOLTZ, N. E., RICHARDS, G. and WHITEHEAD, J. P. (2016). Ergodicity in randomly forced Rayleigh–Bénard convection. *Nonlinearity* **29** 3309–3345. [MR3567091](#) <https://doi.org/10.1088/0951-7715/29/11/3309>
- [17] GLATT-HOLTZ, N., KUKAVICA, I., VICOL, V. and ZIANE, M. (2014). Existence and regularity of invariant measures for the three dimensional stochastic primitive equations. *J. Math. Phys.* **55** 051504, 1–34. [MR3390614](#) <https://doi.org/10.1063/1.4875104>
- [18] GLATT-HOLTZ, N., MATTINGLY, J. C. and RICHARDS, G. (2017). On unique ergodicity in nonlinear stochastic partial differential equations. *J. Stat. Phys.* **166** 618–649. [MR3607584](#) <https://doi.org/10.1007/s10955-016-1605-x>
- [19] HAIRER, M. (2002). Exponential mixing properties of stochastic PDEs through asymptotic coupling. *Probab. Theory Related Fields* **124** 345–380. [MR1939651](#) <https://doi.org/10.1007/s004400200216>
- [20] HAIRER, M. (2006). Ergodic properties of Markov processes. Lecture notes, Univ. Warwick. Available at <http://www.hairer.org/notes/Markov.pdf>.
- [21] HAIRER, M. and MATTINGLY, J. C. (2008). Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations. *Ann. Probab.* **36** 2050–2091. [MR2478676](#) <https://doi.org/10.1214/08-AOP392>
- [22] HAIRER, M. and MATTINGLY, J. C. (2011). A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. J. Probab.* **16** 658–738. [MR2786645](#) <https://doi.org/10.1214/EJP.v16-875>
- [23] HAIRER, M., MATTINGLY, J. C. and SCHEUTZOW, M. (2011). Asymptotic coupling and a general form of Harris’ theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149** 223–259. [MR2773030](#) <https://doi.org/10.1007/s00440-009-0250-6>
- [24] HOKE, J. and ANTHES, R. A. (1976). The initialization of numerical models by a dynamic initialization technique. *Mon. Weather Rev.* **104** 1551–1556.
- [25] KUKSIN, S. and SHIRIKYAN, A. (2012). *Mathematics of Two-Dimensional Turbulence. Cambridge Tracts in Mathematics* **194**. Cambridge Univ. Press, Cambridge. [MR3443633](#) <https://doi.org/10.1017/CBO9781139137119>
- [26] KULIK, A. (2018). *Ergodic Behavior of Markov Processes: With Applications to Limit Theorems. De Gruyter Studies in Mathematics* **67**. de Gruyter, Berlin. [MR3791835](#)
- [27] KULIK, A. and SCHEUTZOW, M. (2018). Generalized couplings and convergence of transition probabilities. *Probab. Theory Related Fields* **171** 333–376. [MR3800835](#) <https://doi.org/10.1007/s00440-017-0779-8>
- [28] LADYŽENSKAJA, O. A. (1959). Solution “in the large” of the nonstationary boundary value problem for the Navier–Stokes system with two space variables. *Comm. Pure Appl. Math.* **12** 427–433. [MR0108962](#) <https://doi.org/10.1002/cpa.3160120303>
- [29] LIPTSER, R. S. and SHIRYAYEV, A. N. (2013). *Statistics of Random Processes. I*. Springer, Berlin. [MR474486](#)
- [30] MATTINGLY, J. C. (1999). Ergodicity of 2D Navier–Stokes equations with random forcing and large viscosity. *Comm. Math. Phys.* **206** 273–288. [MR1722141](#) <https://doi.org/10.1007/s002200050706>
- [31] MATTINGLY, J. C. (2002). Exponential convergence for the stochastically forced Navier–Stokes equations and other partially dissipative dynamics. *Comm. Math. Phys.* **230** 421–462. [MR1937652](#) <https://doi.org/10.1007/s00220-002-0688-1>

- [32] MATTINGLY, J. C. (2003). On recent progress for the stochastic Navier Stokes equations. In *Journées “Équations aux Dérivées Partielles”* 1–52. MR2050597
- [33] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. With a prologue by Peter W. Glynn. MR2509253 <https://doi.org/10.1017/CBO9780511626630>
- [34] NUMMELIN, E. (2004). *General Irreducible Markov Chains and Nonnegative Operators*, 2nd ed. Cambridge Univ. Press, Cambridge. MR776608 <https://doi.org/10.1017/CBO9780511526237>
- [35] OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864. MR2484937 <https://doi.org/10.1016/j.jfa.2008.11.001>
- [36] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, Berlin. MR1725357
- [37] SCHEUTZOW, M. (1984). Qualitative behaviour of stochastic delay equations with a bounded memory. *Stochastics* **12** 41–80. MR0738934 <https://doi.org/10.1080/17442508408833294>
- [38] SCHEUTZOW, M. (2005). Exponential growth rates for stochastic delay differential equations. *Stoch. Dyn.* **5** 163–174. MR2147280 <https://doi.org/10.1142/S0219493705001468>
- [39] TEMAM, R. (1997). *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, 2nd ed. *Applied Mathematical Sciences* **68**. Springer, New York. MR1441312 <https://doi.org/10.1007/978-1-4612-0645-3>
- [40] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 <https://doi.org/10.1007/b13794>
- [41] ÜSTÜNEL, A. S. (2009). Entropy, invertibility and variational calculus of adapted shifts on Wiener space. *J. Funct. Anal.* **257** 3655–3689. MR2572265 <https://doi.org/10.1016/j.jfa.2009.03.015>
- [42] VILLANI, C. (2009). *Optimal Transport: Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin. MR2459454 <https://doi.org/10.1007/978-3-540-71050-9>
- [43] VON RENESSE, M.-K. and SCHEUTZOW, M. (2010). Existence and uniqueness of solutions of stochastic functional differential equations. *Random Oper. Stoch. Equ.* **18** 267–284. MR2718125 <https://doi.org/10.1515/ROSE.2010.015>
- [44] WANG, F.-Y. (2011). Harnack inequality for SDE with multiplicative noise and extension to Neumann semi-group on nonconvex manifolds. *Ann. Probab.* **39** 1449–1467. MR2857246 <https://doi.org/10.1214/10-AOP600>