

# STRONG AVERAGING PRINCIPLE FOR SLOW-FAST STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS WITH LOCALLY MONOTONE COEFFICIENTS

WEI LIU, MICHAEL RÖCKNER, XIAOBIN SUN, AND YINGCHAO XIE

ABSTRACT. This paper is devoted to proving the strong averaging principle for slow-fast stochastic partial differential equations with locally monotone coefficients, where the slow component is a stochastic partial differential equations with locally monotone coefficients and the fast component is a stochastic partial differential equations with strongly monotone coefficients. The result is applicable to a large class of examples, such as the stochastic porous medium equation, the stochastic  $p$ -Laplace equation, the stochastic Burgers type equation and the stochastic 2D Navier-Stokes equation, which are the nonlinear stochastic partial differential equations. The main techniques are based on time discretization and the variational approach to stochastic partial differential equations.

## 1. INTRODUCTION

For  $i = 1, 2$ , let  $(H_i, \|\cdot\|_{H_i})$  be a separable Hilbert spaces with inner product  $\langle \cdot, \cdot \rangle_{H_i}$  and  $H_i^*$  its dual. Let  $(V_i, \|\cdot\|_{V_i})$  be a reflexive Banach space, such that  $V_i \subseteq H_i$  continuously and densely. Then for its dual space  $V_i^*$  it follows that  $H_i^* \subseteq V_i^*$  continuously and densely. Identifying  $H_i$  and  $H_i^*$  via the Riesz isomorphism we have that

$$V_i \subseteq H_i \equiv H_i^* \subseteq V_i^*$$

is a Gelfand triple. Let  ${}_{V_i^*}\langle \cdot, \cdot \rangle_{V_i}$  be the dualization between  $V_i^*$  and  $V_i$ . Then it follows that

$${}_{V_i^*}\langle z_i, v_i \rangle_{V_i} = \langle z_i, v_i \rangle_{H_i}, \quad \text{for all } z_i \in H_i, v_i \in V_i.$$

For  $i = 1, 2$ , let  $\{W_t^i\}_{t \geq 0}$  be a cylindrical  $\mathcal{F}_t$ -Wiener process in a separable Hilbert space  $(U_i, \|\cdot\|_{U_i})$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with natural filtration  $\mathcal{F}_t$ . Let  $L_2(U_i, H_i)$  be the space of Hilbert-Schmidt operator from  $U_i \rightarrow H_i$ . The norm on  $L_2(U_i, H_i)$  is defined by

$$\|S\|_{L_2(U_i, H_i)}^2 := \sum_{k \in \mathbb{N}} \|S e_{i,k}\|_{H_i}^2, \quad S \in L_2(U_i, H_i),$$

where  $\{e_{i,k}\}_{k \in \mathbb{N}}$  is an orthonormal basis of  $U_i$ . We also assume the processes  $\{W_t^1\}_{t \geq 0}$  and  $\{W_t^2\}_{t \geq 0}$  are independent.

In this paper, we consider the following abstract stochastic partial differential equations (SPDEs)

$$\begin{cases} dX_t^\varepsilon = [A(X_t^\varepsilon) + F(X_t^\varepsilon, Y_t^\varepsilon)] dt + G_1(X_t^\varepsilon) dW_t^1, \\ dY_t^\varepsilon = \frac{1}{\varepsilon} B(X_t^\varepsilon, Y_t^\varepsilon) dt + \frac{1}{\sqrt{\varepsilon}} G_2(X_t^\varepsilon, Y_t^\varepsilon) dW_t^2, \\ X_0^\varepsilon = x \in H_1, Y_0^\varepsilon = y \in H_2, \end{cases} \quad (1.1)$$

---

*Date:* November 26, 2022.

*2000 Mathematics Subject Classification.* 60H15; 35Q30; 70K70.

*Key words and phrases.* Local monotonicity; Averaging principle; SPDE; Strong convergence; Slow-fast.

where  $\varepsilon > 0$  is a small parameter describing the ratio of the time scale between the slow component  $X_t^\varepsilon$  and the fast component  $Y_t^\varepsilon$ , and the coefficients

$$A : V_1 \rightarrow V_1^*; \quad F : H_1 \times H_2 \rightarrow H_1; \quad G_1 : V_1 \rightarrow L_2(U_1; H_1);$$

and

$$B : H_1 \times V_2 \rightarrow V_2^*; \quad G_2 : H_1 \times V_2 \rightarrow L_2(U_2; H_2)$$

are measurable.

The averaging principle has a long and rich history in multiscale models, which has wide applications in material sciences, chemistry, fluid dynamics, biology, ecology and climate dynamics, see, e.g., [1, 10, 17, 22] and the references therein. Usually, a multiscale model can be described through coupled equations, which correspond to the "slow" and "fast" component, respectively. The averaging principle is essential to describe the asymptotic behavior of the slow component, i.e., the slow component will convergence to the so-called averaged equation. Bogoliubov and Mitropolsky [2] first studied the averaging principle for deterministic systems, which then was extended to stochastic differential equations by Khasminskii [18].

Since the averaging principle for a general class of stochastic reaction-diffusion systems with two time-scales were investigated by Cerrai and Freidlin in [6], the averaging principle for slow-fast SPDEs has initiated further studies in the past decade, including other types of SPDEs, various ways of convergence and rates of convergence. For instance, Bréhier obtained the strong and weak orders in averaging for stochastic evolution equation of parabolic type with slow and fast time scales in [3]. Fu, Wan and Liu proved the strong averaging principle for stochastic hyperbolic-parabolic equations with slow and fast time-scales in [13]. Cerrai and Lunardi studied the averaging principle for nonautonomous slow-fast systems of stochastic reaction-diffusion equations in [7]. For some further results on this topic, we refer to [4, 11, 12, 20, 24, 25] and the references therein.

However, the references we mentioned above always assume that the coefficients satisfy Lipschitz conditions, and there are few results on the average principle for SPDEs with nonlinear terms. For example, stochastic reaction-diffusion equations with polynomial coefficients [5], stochastic Burgers equation [9], stochastic two dimensional Navier-Stokes equations [19], stochastic Kuramoto-Sivashinsky equation [14], stochastic Schrödinger equation [15] and stochastic Klein-Gordon equation [16]. But all these papers consider semilinear SPDEs (i.e., for operators  $A = A_1 + A_2$  with  $A_1$  a linear operator and  $A_2$  a nonlinear perturbation), and use the mild solution approach to SPDEs exploiting the smoothing properties of the  $C_0$ - semigroup  $e^{A_1 t}$  generated by the linear operator  $A_1$  in an essential way. To the best of our knowledge, the case of the operator  $A$  has no linear part hasn't been studied yet, such as the porous medium operator and the  $p$ -Laplace operator.

Hence, the main purpose of this paper is to prove the strong averaging principle for slow-fast SPDEs within the (generalized) variational framework, i.e., locally monotone and strongly monotone coefficients for the slow and fast equations respectively. Our result covers a large class of examples (see [21, Sections 4 and 5]), especially for the case that the slow equation is a quasilinear SPDEs, such as the stochastic porous medium equation or the stochastic  $p$ -Laplace equation. Our result is also applicable to the stochastic Burgers type equation and stochastic two dimensional Navier-Stokes equation, whose coefficients only satisfy the local monotonicity conditions.

The main difficulty here is how to avoid applying the techniques which only work in the case of the mild solution approach, and use the techniques from the variational approach. More precisely, we will use the variational approach to estimate the integral of the time increment of  $X_t^\varepsilon$  instead of studying the Hölder continuity of time, which is strong enough for our purpose. We will also use the variational approach to obtain some apriori estimates of the solution, which are crucial to construct a proper stopping time to deal with the nonlinear terms.

The rest of the paper is organized as follows. In Section 2, under some suitable assumptions, we formulate our main result. Section 3 is devoted to proving the main result. In Section 4, we will give some examples to illustrate the wide applicability of our result. In the Appendix, we give the detailed proof of the existence and uniqueness of solutions to system (1.1).

Throughout the paper,  $C$ ,  $C_T$  and  $C_{p,T}$  denote positive constants which may change from line to line, where  $C_T$  and  $C_{p,T}$  are used to emphasize that the constants depend on  $T$  and  $p, T$  respectively.

## 2. MAIN RESULT

For the coefficients of the slow equation, we suppose that there exist constants  $\alpha \in (1, \infty)$ ,  $\beta \in [0, \infty)$ ,  $\theta \in (0, \infty)$  and  $C > 0$  such that the following conditions hold for all  $u, v, w \in V_1$ ,  $u_1, v_1 \in H_1$  and  $u_2, v_2 \in H_2$ :

**A1.** (*Hemicontinuity*) *The map  $\lambda \rightarrow v_1^* \langle A(u + \lambda v), w \rangle_{V_1}$  is continuous on  $\mathbb{R}$ .*

**A2.** (*Local monotonicity*)

$$2v_1^* \langle A(u) - A(v), u - v \rangle_{V_1} + \|G_1(u) - G_1(v)\|_{L_2(U_1, H_1)}^2 \leq \rho(v) \|u - v\|_{H_1}^2.$$

where  $\rho : V_1 \rightarrow [0, \infty)$  is a measurable hemicontinuous function satisfying

$$|\rho(v)| \leq C \left[ (1 + \|v\|_{V_1}^\alpha)(1 + \|v\|_{H_1}^\beta) \right].$$

Furthermore,

$$\|F(u_1, u_2) - F(v_1, v_2)\|_{H_1} \leq C(\|u_1 - v_1\|_{H_1} + \|u_2 - v_2\|_{H_2}). \quad (2.1)$$

**A3.** (*Coercivity*)

$$v_1^* \langle A(v), v \rangle_{V_1} \leq C\|v\|_{H_1}^2 - \theta\|v\|_{V_1}^\alpha + C.$$

**A4.** (*Growth*)

$$\|A(v)\|_{V_1^*}^{\frac{\alpha}{\alpha-1}} \leq C(1 + \|v\|_{V_1}^\alpha)(1 + \|v\|_{H_1}^\beta)$$

and

$$\|G_1(v)\|_{L_2(U_1, H_1)} \leq C(1 + \|v\|_{H_1}).$$

For the coefficients of the fast equation, we suppose that there exist constants  $\kappa \in (1, \infty)$ ,  $\gamma, \eta \in (0, \infty)$ ,  $\zeta \in (0, 1)$  and  $C > 0$  such that the following conditions hold for all  $u, v, w \in V_2$ ,  $u_1, v_1 \in H_1$ :

**B1.** (*Hemicontinuity*) *The map  $\lambda \rightarrow v_2^* \langle B(u_1 + \lambda v_1, u + \lambda v), w \rangle_{V_2}$  is continuous on  $\mathbb{R}$ .*

**B2.** (*Strong monotonicity*)

$$\begin{aligned} & 2_{V_2^*} \langle B(u_1, u) - B(v_1, v), u - v \rangle_{V_2} + \|G_2(u_1, u) - G_2(v_1, v)\|_{L_2(U_2, H_2)}^2 \\ & \leq -\gamma \|u - v\|_{H_2}^2 + C \|u_1 - v_1\|_{H_1}^2. \end{aligned} \quad (2.2)$$

**B3.** (*Coercivity*)

$$_{V_2^*} \langle B(u_1, v), v \rangle_{V_2} \leq C \|v\|_{H_2}^2 - \eta \|v\|_{V_2}^\kappa + C(1 + \|u_1\|_{H_1}^2).$$

**B4.** (*Growth*)

$$\|B(u_1, v)\|_{V_2^*} \leq C \left( 1 + \|v\|_{V_2}^{\kappa-1} + \|u_1\|_{H_1}^{\frac{2(\kappa-1)}{\kappa}} \right)$$

and

$$\|G_2(u_1, v)\|_{L_2(U_2, H_2)} \leq C(1 + \|u_1\|_{H_1} + \|v\|_{H_2}^\zeta). \quad (2.3)$$

**Remark 2.1.** We here give some comments for the assumptions above.

- Condition (2.2) is also called the dissipativity condition, which guarantees that there exists a unique invariant measure for the frozen equation and the exponential ergodicity holds.
- The  $\zeta \in (0, 1)$  in condition (2.3) is used to prove the  $p$ -th moments of the solution  $(X_t^\varepsilon, Y_t^\varepsilon)$  are finite, when  $p$  is large enough, which could be removed if we assume the Lipschitz constant of  $G_2$  is sufficiently small.

Now, we recall the definition of a variational solution in [21].

**Definition 2.2.** For any given  $\varepsilon > 0$ , a continuous  $H_1 \times H_2$ -valued  $\mathcal{F}_t$ -adapted process  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \in [0, T]}$  is called a solution of system (1.1), if for its  $dt \otimes \mathbb{P}$ -equivalence class  $(\tilde{X}^\varepsilon, \tilde{Y}^\varepsilon)$  we have  $\tilde{X}^\varepsilon \in L^\alpha([0, T] \times \Omega, dt \otimes \mathbb{P}; V_1) \cap L^2([0, T] \times \Omega, dt \otimes \mathbb{P}; H_1)$  with  $\alpha$  as in A3,  $\tilde{Y}^\varepsilon \in L^\kappa([0, T] \times \Omega, dt \otimes \mathbb{P}; V_2) \cap L^2([0, T] \times \Omega, dt \otimes \mathbb{P}; H_2)$  with  $\kappa$  as in B3 and  $\mathbb{P}$ -a.s.

$$\begin{cases} X_t^\varepsilon = X_0^\varepsilon + \int_0^t A(\tilde{X}_s^\varepsilon) ds + \int_0^t F(X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t G(\tilde{X}_s^\varepsilon) dW_s^1, \\ Y_t^\varepsilon = Y_0^\varepsilon + \frac{1}{\varepsilon} \int_0^t B(X_s^\varepsilon, \tilde{Y}_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^t G_2(X_s^\varepsilon, \tilde{Y}_s^\varepsilon) dW_s^2, \end{cases} \quad (2.4)$$

where  $(\tilde{X}^\varepsilon, \tilde{Y}^\varepsilon)$  is any  $V_1 \times V_2$ -valued progressively measurable  $dt \otimes \mathbb{P}$ -version of  $(\tilde{X}^\varepsilon, \tilde{Y}^\varepsilon)$ .

Using the variational approach in infinite dimensional space, we have the following well-posedness result, whose proof will be presented in the Appendix.

**Theorem 2.3.** Assume the conditions A1-A4, B1-B4 hold. Then for any  $\varepsilon > 0$  and initial values  $(x, y) \in H_1 \times H_2$ , the system (1.1) has a unique solution  $(X^\varepsilon, Y^\varepsilon)$ .

The following is the main result of this work.

**Theorem 2.4.** Assume the conditions A1-A4, B1-B4 hold. Then for any initial values  $(x, y) \in H_1 \times H_2$ ,  $p \geq 1$  and  $T > 0$ , we have

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p} \right) = 0, \quad (2.5)$$

where  $\bar{X}_t$  is the solution of the corresponding averaged equation:

$$\begin{cases} d\bar{X}_t = A(\bar{X}_t)dt + \bar{F}(\bar{X}_t)dt + G_1(\bar{X}_t)dW_t^1, \\ \bar{X}_0 = x, \end{cases} \quad (2.6)$$

with the average  $\bar{F}(x) = \int_{H_2} F(x, y)\mu^x(dy)$ .  $\mu^x$  is the unique invariant measure of the transition semigroup of the frozen equation

$$\begin{cases} dY_t = B(x, Y_t)dt + G_2(x, Y_t)d\bar{W}_t^2, \\ Y_0 = y, \end{cases}$$

where  $\{\bar{W}_t^2\}_{t \geq 0}$  is a  $\tilde{\mathcal{F}}_t$ -cylindrical Wiener process in a separable Hilbert space  $U_2$  on another probability space, with natural filtration  $\tilde{\mathcal{F}}_t$ .

**Remark 2.5.** The advantage of using the variational approach is that it can cover some non-linear SPDEs for slow component, such as stochastic power law fluids, and some quasilinear SPDEs for slow component, such as the stochastic porous medium equation and the stochastic  $p$ -Laplace equation, which can not be handled by the mild solution approach and thus have not been studied yet. Furthermore, our result also generalizes some known results of the cases that the slow component is a semilinear stochastic partial differential equation, such as the stochastic Burgers equation (see [9]) and stochastic two dimensional Navier-Stokes equation (see [19]). Besides some known results, our result can also be applied to many other unstudied hydrodynamical models in [8], such as the stochastic magneto-hydrodynamic equations, the stochastic Boussinesq model for the Bénard convection, the stochastic 2D magnetic Bénard problem, the stochastic 3D Leray- $\alpha$  model and some stochastic shell models of turbulence.

### 3. PROOF OF THE MAIN RESULT

This section is devoted to proving Theorem 2.4. The proof consists of the following four subsections: In Subsection 3.1, we give some apriori estimates for the solution  $(X_t^\varepsilon, Y_t^\varepsilon)$ . Using the apriori estimates, we get an estimate for the time increments for  $X_t^\varepsilon$ , which plays an important role in the proof of the main result. In Subsection 3.2, we will use the technique of time discretization to construct an auxiliary process  $\hat{Y}_t^\varepsilon$  and give an estimate of the difference process  $Y_t^\varepsilon - \hat{Y}_t^\varepsilon$ . In Subsection 3.3, by constructing a stopping time  $\tau_R$ , we prove that  $X_t^\varepsilon$  strongly converges to  $\bar{X}_t$  for  $t < \tau_R$ . Finally, the apriori estimates for the solution control the difference  $X_t^\varepsilon - \bar{X}_t$  after the stopping time. Note that we always assume conditions **A1-A4** and **B1-B4** hold and from now on we fix an initial value  $(x, y) \in H_1 \times H_2$  in this section.

**3.1. Some apriori estimates of  $(X_t^\varepsilon, Y_t^\varepsilon)$ .** At first, we prove uniform bounds with respect to  $\varepsilon \in (0, 1)$  for the moments of the solution  $(X_t^\varepsilon, Y_t^\varepsilon)$  to the system (1.1).

**Lemma 3.1.** *For any  $T > 0$  and  $p \geq 1$ , there exists a constant  $C_{p,T} > 0$  such that*

$$\sup_{\varepsilon \in (0,1)} \mathbb{E} \left( \sup_{t \in [0,T]} \|X_t^\varepsilon\|_{H_1}^{2p} \right) + \sup_{\varepsilon \in (0,1)} \mathbb{E} \left( \int_0^T \|X_t^\varepsilon\|_{H_1}^{2p-2} \|\tilde{X}_t^\varepsilon\|_{V_1}^\alpha dt \right) \leq C_{p,T} \left( 1 + \|x\|_{H_1}^{2p} + \|y\|_{H_2}^{2p} \right) \quad (3.1)$$

and

$$\sup_{\varepsilon \in (0,1)} \sup_{t \in [0,T]} \mathbb{E} \|Y_t^\varepsilon\|_{H_2}^{2p} \leq C_{p,T} \left( 1 + \|x\|_{H_1}^{2p} + \|y\|_{H_2}^{2p} \right). \quad (3.2)$$

*Proof.* Applying Itô's formula (see e.g. [21, Theorem 6.1.1]), we have

$$\begin{aligned} \|Y_t^\varepsilon\|_{H_2}^2 &= \|y\|_{H_2}^2 + \frac{2}{\varepsilon} \int_0^t v_2^* \langle B(X_s^\varepsilon, \tilde{Y}_s^\varepsilon), \tilde{Y}_s^\varepsilon \rangle_{V_2} ds \\ &\quad + \frac{1}{\varepsilon} \int_0^t \|G_2(X_s^\varepsilon, \tilde{Y}_s^\varepsilon)\|_{L_2(U_2, H_2)}^2 ds + \frac{2}{\sqrt{\varepsilon}} \int_0^t \langle G_2(X_s^\varepsilon, \tilde{Y}_s^\varepsilon) dW_s^2, Y_s^\varepsilon \rangle_{H_2}. \end{aligned}$$

Then applying Itô's formula for  $f(z) = (z)^p$  with  $z_t = \|Y_t^\varepsilon\|_{H_2}^2$ , and taking expectation on both sides, we obtain

$$\begin{aligned} \mathbb{E}\|Y_t^\varepsilon\|_{H_2}^{2p} &= \|y\|_{H_2}^{2p} + \frac{2p}{\varepsilon} \mathbb{E} \left[ \int_0^t \|Y_s^\varepsilon\|_{H_2}^{2p-2} v_2^* \langle B(X_s^\varepsilon, \tilde{Y}_s^\varepsilon), \tilde{Y}_s^\varepsilon \rangle_{V_2} ds \right] \\ &\quad + \frac{p}{\varepsilon} \mathbb{E} \left[ \int_0^t \|Y_s^\varepsilon\|_{H_2}^{2p-2} \|G_2(X_s^\varepsilon, \tilde{Y}_s^\varepsilon)\|_{L_2(U_2, H_2)}^2 ds \right] \\ &\quad + \frac{2p(p-1)}{\varepsilon} \mathbb{E} \left[ \int_0^t \|Y_s^\varepsilon\|_{H_2}^{2p-4} \|G_2(X_s^\varepsilon, \tilde{Y}_s^\varepsilon)^* Y_s^\varepsilon\|_{U_2}^2 ds \right] \end{aligned}$$

By conditions **B2-B4** and a similar argument in the proof of [21, Lemma 4.3.8], there exists a constant  $\hat{\gamma} \in (0, \gamma)$  such that for any  $u \in H_1, v \in V_2$ ,

$$2v_2^* \langle B(u, v), v \rangle_{V_2} \leq -\hat{\gamma} \|v\|_{H_2}^2 + C(1 + \|u\|_{H_1}^2). \quad (3.3)$$

Then applying Young's inequality and estimate (3.3), we get

$$\begin{aligned} \frac{d}{dt} \mathbb{E}\|Y_t^\varepsilon\|_{H_2}^{2p} &= \frac{2p}{\varepsilon} \mathbb{E} \left[ \|Y_t^\varepsilon\|_{H_2}^{2p-2} v_2^* \langle B(X_t^\varepsilon, \tilde{Y}_t^\varepsilon), \tilde{Y}_t^\varepsilon \rangle_{V_2} \right] + \frac{p}{\varepsilon} \mathbb{E} \left[ \|Y_t^\varepsilon\|_{H_2}^{2p-2} \|G_2(X_t^\varepsilon, \tilde{Y}_t^\varepsilon)\|_{L_2(U_2, H_2)}^2 \right] \\ &\quad + \frac{2p(p-1)}{\varepsilon} \mathbb{E} \left[ \|Y_t^\varepsilon\|_{H_2}^{2p-4} \|G_2(X_t^\varepsilon, \tilde{Y}_t^\varepsilon)^* Y_t^\varepsilon\|_{U_2}^2 \right] \\ &\leq \frac{2p}{\varepsilon} \mathbb{E} \left[ \|Y_t^\varepsilon\|_{H_2}^{2p-2} (-\hat{\gamma} \|Y_t^\varepsilon\|_{H_2}^2 + C \|X_t^\varepsilon\|_{H_2}^2 + C) \right] \\ &\quad + \frac{C_p}{\varepsilon} \mathbb{E} \left[ \|Y_t^\varepsilon\|_{H_2}^{2p-2} (1 + \|X_t^\varepsilon\|_{H_1}^2 + \|Y_t^\varepsilon\|_{H_2}^{2\zeta}) \right] \\ &\leq -\frac{C_p}{\varepsilon} \mathbb{E}\|Y_t^\varepsilon\|_{H_2}^{2p} + \frac{C_p}{\varepsilon} \mathbb{E}\|X_t^\varepsilon\|_{H_1}^{2p} + \frac{C_p}{\varepsilon}. \end{aligned}$$

Multiplication by the integrating factor  $e^{\frac{C_p t}{\varepsilon}}$  yields that

$$\frac{d}{dt} \left( e^{\frac{C_p t}{\varepsilon}} \mathbb{E}\|Y_t^\varepsilon\|_{H_2}^{2p} \right) \leq \frac{C_p}{\varepsilon} e^{\frac{C_p t}{\varepsilon}} \left( 1 + \mathbb{E}\|X_t^\varepsilon\|_{H_1}^{2p} \right).$$

Integrate this from 0 to  $t$ , we get

$$\mathbb{E}\|Y_t^\varepsilon\|_{H_2}^{2p} \leq \|y\|_{H_2}^{2p} e^{-\frac{C_p t}{\varepsilon}} + \frac{C_p}{\varepsilon} \int_0^t e^{-\frac{C_p}{\varepsilon}(t-s)} \left( 1 + \mathbb{E}\|X_s^\varepsilon\|_{H_1}^{2p} \right) ds. \quad (3.4)$$

On the other hand, applying Itô's formula, we also have

$$\begin{aligned} \|X_t^\varepsilon\|_{H_1}^{2p} &= \|x\|_{H_1}^{2p} + 2p \int_0^t \|X_s^\varepsilon\|_{H_1}^{2p-2} v_1^* \langle A(\tilde{X}_s^\varepsilon), \tilde{X}_s^\varepsilon \rangle_{V_1} ds + 2p \int_0^t \|X_s^\varepsilon\|_{H_1}^{2p-2} \langle F(X_s^\varepsilon, Y_s^\varepsilon), X_s^\varepsilon \rangle_{H_1} ds \\ &\quad + 2p \int_0^t \|X_s^\varepsilon\|_{H_1}^{2p-2} \langle X_s^\varepsilon, G_1(\tilde{X}_s^\varepsilon) dW_s^1 \rangle_{H_1} + p \int_0^t \|X_s^\varepsilon\|_{H_1}^{2p-2} \|G_1(\tilde{X}_s^\varepsilon)\|_{L_2(U_1, H_1)}^2 ds \\ &\quad + 2p(p-1) \int_0^t \|X_s^\varepsilon\|_{H_1}^{2p-4} \|G_1(\tilde{X}_s^\varepsilon)^* X_s^\varepsilon\|_{U_1}^2 ds. \end{aligned}$$

Note that

$$\left\{ \int_0^t \|X_s^\varepsilon\|_{H_1}^{2p-2} \langle F(X_s^\varepsilon, Y_s^\varepsilon), X_s^\varepsilon \rangle_{H_1} ds \right\}_{0 \leq t \leq T}$$

is a local martingale, then applying Burkholder-Davis-Gundy inequality (see e.g. [21, Theorem 6.1.2]), conditions **A2-A4** and (3.4), we get

$$\begin{aligned} & \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon\|_{H_1}^{2p} \right) + 2p\theta \mathbb{E} \left( \int_0^T \|X_t^\varepsilon\|_{H_1}^{2p-2} \|\tilde{X}_t^\varepsilon\|_{V_1}^\alpha dt \right) \\ & \leq C_P (\|x\|_{H_1}^{2p} + 1) + C_P \int_0^T \mathbb{E} \|X_t^\varepsilon\|_{H_1}^{2p} dt + C_P \int_0^T \mathbb{E} \|Y_t^\varepsilon\|_{H_2}^{2p} dt \\ & \leq C_P (\|x\|_{H_1}^{2p} + \|y\|_{H_2}^{2p} + 1) + C_P \int_0^T \mathbb{E} \|X_t^\varepsilon\|_{H_1}^{2p} dt + \frac{C_P}{\varepsilon} \int_0^T \int_0^t e^{-\frac{C_P}{\varepsilon}(t-s)} (1 + \mathbb{E} \|X_s^\varepsilon\|_{H_1}^{2p}) ds dt \\ & \leq C_P (\|x\|_{H_1}^{2p} + \|y\|_{H_2}^{2p} + 1) + C_P \int_0^T \mathbb{E} \|X_t^\varepsilon\|_{H_1}^{2p} dt. \end{aligned}$$

Hence, applying Gronwall's inequality (see e.g. [23, Exercise 5.17]), we obtain

$$\mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon\|_{H_1}^{2p} \right) + 2p \mathbb{E} \left( \int_0^T \|X_t^\varepsilon\|_{H_1}^{2p-2} \|\tilde{X}_t^\varepsilon\|_{V_1}^\alpha dt \right) \leq C_{p,T} (\|x\|_{H_1}^{2p} + \|y\|_{H_2}^{2p} + 1),$$

which also gives that

$$\sup_{t \in [0, T]} \mathbb{E} \|Y_t^\varepsilon\|_{H_2}^{2p} \leq C_{p,T} (1 + \|x\|_{H_1}^{2p} + \|y\|_{H_2}^{2p}).$$

The proof is complete.  $\square$

Because the method of time discretization is used in this paper, the following estimate about the integral of the time increment plays an important role in the proof of our main result, which has been proved in the case of the stochastic 2D Navier-Stokes equation in [19].

**Lemma 3.2.** *For any  $T > 0$ ,  $\varepsilon \in (0, 1)$  and  $\delta > 0$  small enough, there exist constants  $C_T, m > 0$  such that*

$$\mathbb{E} \left[ \int_0^T \|X_t^\varepsilon - X_{t(\delta)}^\varepsilon\|_{H_1}^2 dt \right] \leq C_T (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m) \delta^{1/2}, \quad (3.5)$$

where  $t(\delta) := [\frac{t}{\delta}]\delta$  and  $[s]$  denotes the largest integer which is smaller than  $s$ .

*Proof.* Note that

$$\begin{aligned} \mathbb{E} \left[ \int_0^T \|X_t^\varepsilon - X_{t(\delta)}^\varepsilon\|_{H_1}^2 dt \right] &= \mathbb{E} \left( \int_0^\delta \|X_t^\varepsilon - x\|_{H_1}^2 dt \right) + \mathbb{E} \left[ \int_\delta^T \|X_t^\varepsilon - X_{t(\delta)}^\varepsilon\|_{H_1}^2 dt \right] \\ &\leq C(1 + \|x\|_{H_1}^2 + \|y\|_{H_2}^2) \delta + 2 \mathbb{E} \left( \int_\delta^T \|X_t^\varepsilon - X_{t-\delta}^\varepsilon\|_{H_1}^2 dt \right) \\ &\quad + 2 \mathbb{E} \left( \int_\delta^T \|X_{t(\delta)}^\varepsilon - X_{t-\delta}^\varepsilon\|_{H_1}^2 dt \right). \end{aligned} \quad (3.6)$$

Then applying Itô's formula we have

$$\begin{aligned} \|X_t^\varepsilon - X_{t-\delta}^\varepsilon\|_{H_1}^2 &= 2 \int_{t-\delta}^t V_1^* \langle A(\tilde{X}_s^\varepsilon), \tilde{X}_s^\varepsilon - \tilde{X}_{t-\delta}^\varepsilon \rangle_{V_1} ds + 2 \int_{t-\delta}^t \langle F(X_s^\varepsilon, Y_s^\varepsilon), X_s^\varepsilon - X_{t-\delta}^\varepsilon \rangle_{H_1} ds \\ &\quad + \int_{t-\delta}^t \|G_1(\tilde{X}_s^\varepsilon)\|_{L_2(U_1, H_1)}^2 ds + 2 \int_{t-\delta}^t \langle X_s^\varepsilon - X_{t-\delta}^\varepsilon, G_1(\tilde{X}_s^\varepsilon) dW_s^1 \rangle_{H_1} \end{aligned}$$

$$:= I_1(t) + I_2(t) + I_3(t) + I_4(t). \quad (3.7)$$

For the term  $I_1(t)$ , by condition **A4** and applying Hölder's inequality, there exist constants  $m, C_T > 0$  such that

$$\begin{aligned} \mathbb{E} \left( \int_{\delta}^T |I_1(t)| dt \right) &\leq C \mathbb{E} \left( \int_{\delta}^T \int_{t-\delta}^t \|A(\tilde{X}_s^\varepsilon)\|_{V_1^*} \|\tilde{X}_s^\varepsilon - \tilde{X}_{t-\delta}^\varepsilon\|_{V_1} ds dt \right) \\ &\leq C \left[ \mathbb{E} \int_{\delta}^T \int_{t-\delta}^t \|A(\tilde{X}_s^\varepsilon)\|_{V_1^*}^{\frac{\alpha}{\alpha-1}} ds dt \right]^{\frac{\alpha-1}{\alpha}} \left[ \mathbb{E} \int_{\delta}^T \int_{t-\delta}^t \|\tilde{X}_s^\varepsilon - \tilde{X}_{t-\delta}^\varepsilon\|_{V_1}^{\alpha} ds dt \right]^{\frac{1}{\alpha}} \\ &\leq C \left[ \delta \mathbb{E} \int_0^T (1 + \|\tilde{X}_s^\varepsilon\|_{V_1}^{\alpha})(1 + \|X_s^\varepsilon\|_{H_1}^{\beta}) ds \right]^{\frac{\alpha-1}{\alpha}} \cdot \left[ \delta \mathbb{E} \int_0^T \|\tilde{X}_s^\varepsilon\|_{V_1}^{\alpha} ds \right]^{\frac{1}{\alpha}} \\ &\leq C_T (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m) \delta, \end{aligned} \quad (3.8)$$

where we applied Fubini's theorem and (3.1) in the third and fourth inequality respectively.

For terms  $I_2(t)$  and  $I_3(t)$ , by condition (2.1), estimates (3.1) and (3.2), we get

$$\begin{aligned} &\mathbb{E} \left( \int_{\delta}^T |I_2(t)| dt \right) \\ &\leq C \mathbb{E} \left[ \int_{\delta}^T \int_{t-\delta}^t (1 + \|X_s^\varepsilon\|_{H_1} + \|Y_s^\varepsilon\|_{H_2})(\|X_s^\varepsilon\|_{H_1} + \|X_{t-\delta}^\varepsilon\|_{H_1}) ds dt \right] \\ &\leq C \delta \mathbb{E} \left[ \sup_{s \in [0, T]} (1 + \|X_s^\varepsilon\|_{H_1}^2) \right] + C \mathbb{E} \left[ \sup_{s \in [0, T]} \|X_s^\varepsilon\|_{H_1} \int_{\delta}^T \int_{t-\delta}^t \|Y_s^\varepsilon\|_{H_2} ds dt \right] \\ &\leq C \delta \mathbb{E} \left[ \sup_{s \in [0, T]} (1 + \|X_s^\varepsilon\|_{H_1}^2) \right] + C_T \delta^{1/2} \mathbb{E} \left[ \sup_{s \in [0, T]} \|X_s^\varepsilon\|_{H_1}^2 \right]^{1/2} \left( \int_{\delta}^T \int_{t-\delta}^t \|Y_s^\varepsilon\|_{H_2}^2 ds dt \right)^{1/2} \\ &\leq C_T \delta (1 + \|x\|_{H_1}^2 + \|y\|_{H_2}^2) \end{aligned} \quad (3.9)$$

and

$$\begin{aligned} \mathbb{E} \left( \int_{\delta}^T |I_3(t)| dt \right) &\leq C \mathbb{E} \left( \int_{\delta}^T \int_{t-\delta}^t (1 + \|X_s^\varepsilon\|_{H_1}^2) ds dt \right) \\ &\leq C_T \delta \mathbb{E} \left[ \sup_{s \in [0, T]} (1 + \|X_s^\varepsilon\|_{H_1}^2) \right] \\ &\leq C_T \delta (1 + \|x\|_{H_1}^2 + \|y\|_{H_2}^2). \end{aligned} \quad (3.10)$$

For the term  $I_4(t)$ , note that

$$\left\{ \int_{t-\delta}^u \langle X_s^\varepsilon - X_{t-\delta}^\varepsilon, G_1(\tilde{X}_s^\varepsilon) dW_s^1 \rangle_{H_1} \right\}_{t-\delta \leq u \leq T}$$

is a local martingale, then applying Burkholder-Davies-Gundy inequality, it follows

$$\begin{aligned} \mathbb{E} \left( \int_{\delta}^T |I_4(t)| dt \right) &\leq C \mathbb{E} \int_{\delta}^T \left[ \int_{t-\delta}^t \|G_1(\tilde{X}_s^\varepsilon)\|_{L_2(U_1, H_1)}^2 \|X_s^\varepsilon - X_{t-\delta}^\varepsilon\|_{H_1}^2 ds \right]^{1/2} dt \\ &\leq C_T \left[ \mathbb{E} \int_{\delta}^T \int_{t-\delta}^t (1 + \|X_s^\varepsilon\|_{H_1}^2) \|X_s^\varepsilon - X_{t-\delta}^\varepsilon\|_{H_1}^2 ds dt \right]^{1/2} \end{aligned}$$

$$\begin{aligned}
 &\leq C_T \delta^{1/2} \left[ \mathbb{E} \sup_{s \in [0, T]} \left( 1 + \|X_s^\varepsilon\|_{H_1}^4 \right) \right]^{1/2} \\
 &\leq C_T \delta^{1/2} (1 + \|x\|_{H_1}^2 + \|y\|_{H_2}^2).
 \end{aligned} \tag{3.11}$$

Combining estimates (3.8)-(3.11), we obtain

$$\mathbb{E} \left( \int_\delta^T \|X_t^\varepsilon - X_{t-\delta}^\varepsilon\|_{H_1}^2 dt \right) \leq C_T (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m) \delta^{1/2}. \tag{3.12}$$

By the same argument above, we also have

$$\mathbb{E} \left( \int_\delta^T \|X_{t(\delta)}^\varepsilon - X_{t-\delta}^\varepsilon\|_{H_1}^2 dt \right) \leq C_T (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m) \delta^{1/2}. \tag{3.13}$$

Hence, the result (3.5) holds by estimates (3.6), (3.12) and (3.13). The proof is complete.  $\square$

**3.2. Construction of the auxiliary process.** Based on the method of time discretization, which is inspired by [18], we first construct an auxiliary process  $\hat{Y}_t^\varepsilon \in H_2$  satisfying the following equation

$$d\hat{Y}_t^\varepsilon = \frac{1}{\varepsilon} B(X_{t(\delta)}^\varepsilon, \hat{Y}_t^\varepsilon) dt + \frac{1}{\sqrt{\varepsilon}} G_2(X_{t(\delta)}^\varepsilon, \hat{Y}_t^\varepsilon) dW_t^2, \quad \hat{Y}_0^\varepsilon = y \in H_2,$$

where  $\delta$  is a fixed positive number depending on  $\varepsilon$  and will be chosen later. Then for its  $dt \otimes \mathbb{P}$ -equivalence class  $\check{Y}^\varepsilon$  we have  $\check{Y}^\varepsilon \in L^\kappa([0, T] \times \Omega, dt \otimes \mathbb{P}; V_2) \cap L^2([0, T] \times \Omega, dt \otimes \mathbb{P}; H_2)$  with  $\kappa$  as in B3, and for any  $k \in \mathbb{N}$  and  $t \in [k\delta, \min((k+1)\delta, T)]$ ,  $\mathbb{P}$ -a.s.

$$\hat{Y}_t^\varepsilon = \hat{Y}_{k\delta}^\varepsilon + \frac{1}{\varepsilon} \int_{k\delta}^t B(X_{k\delta}^\varepsilon, \tilde{Y}_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_{k\delta}^t G_2(X_{k\delta}^\varepsilon, \tilde{Y}_s^\varepsilon) dW_s^2, \tag{3.14}$$

where  $\tilde{Y}^\varepsilon$  is any  $V_2$ -valued progressively measurable  $dt \otimes \mathbb{P}$ -version of  $\check{Y}^\varepsilon$ .

By the construction of  $\hat{Y}_t^\varepsilon$ , we obtain the following estimates, which will be used later.

**Lemma 3.3.** *For any  $T > 0$  and  $\varepsilon \in (0, 1)$ , there exist a constant  $C_T > 0$  and  $m \in \mathbb{N}$  such that*

$$\sup_{t \in [0, T]} \mathbb{E} \|\hat{Y}_t^\varepsilon\|_{H_2}^2 \leq C_T (1 + \|x\|_{H_1}^2 + \|y\|_{H_2}^2) \tag{3.15}$$

and

$$\mathbb{E} \left( \int_0^T \|Y_t^\varepsilon - \hat{Y}_t^\varepsilon\|_{H_2}^2 dt \right) \leq C_T (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m) \delta^{1/2}. \tag{3.16}$$

*Proof.* Because the proof of estimate (3.15) follows almost the same steps as in the proof of Lemma 3.1, we omit its proof and only prove (3.16) here.

Note that

$$\begin{aligned}
 Y_t^\varepsilon - \hat{Y}_t^\varepsilon &= \frac{1}{\varepsilon} \int_0^t \left[ B(X_s^\varepsilon, \tilde{Y}_s^\varepsilon) - B(X_{s(\delta)}^\varepsilon, \tilde{Y}_s^\varepsilon) \right] ds \\
 &\quad + \frac{1}{\sqrt{\varepsilon}} \int_0^t \left[ G_2(X_s^\varepsilon, \tilde{Y}_s^\varepsilon) - G_2(X_{s(\delta)}^\varepsilon, \tilde{Y}_s^\varepsilon) \right] dW_s^2.
 \end{aligned}$$

Applying Itô's formula, we obtain

$$\mathbb{E} \|Y_t^\varepsilon - \hat{Y}_t^\varepsilon\|_{H_2}^2 = \frac{2}{\varepsilon} \mathbb{E} \int_0^t v_2^* \langle B(X_s^\varepsilon, \tilde{Y}_s^\varepsilon) - B(X_{s(\delta)}^\varepsilon, \tilde{Y}_s^\varepsilon), \tilde{Y}_s^\varepsilon - \hat{Y}_s^\varepsilon \rangle_{V_2} ds$$

$$+\frac{1}{\varepsilon}\mathbb{E}\int_0^t\|G_2(X_s^\varepsilon,\tilde{Y}_s^\varepsilon)-G_2(X_{s(\delta)}^\varepsilon,\tilde{Y}_s^\varepsilon)\|_{L_2(U_2,H_2)}^2ds.$$

Then by condition **B2**, there exists  $\gamma > 0$  such that

$$\frac{d}{dt}\mathbb{E}\|Y_t^\varepsilon-\hat{Y}_t^\varepsilon\|_{H_2}^2\leq-\frac{\gamma}{\varepsilon}\mathbb{E}\|Y_t^\varepsilon-\hat{Y}_t^\varepsilon\|_{H_2}^2+\frac{C}{\varepsilon}\mathbb{E}\|X_t^\varepsilon-X_{t(\delta)}^\varepsilon\|_{H_1}^2.$$

The the same argument used in (3.4), we have

$$\mathbb{E}\|Y_t^\varepsilon-\hat{Y}_t^\varepsilon\|_{H_2}^2\leq\frac{C}{\varepsilon}\int_0^te^{-\frac{\gamma(t-s)}{\varepsilon}}\mathbb{E}\|X_s^\varepsilon-X_{s(\delta)}^\varepsilon\|_{H_1}^2ds.$$

Then applying Fubini's theorem, for any  $T > 0$ ,

$$\begin{aligned}\mathbb{E}\left(\int_0^T\|Y_t^\varepsilon-\hat{Y}_t^\varepsilon\|_{H_2}^2dt\right)&\leq\frac{C}{\varepsilon}\int_0^T\int_0^te^{-\frac{\beta(t-s)}{\varepsilon}}\mathbb{E}\|X_s^\varepsilon-X_{s(\delta)}^\varepsilon\|_{H_1}^2dsdt \\ &=\frac{C}{\varepsilon}\mathbb{E}\left[\int_0^T\|X_s^\varepsilon-X_{s(\delta)}^\varepsilon\|_{H_1}^2\left(\int_s^Te^{-\frac{\beta(t-s)}{\varepsilon}}dt\right)ds\right] \\ &\leq C\mathbb{E}\left(\int_0^T\|X_s^\varepsilon-X_{s(\delta)}^\varepsilon\|_{H_1}^2ds\right).\end{aligned}$$

Therefore, by Lemma 3.2, we obtain

$$\mathbb{E}\left(\int_0^T\|Y_t^\varepsilon-\hat{Y}_t^\varepsilon\|_{H_2}^2dt\right)\leq C_T(1+\|x\|_{H_1}^m+\|y\|_{H_2}^m)\delta^{1/2}.$$

The proof is complete.  $\square$

**3.3. The ergodicity of the frozen equation.** The frozen equation associated to the fast motion for fixed slow component  $x \in H_1$  is as follows,

$$\begin{cases}dY_t=B(x,Y_t)dt+G_2(x,Y_t)d\bar{W}_t^2, \\ Y_0=y\in H_2,\end{cases}\quad (3.17)$$

where  $\{\bar{W}_t^2\}_{t\geq 0}$  is a cylindrical  $\tilde{\mathcal{F}}_t$ -Wiener process in a separable Hilbert space  $U_2$  on another probability space  $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$  with natural filtration  $\tilde{\mathcal{F}}_t$ .

Under the assumptions **B1-B4**, for any fixed  $x \in H_1$  and initial data  $y \in H_2$ , equation (3.17) has a unique variational solution  $Y_t^{x,y}$  in the sense of Definition 2.2, i.e., for its  $dt \otimes \tilde{\mathbb{P}}$ -equivalence class  $\hat{Y}$  we have  $\hat{Y}^{x,y} \in L^\kappa([0,T] \times \tilde{\Omega}, dt \otimes \tilde{\mathbb{P}}; V_2) \cap L^2([0,T] \times \tilde{\Omega}, dt \otimes \tilde{\mathbb{P}}; H_2)$  with  $\kappa$  as in **B3**, we have  $\tilde{\mathbb{P}}$ -a.s.

$$Y_t^{x,y}=y+\int_0^tB(x,\tilde{Y}_s^{x,y})ds+\int_0^tG_2(x,\tilde{Y}_s^{x,y})d\bar{W}_s^2,\quad (3.18)$$

where  $\tilde{Y}^{x,y}$  is any  $V_2$ -valued progressively measurable  $dt \otimes \tilde{\mathbb{P}}$ -version of  $\hat{Y}^{x,y}$ . By the same arguments as in the proof of Lemma 3.1, it is easy to prove that

$$\sup_{t\geq 0}\tilde{\mathbb{E}}\|Y_t^{x,y}\|_{H_2}^2\leq C(1+\|x\|_{H_1}^2+\|y\|_{H_2}^2).$$

Let  $\{P_t^x\}_{t\geq 0}$  be the transition semigroup of the Markov process  $\{Y_t^{x,y}\}_{t\geq 0}$ , that is, for any bounded measurable function  $\varphi$  on  $H_2$ ,

$$P_t^x\varphi(y)=\tilde{\mathbb{E}}[\varphi(Y_t^{x,y})],\quad y\in H_2,\quad t>0,$$

where  $\tilde{\mathbb{E}}$  is the expectation on  $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ . Then we have the following asymptotic behavior of  $P_t^x$ , whose proof can be founded in [21, Theorem 4.3.9].

**Proposition 3.4.** *The transition semigroup  $\{P_t^x\}_{t \geq 0}$  has a unique invariant measure  $\mu^x$ . Moreover, there exists a constant  $C > 0$  such that for any Lipschitz function  $\varphi : H_2 \rightarrow R$ ,*

$$\left| P_t^x \varphi(y) - \int_{H_2} \varphi(z) \mu^x(dz) \right| \leq C(1 + \|x\|_{H_1} + \|y\|_{H_2}) e^{-\frac{\gamma t}{2}} \|\varphi\|_{Lip}, \quad (3.19)$$

where  $\|\varphi\|_{Lip} = \sup_{y_1 \neq y_2 \in H_2} \frac{|\varphi(y_1) - \varphi(y_2)|}{\|y_1 - y_2\|_{H_2}}$ .

**3.4. The averaged equation.** We consider the corresponding averaged equation, i.e.

$$\begin{cases} d\bar{X}_t = A(\bar{X}_t)dt + \bar{F}(\bar{X}_t)dt + G_1(\bar{X}_t)dW_t^1, \\ \bar{X}_0 = x \in H_1, \end{cases} \quad (3.20)$$

with the averaged coefficient

$$\bar{F}(x) := \int_{H_2} F(x, y) \mu^x(dy), \quad x \in H_1,$$

where  $\mu^x$  is the unique invariant measure for the transition semigroup  $\{P_t^x\}_{t \geq 0}$ .

Since  $F$  is Lipschitz continuous, it is easy to check  $\bar{F}$  is also Lipschitz continuous, i.e.

$$\|\bar{F}(u) - \bar{F}(v)\|_{H_1} \leq C\|u - v\|_{H_1}, \quad u, v \in H_1.$$

Then equation (3.20) has a unique variational solution  $\bar{X}$  in the sense of Definition 2.2, i.e., for its  $dt \otimes \mathbb{P}$ -equivalence class  $\check{\bar{X}}$  we have  $\check{\bar{X}} \in L^\alpha([0, T] \times \Omega, dt \otimes \mathbb{P}; V_1) \cap L^2([0, T] \times \Omega, dt \otimes \mathbb{P}; H_1)$  with  $\alpha$  as in A3, we have  $\mathbb{P}$ -a.s.

$$\bar{X}_t = x + \int_0^t A(\check{\bar{X}}_s)ds + \int_0^t \bar{F}(\bar{X}_s)ds + \int_0^t G_1(\check{\bar{X}}_s)dW_s^1, \quad (3.21)$$

where  $\check{\bar{X}}$  is any  $V_1$ -valued progressively measurable  $dt \otimes \mathbb{P}$ -version of  $\check{\bar{X}}$ . Moreover, we also have the following estimates. Because their proofs follows almost the same steps in the proof of Lemmas 3.1 and 3.2, we omit them here.

**Lemma 3.5.** *For any  $T > 0$ ,  $p \geq 1$ , there exist constants  $C_{p,T}, C_T > 0$  and  $m > 0$  such that for any  $x \in H_1$ ,*

$$\mathbb{E} \left( \sup_{t \in [0, T]} \|\bar{X}_t\|_{H_1}^{2p} \right) + \mathbb{E} \left( \int_0^T \|\bar{X}_t\|_{H_1}^{2p-2} \|\check{\bar{X}}_t\|_{V_1}^\alpha dt \right) \leq C_{p,T}(1 + \|x\|_{H_1}^{2p})$$

and

$$\mathbb{E} \left[ \int_0^T \|\bar{X}_t - \bar{X}_{t(\delta)}\|_{H_1}^2 dt \right] \leq C_T \delta^{1/2} (1 + \|x\|_{H_1}^m). \quad (3.22)$$

Next, we intend to prove that  $X_t^\varepsilon$  strongly converges to  $\bar{X}_t$  for  $t < \tau_R$  firstly, then the proof of the main result will follow from the fact that the difference process  $X_t^\varepsilon - \bar{X}_t$  after the stopping time is sufficient small when  $R$  is large enough, whose proof is given in Subsection 3.5.

**Proposition 3.6.** *For any  $(x, y) \in H_1 \times H_2$ ,  $T, R > 0$  and  $\varepsilon \in (0, 1)$ , then there exist constants  $C_{R,T}, m > 0$  such that*

$$\mathbb{E} \left( \sup_{t \in [0, T \wedge \tau_R]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right) \leq C_{R,T} (1 + \|x\|_{H_1}^m + \|y\|_{H_1}^m) \left( \frac{\varepsilon}{\delta} + \frac{\varepsilon^{1/2}}{\delta^{1/2}} + \delta^{1/2} + \delta^{1/4} \right), \quad (3.23)$$

where

$$\tau_R := \inf \left\{ t \geq 0 : \int_0^t (1 + \|\tilde{X}_s\|_{V_1}^\alpha) (1 + \|\bar{X}_s\|_{H_1}^\beta) ds \geq R \right\}.$$

*Proof.* We will divide the proof into three steps.

**Step 1.** We note that

$$X_t^\varepsilon - \bar{X}_t = \int_0^t [A(\tilde{X}_s^\varepsilon) - A(\tilde{X}_s)] ds + [F(X_s^\varepsilon, Y_s^\varepsilon) - \bar{F}(\bar{X}_s)] ds + [G_1(\tilde{X}_s^\varepsilon) - G_1(\tilde{X}_s)] dW_s^1.$$

Applying Itô's formula, we have

$$\begin{aligned} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 &= 2 \int_0^t \langle A(\tilde{X}_s^\varepsilon) - A(\tilde{X}_s), \tilde{X}_s^\varepsilon - \tilde{X}_s \rangle_{V_1} ds + \int_0^t \|G_1(\tilde{X}_s^\varepsilon) - G_1(\tilde{X}_s)\|_{L_2(U_1, H_1)}^2 ds \\ &\quad + 2 \int_0^t \langle [F(X_s^\varepsilon, Y_s^\varepsilon) - \bar{F}(\bar{X}_s)], X_s^\varepsilon - \bar{X}_s \rangle_{H_1} ds \\ &\quad + 2 \int_0^t \langle X_s^\varepsilon - \bar{X}_s, [G_1(\tilde{X}_s^\varepsilon) - G_1(\tilde{X}_s)] dW_s^1 \rangle_{H_1} \\ &= 2 \int_0^t \langle A(\tilde{X}_s^\varepsilon) - A(\tilde{X}_s), \tilde{X}_s^\varepsilon - \tilde{X}_s \rangle_{V_1} ds + \int_0^t \|G_1(\tilde{X}_s^\varepsilon) - G_1(\tilde{X}_s)\|_{L_2(U_1, H_1)}^2 ds \\ &\quad + 2 \int_0^t \langle [\bar{F}(X_s^\varepsilon) - \bar{F}(\bar{X}_s)], X_s^\varepsilon - \bar{X}_s \rangle_{H_1} ds \\ &\quad + 2 \int_0^t \langle [F(X_s^\varepsilon, Y_s^\varepsilon) - \bar{F}(X_s^\varepsilon) - F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) + \bar{F}(X_{s(\delta)}^\varepsilon)], X_s^\varepsilon - \bar{X}_s \rangle_{H_1} ds \\ &\quad + 2 \int_0^t \langle [F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon)], X_s^\varepsilon - X_{s(\delta)}^\varepsilon - \bar{X}_s + \bar{X}_{s(\delta)} \rangle_{H_1} ds \\ &\quad + 2 \int_0^t \langle [F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon)], X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \\ &\quad + 2 \int_0^t \langle X_s^\varepsilon - \bar{X}_s, [G_1(\tilde{X}_s^\varepsilon) - G_1(\tilde{X}_s)] dW_s^1 \rangle_{H_1}. \end{aligned}$$

Then conditions **A2** and **A3** imply  $\mathbb{P}$ -a.s.,

$$\begin{aligned} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 &\leq C \int_0^t \|X_s^\varepsilon - \bar{X}_s\|_{H_1}^2 (1 + \|\tilde{X}_s\|_{V_1}^\alpha) (1 + \|\bar{X}_s\|_{H_1}^\beta) ds \\ &\quad + C \int_0^t (\|X_s^\varepsilon - X_{s(\delta)}^\varepsilon\|_{H_1}^2 + \|Y_s^\varepsilon - \hat{Y}_s^\varepsilon\|_{H_2}^2 + \|\bar{X}_s - \bar{X}_{s(\delta)}\|_{H_1}^2) ds \\ &\quad + C \left[ \int_0^t \|F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon)\|_{H_1}^2 ds \right]^{1/2} \\ &\quad \quad \times \left[ \int_0^t (\|X_s^\varepsilon - X_{s(\delta)}^\varepsilon\|_{H_1}^2 + \|\bar{X}_s - \bar{X}_{s(\delta)}\|_{H_1}^2) ds \right]^{1/2} \\ &\quad + 2 \left| \int_0^t \langle [F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon)], X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \right| \\ &\quad + 2 \left| \int_0^t \langle X_s^\varepsilon - \bar{X}_s, [G_1(X_s^\varepsilon) - G_1(\tilde{X}_s)] dW_s^1 \rangle_{H_1} \right|. \end{aligned}$$

Applying Gronwall's inequality and the definition of the stopping time  $\tau_R$ , we deduce that

$$\begin{aligned} \sup_{t \in [0, T \wedge \tau_R]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 &\leq C_{R,T} \left\{ \int_0^T \left( \|X_s^\varepsilon - X_{s(\delta)}^\varepsilon\|_{H_1}^2 + \|Y_s^\varepsilon - \hat{Y}_s\|_{H_2}^2 + \|\bar{X}_s - \bar{X}_{s(\delta)}\|_{H_1}^2 \right) ds \right. \\ &\quad + \left[ \int_0^T \left( 1 + \|X_{s(\delta)}^\varepsilon\|_{H_1}^2 + \|\hat{Y}_s^\varepsilon\|_{H_2}^2 \right) ds \right]^{1/2} \\ &\quad \times \left[ \int_0^t \|X_s^\varepsilon - X_{s(\delta)}^\varepsilon\|_{H_1}^2 + \|\bar{X}_s - \bar{X}_{s(\delta)}\|_{H_1}^2 ds \right]^{1/2} \\ &\quad + \sup_{t \in [0, T]} \left| \int_0^t \left\langle [F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon)], X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \right\rangle_{H_1} ds \right| \\ &\quad \left. + \sup_{t \in [0, T \wedge \tau_R]} \left| \int_0^t \left\langle X_s^\varepsilon - \bar{X}_s, [G_1(X_s^\varepsilon) - G_1(\tilde{X}_s)] dW_s^1 \right\rangle_{H_1} \right| \right\}. \end{aligned}$$

Note that

$$\left\{ \int_0^t \left\langle X_s^\varepsilon - \bar{X}_s, [G_1(X_s^\varepsilon) - G_1(\tilde{X}_s)] dW_s^1 \right\rangle_{H_1} \right\}_{0 \leq t \leq T}$$

is a local martingale and  $T \wedge \tau_R$  is a stopping time, thus applying Burkholder-Davis inequality (see e.g. [21, Proposition D.0.1]), estimates (3.5), (3.16) and (3.22), there exists  $m > 0$  such that

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0, T \wedge \tau_R]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right) &\leq C_{R,T} \left( 1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m \right) \delta^{1/4} + \frac{1}{2} \mathbb{E} \left( \sup_{t \in [0, T \wedge \tau_R]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right) \\ &\quad + C_{R,T} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \int_0^t \left\langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \right\rangle_{H_1} ds \right| \right] \\ &\quad + C_{R,T} \mathbb{E} \left( \int_0^{T \wedge \tau_R} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 dt \right), \end{aligned}$$

which implies

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0, T \wedge \tau_R]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right) &\leq C_{R,T} \left( 1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m \right) \delta^{1/4} \\ &\quad + C_{R,T} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \int_0^t \left\langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \right\rangle_{H_1} ds \right| \right] \\ &\quad + C_{R,T} \int_0^T \mathbb{E} \left( \sup_{s \in [0, t \wedge \tau_R]} \|X_s^\varepsilon - \bar{X}_s\|_{H_1}^2 \right) dt. \end{aligned}$$

Applying Gronwall's inequality, we finally get

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0, T \wedge \tau_R]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right) &\leq C_{R,T} \left( 1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m \right) \delta^{1/4} \\ &\quad + C_{R,T} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \int_0^t \left\langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \right\rangle_{H_1} ds \right| \right]. \end{aligned}$$

Hence, the proof will be completed by the following estimate:

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left| \int_0^t \left\langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \right\rangle_{H_1} ds \right| \right]$$

$$\leq C_T \left( 1 + \|x\|_{H_1}^2 + \|y\|_{H_2}^2 \right) \left( \frac{\varepsilon}{\delta} + \frac{\varepsilon^{1/2}}{\delta^{1/2}} + \delta^{1/2} \right), \quad (3.24)$$

whose proof will be given in Step 2.

**Step 2.** We note that

$$\begin{aligned} & \left| \int_0^t \langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \right| \\ &= \left| \sum_{k=0}^{\lfloor t/\delta \rfloor - 1} \int_{k\delta}^{(k+1)\delta} \langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \right. \\ & \quad \left. + \int_{t(\delta)}^t \langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \right| \\ &\leq \sum_{k=0}^{\lfloor t/\delta \rfloor - 1} \left| \int_{k\delta}^{(k+1)\delta} \langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \right| \\ & \quad + \left| \int_{t(\delta)}^t \langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \right| \\ &:= J_1(t) + J_2(t). \end{aligned} \quad (3.25)$$

For the term  $J_2(t)$ , it is easy to see

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0, T]} J_2(t) \right] &\leq C \left[ \mathbb{E} \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right]^{1/2} \left[ \mathbb{E} \sup_{t \in [0, T]} \left| \int_{t(\delta)}^t (1 + \|X_{s(\delta)}^\varepsilon\|_{H_1} + \|\hat{Y}_s^\varepsilon\|_{H_2}) ds \right|^2 \right]^{1/2} \\ &\leq C \left[ \mathbb{E} \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right]^{1/2} \left[ \mathbb{E} \int_0^T (1 + \|X_{s(\delta)}^\varepsilon\|_{H_1}^2 + \|\hat{Y}_s^\varepsilon\|_{H_2}^2) ds \right]^{1/2} \delta^{1/2} \\ &\leq C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) \delta^{1/2}. \end{aligned} \quad (3.26)$$

For the term  $J_1(t)$ , we have

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0, T]} J_1(t) \right] &\leq \mathbb{E} \sum_{k=0}^{\lfloor T/\delta \rfloor - 1} \left| \int_{k\delta}^{(k+1)\delta} \langle F(X_{k\delta}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{k\delta}^\varepsilon), X_{k\delta}^\varepsilon - \bar{X}_{k\delta} \rangle_{H_1} ds \right| \\ &\leq \frac{C_T}{\delta} \max_{0 \leq k \leq \lfloor T/\delta \rfloor - 1} \mathbb{E} \left| \int_{k\delta}^{(k+1)\delta} \langle F(X_{k\delta}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{k\delta}^\varepsilon), X_{k\delta}^\varepsilon - \bar{X}_{k\delta} \rangle_{H_1} ds \right| \\ &\leq \frac{C_T \varepsilon}{\delta} \max_{0 \leq k \leq \lfloor T/\delta \rfloor - 1} \left[ \mathbb{E} \|X_{k\delta}^\varepsilon - \bar{X}_{k\delta}\|_{H_1}^2 \right]^{1/2} \left[ \mathbb{E} \left\| \int_0^{\frac{\delta}{\varepsilon}} F(X_{k\delta}^\varepsilon, \hat{Y}_{s\varepsilon+k\delta}^\varepsilon) - \bar{F}(X_{k\delta}^\varepsilon) ds \right\|_{H_1}^2 \right]^{1/2} \\ &\leq \frac{C_T (1 + \|x\|_{H_1} + \|y\|_{H_2}) \varepsilon}{\delta} \max_{0 \leq k \leq \lfloor T/\delta \rfloor - 1} \left[ \int_0^{\frac{\delta}{\varepsilon}} \int_r^{\frac{\delta}{\varepsilon}} \Psi_k(s, r) ds dr \right]^{1/2}, \end{aligned}$$

where for any  $0 \leq r \leq s \leq \frac{\delta}{\varepsilon}$ ,

$$\Psi_k(s, r) := \mathbb{E} \left[ \langle F(X_{k\delta}^\varepsilon, \hat{Y}_{s\varepsilon+k\delta}^\varepsilon) - \bar{F}(X_{k\delta}^\varepsilon), F(X_{k\delta}^\varepsilon, \hat{Y}_{r\varepsilon+k\delta}^\varepsilon) - \bar{F}(X_{k\delta}^\varepsilon) \rangle_{H_1} \right],$$

and

$$\Psi_k(s, r) \leq C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) e^{-\frac{(s-r)\gamma}{2}}, \quad (3.27)$$

whose proof will be presented in Step 3. Hence, we get

$$\begin{aligned}
 & \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \int_0^t \langle F(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{F}(X_{s(\delta)}^\varepsilon), X_{s(\delta)}^\varepsilon - \bar{X}_{s(\delta)} \rangle_{H_1} ds \right| \right] \\
 & \leq C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) \frac{\varepsilon}{\delta} \left[ \int_0^{\frac{\delta}{\varepsilon}} \int_r^{\frac{\delta}{\varepsilon}} e^{-\frac{(s-r)\gamma}{2}} ds dr \right]^{1/2} + C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) \delta^{1/2} \\
 & = C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) \frac{\varepsilon}{\delta} \left( \frac{\delta}{\gamma \varepsilon} - \frac{1}{\gamma^2} + \frac{1}{\gamma^2} e^{-\frac{\gamma \delta}{\varepsilon}} \right)^{1/2} + C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) \delta^{1/2} \\
 & \leq C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) \left( \frac{\varepsilon}{\delta} + \frac{\varepsilon^{1/2}}{\delta^{1/2}} + \delta^{1/2} \right),
 \end{aligned}$$

which completes the proof of estimate (3.24).

**Step 3.** For any  $s > 0$ , and any  $\mathcal{F}_s$ -measurable  $H_1$ -valued random variable  $X$  and  $H_2$ -valued random variable  $Y$ , we consider the following equation:

$$\begin{cases} dY_t = \frac{1}{\varepsilon} B(X, Y_t) dt + \frac{1}{\sqrt{\varepsilon}} G_2(X, Y_t) dW_t^2, & t \geq s, \\ Y_s = Y, \end{cases}$$

which has a unique solution  $\tilde{Y}_t^{\varepsilon, s, X, Y}$ . Then by the construction of  $\hat{Y}_t^\varepsilon$ , for any  $k \in \mathbb{N}_*$  and  $t \in [k\delta, (k+1)\delta]$  we have  $\mathbb{P}$ -a.s.,

$$\hat{Y}_t^\varepsilon = \tilde{Y}_t^{\varepsilon, k\delta, X_{k\delta}^\varepsilon, \hat{Y}_{k\delta}^\varepsilon},$$

which implies

$$\begin{aligned}
 \Psi_k(s, r) &= \mathbb{E} \left[ \langle F(X_{k\delta}^\varepsilon, \tilde{Y}_{s\varepsilon+k\delta}^{\varepsilon, k\delta, X_{k\delta}^\varepsilon, \hat{Y}_{k\delta}^\varepsilon}) - \bar{F}(X_{k\delta}^\varepsilon), F(X_{k\delta}^\varepsilon, \tilde{Y}_{r\varepsilon+k\delta}^{\varepsilon, k\delta, X_{k\delta}^\varepsilon, \hat{Y}_{k\delta}^\varepsilon}) - \bar{F}(X_{k\delta}^\varepsilon) \rangle_{H_1} \right] \\
 &= \int_{\Omega} \mathbb{E} \left[ \langle F(X_{k\delta}^\varepsilon, \tilde{Y}_{s\varepsilon+k\delta}^{\varepsilon, k\delta, X_{k\delta}^\varepsilon, \hat{Y}_{k\delta}^\varepsilon}) - \bar{F}(X_{k\delta}^\varepsilon), \right. \\
 &\quad \left. F(X_{k\delta}^\varepsilon, \tilde{Y}_{r\varepsilon+k\delta}^{\varepsilon, k\delta, X_{k\delta}^\varepsilon, \hat{Y}_{k\delta}^\varepsilon}) - \bar{F}(X_{k\delta}^\varepsilon) \rangle_{H_1} \middle| \mathcal{F}_{k\delta} \right] (\omega) \mathbb{P}(d\omega) \\
 &= \int_{\Omega} \mathbb{E} \left[ \langle F(X_{k\delta}^\varepsilon, \tilde{Y}_{s\varepsilon+k\delta}^{\varepsilon, k\delta, X_{k\delta}^\varepsilon(\omega), \hat{Y}_{k\delta}^\varepsilon(\omega)}) - \bar{F}(X_{k\delta}^\varepsilon(\omega)), \right. \\
 &\quad \left. F(X_{k\delta}^\varepsilon(\omega), \tilde{Y}_{r\varepsilon+k\delta}^{\varepsilon, k\delta, X_{k\delta}^\varepsilon(\omega), \hat{Y}_{k\delta}^\varepsilon(\omega)}) - \bar{F}(X_{k\delta}^\varepsilon(\omega)) \rangle_{H_1} \right] \mathbb{P}(d\omega),
 \end{aligned}$$

where the last equality comes from the fact that  $X_{k\delta}^\varepsilon$  and  $\hat{Y}_{k\delta}^\varepsilon$  are  $\mathcal{F}_{k\delta}$ -measurable, and for any fixed  $(x, y) \in H_1 \times H_2$ ,  $\{\tilde{Y}_{s\varepsilon+k\delta}^{\varepsilon, k\delta, x, y}\}_{s \geq 0}$  is independent of  $\mathcal{F}_{k\delta}$ .

By the definition of process  $\tilde{Y}_t^{\varepsilon, k\delta, x, y}$ , for its  $dt \otimes \mathbb{P}$ -equivalence class  $\check{Y}^{\varepsilon, k\delta, x, y}$  we have  $\check{Y}^{\varepsilon, k\delta, x, y} \in L^\kappa([k\delta, T] \times \Omega, dt \otimes \mathbb{P}; V_2) \cap L^2([k\delta, T] \times \Omega, dt \otimes \mathbb{P}; H_2)$  with  $\kappa$  as in **B3** and  $\mathbb{P}$ -a.s.

$$\begin{aligned}
 \tilde{Y}_{s\varepsilon+k\delta}^{\varepsilon, k\delta, x, y} &= y + \frac{1}{\varepsilon} \int_{k\delta}^{s\varepsilon+k\delta} B(x, \tilde{Y}_r^{\varepsilon, k\delta, x, y}) dr + \frac{1}{\sqrt{\varepsilon}} \int_{k\delta}^{s\varepsilon+k\delta} G_2(x, \tilde{Y}_r^{\varepsilon, k\delta, x, y}) dW_r^2 \\
 &= y + \frac{1}{\varepsilon} \int_0^{s\varepsilon} B(x, \tilde{Y}_{r+k\delta}^{\varepsilon, k\delta, x, y}) dr + \frac{1}{\sqrt{\varepsilon}} \int_0^{s\varepsilon} G_2(x, \tilde{Y}_{r+k\delta}^{\varepsilon, k\delta, x, y}) dW_r^{2, k\delta} \\
 &= y + \int_0^s B(x, \tilde{Y}_{r\varepsilon+k\delta}^{\varepsilon, k\delta, x, y}) dr + \int_0^s G_2(x, \tilde{Y}_{r\varepsilon+k\delta}^{\varepsilon, k\delta, x, y}) d\hat{W}_r^{2, k\delta}, \tag{3.28}
 \end{aligned}$$

where  $\tilde{Y}^{\varepsilon,k\delta,x,y}$  is any  $V_2$ -valued progressively measurable  $dt \otimes \mathbb{P}$ -version of  $\check{Y}^{\varepsilon,k\delta,x,y}$ ,  $\{W_r^{2,k\delta} := W_{r+k\delta}^2 - W_{k\delta}^2\}_{r \geq 0}$  and  $\{\hat{W}_t^{2,k\delta} := \frac{1}{\sqrt{\varepsilon}} W_{t\varepsilon}^{2,k\delta}\}_{t \geq 0}$ .

The uniqueness of the solution of equations (3.28) and (3.18) implies that the distribution of  $(\tilde{Y}_{s\varepsilon+k\delta}^{\varepsilon,k\delta,x,y})_{0 \leq s \leq \delta/\varepsilon}$  coincides with the distribution of  $(Y_s^{x,y})_{0 \leq s \leq \delta/\varepsilon}$ . Then by Proposition 3.4, estimates (3.1) and (3.15), we have

$$\begin{aligned}
\Psi_k(s, r) &= \int_{\Omega} \left[ \tilde{\mathbb{E}} \left\langle F \left( X_{k\delta}^{\varepsilon}(\omega), Y_s^{X_{k\delta}^{\varepsilon}(\omega), \hat{Y}_{k\delta}^{\varepsilon}(\omega)} \right) - \bar{F}(X_{k\delta}^{\varepsilon}(\omega)), \right. \right. \\
&\quad \left. \left. F \left( X_{k\delta}^{\varepsilon}(\omega), Y_r^{X_{k\delta}^{\varepsilon}(\omega), \hat{Y}_{k\delta}^{\varepsilon}(\omega)} \right) - \bar{F}(X_{k\delta}^{\varepsilon}(\omega)) \right\rangle_{H_1} \right] \mathbb{P}(d\omega) \\
&= \int_{\Omega} \int_{\tilde{\Omega}} \left\langle \tilde{\mathbb{E}} \left[ F \left( X_{k\delta}^{\varepsilon}(\omega), Y_{s-r}^{X_{k\delta}^{\varepsilon}(\omega), \hat{Y}_{k\delta}^{\varepsilon}(\omega)}(\tilde{\omega}) \right) - \bar{F}(X_{k\delta}^{\varepsilon}(\omega)) \right], \right. \\
&\quad \left. F \left( X_{k\delta}^{\varepsilon}(\omega), Y_r^{X_{k\delta}^{\varepsilon}(\omega), \hat{Y}_{k\delta}^{\varepsilon}(\omega)}(\tilde{\omega}) \right) - \bar{F}(X_{k\delta}^{\varepsilon}(\omega)) \right\rangle_{H_1} \tilde{\mathbb{P}}(d\tilde{\omega}) \mathbb{P}(d\omega) \\
&\leq \int_{\Omega} \int_{\tilde{\Omega}} \left[ 1 + \|X_{k\delta}^{\varepsilon}(\omega)\|_{H_1} + \|Y_r^{X_{k\delta}^{\varepsilon}(\omega), \hat{Y}_{k\delta}^{\varepsilon}(\omega)}(\tilde{\omega})\|_{H_2} \right] e^{-\frac{(s-r)\gamma}{2}} \\
&\quad \cdot \left[ (1 + \|X_{k\delta}^{\varepsilon}(\omega)\|_{H_1} + \|Y_r^{X_{k\delta}^{\varepsilon}(\omega), \hat{Y}_{k\delta}^{\varepsilon}(\omega)}(\tilde{\omega})\|_{H_2}) \right] \tilde{\mathbb{P}}(d\tilde{\omega}) \mathbb{P}(d\omega) \\
&\leq C_T \int_{\Omega} \left[ (1 + \|X_{k\delta}^{\varepsilon}(\omega)\|_{H_1}^2 + \|\hat{Y}_{k\delta}^{\varepsilon}(\omega)\|_{H_2}^2) \right] \mathbb{P}(d\omega) e^{-\frac{(s-r)\gamma}{2}} \\
&\leq C_T (\|x\|_{H_1}^2 + \|y\|_{H_2}^2 + 1) e^{-\frac{(s-r)\gamma}{2}},
\end{aligned}$$

which gives estimate (3.27). The proof is complete.  $\square$

**3.5. Proof of Theorem 2.4:** Applying Chebyshev's inequality, Lemmas 3.1 and 3.5, we have

$$\begin{aligned}
\mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^{\varepsilon} - \bar{X}_t\|_{H_1}^2 1_{\{T > \tau_R\}} \right) &\leq \left[ \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^{\varepsilon} - \bar{X}_t\|_{H_1}^4 \right) \right]^{1/2} \cdot [\mathbb{P}(T > \tau_R)]^{1/2} \\
&\leq \frac{C_T (1 + \|x\|_{H_1}^2 + \|y\|_{H_2}^2)}{\sqrt{R}} \\
&\quad \times \left[ \mathbb{E} \int_0^T (1 + \|\tilde{X}_s\|_{V_1}^{\alpha}) (1 + \|\bar{X}_s\|_{H_1}^{\beta}) ds \right]^{1/2} \\
&\leq \frac{C_T (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m)}{\sqrt{R}}, \tag{3.29}
\end{aligned}$$

where  $m$  is a positive constant. Then taking  $\delta = \varepsilon^{\frac{2}{3}}$ , estimates (3.23) and (3.29) give

$$\begin{aligned}
\mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^{\varepsilon} - \bar{X}_t\|_{H_1}^2 \right) &\leq \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^{\varepsilon} - \bar{X}_t\|_{H_1}^2 1_{\{T \leq \tau_R\}} \right) + \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^{\varepsilon} - \bar{X}_t\|_{H_1}^2 1_{\{T > \tau_R\}} \right) \\
&\leq C_{R, T} (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m) \varepsilon^{\frac{1}{6}} + \frac{C_T (1 + \|x\|_{H_1}^m + \|y\|_{H_2}^m)}{\sqrt{R}}.
\end{aligned}$$

Now, letting  $\varepsilon \rightarrow 0$  first, then  $R \rightarrow \infty$ , we have

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^{\varepsilon} - \bar{X}_t\|_{H_1}^2 \right) = 0. \tag{3.30}$$

Note that for any  $p > 1$ , by Lemmas 3.1 and 3.5, it is easy to see that

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p} \right) &= \mathbb{E} \left[ \sup_{t \in [0, T]} \left( \|X_t^\varepsilon - \bar{X}_t\|_{H_1} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p-1} \right) \right] \\ &\leq C_p \left[ \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right) \right]^{1/2} \left[ \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{4p-2} \right) \right]^{1/2} \\ &\leq C_{p, T} (1 + \|x\|_{H_1}^{2p-1} + \|y\|_{H_2}^{2p-1}) \left[ \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^2 \right) \right]^{1/2}. \end{aligned}$$

Hence, by (3.30), we finally get

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p} \right) = 0.$$

The proof is complete.  $\square$

#### 4. APPLICATION TO EXAMPLES

In this section we will apply our main result to establish the averaging principle for stochastic porous medium equations,  $p$ -Laplace equations, Burgers equations and 2D Navier-Stokes equations with slow and fast time-scales. Note that we here mainly focus on the nonlinear operator  $A$ , so we take the stochastic porous medium equation,  $p$ -Laplace equations, Burgers equations, or 2D Navier-Stokes equations for the slow component and stochastic heat equation with Lipschitz drift for the fast component for the simplicity.

Let  $\Lambda \subset \mathbb{R}^d$  be an open bounded domain and  $\Delta$  be the Laplace operator on  $\Lambda$  with Dirichlet boundary conditions, and for  $p \in [1, \infty)$  we use  $L^p(\Lambda)$  and  $H_0^{n, p}$  to denote the space of  $p$ -Lebesgue integrable functions on  $\Lambda$  and the Sobolev space of order  $n$  in  $L^p(\Lambda)$  with Dirichlet boundary conditions. Recall that  $X^*$  denotes the dual space of a Banach space  $X$ .

**4.1. Stochastic porous medium equations.** Let  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  be a function having the following properties :

( $\Psi$ 1)  $\Psi$  is continuous.

( $\Psi$ 2) For all  $s, t \in \mathbb{R}$

$$(t - s)(\Psi(t) - \Psi(s)) \geq 0.$$

( $\Psi$ 3) There exist  $p \in [2, \infty)$ ,  $c_1 \in (0, \infty)$ ,  $c_2 \in [0, \infty)$  such that for all  $s \in \mathbb{R}$

$$s\Psi(s) \geq c_1|s|^p - c_2.$$

( $\Psi$ 4) There exist  $c_3, c_4 \in (0, \infty)$  such that for all  $s \in \mathbb{R}$

$$|\Psi(s)| \leq c_3|s|^{p-1} + c_4,$$

where  $p$  is as in ( $\Psi$ 3).

Considering the Gelfand triple for the slow equation

$$V_1 := L^p(\Lambda) \subseteq H_1 := (H_0^{1,2}(\Lambda))^* \subseteq V_1^* := (L^p(\Lambda))^*$$

and the Gelfand triple for the fast equation

$$V_2 := H_0^{1,2}(\Lambda) \subseteq H_2 := L^2(\Lambda) \subseteq V_2^* := (H_0^{1,2}(\Lambda))^*.$$

Then we introduce the porous medium operator  $A(u) : V_1 \rightarrow V_1^*$  (see [21, Section 4.1] for details) by

$$A(u) = \Delta \Psi(u), \quad u \in V_1.$$

Now, we consider the slow-fast stochastic porous medium-heat equations

$$\begin{cases} dX_t^\varepsilon = [\Delta \Psi(X_t^\varepsilon) + F(X_t^\varepsilon, Y_t^\varepsilon)] dt + G_1(X_t^\varepsilon) dW_t^1, \\ dY_t^\varepsilon = \frac{1}{\varepsilon} [\Delta Y_t^\varepsilon + B_2(X_t^\varepsilon, Y_t^\varepsilon)] dt + \frac{1}{\sqrt{\varepsilon}} G_2(X_t^\varepsilon, Y_t^\varepsilon) dW_t^2, \\ X_0^\varepsilon = x \in H_1, Y_0^\varepsilon = y \in H_2, \end{cases} \quad (4.1)$$

where

$$F : H_1 \times H_2 \rightarrow H_1; \quad G_1 : V_1 \rightarrow L_2(U_1; H_1);$$

are measurable mappings and

$$B_2 : H_1 \times V_2 \rightarrow V_2^*; \quad G_2 : H_1 \times V_2 \rightarrow L_2(U_2; H_2)$$

are Lipschitz continuous. More precisely,

$$\|F(u_1, u_2) - F(v_1, v_2)\|_{H_1} \leq C(\|u_1 - v_1\|_{H_1} + \|u_2 - v_2\|_{H_2}); \quad (4.2)$$

$$\|G_1(u) - G_1(v)\|_{L_2(U_1, H_1)}^2 \leq C\|u - v\|_{H_1}^2; \quad (4.3)$$

$$\|B_2(u_1, u_2) - B_2(v_1, v_2)\|_{H_2} \leq C\|u_1 - v_1\|_{H_1} + L_{B_2}\|u_2 - v_2\|_{H_2}; \quad (4.4)$$

$$\|G_2(u_1, u) - G_2(v_1, v)\|_{L_2(U_2, H_2)} \leq C\|u_1 - v_1\|_{H_1} + L_{G_2}\|u - v\|_{H_2}. \quad (4.5)$$

Moreover, there exists  $\zeta \in (0, 1)$  such that

$$\|G_2(u_1, v)\|_{L_2(U_2, H_2)} \leq C(1 + \|u_1\|_{H_1} + \|v\|_{H_2}^\zeta) \quad (4.6)$$

and for the smallest eigenvalue  $\lambda_1$  of  $-\Delta$  in  $H_2$ , the Lipschitz constants  $L_{B_2}, L_{G_2}$  satisfy

$$2\lambda_1 - 2L_g - L_{\sigma_2}^2 > 0. \quad (4.7)$$

It is well known that the porous medium operator  $A$  satisfies the monotonicity and coercivity properties (see, e.g., [21, Pages 87-88]). So it is easy to check all the conditions **A1-A4** hold. Furthermore, the condition **B2** holds by condition (4.7) and the conditions **B1, B3** and **B4** hold obviously. Hence, by Theorem 2.4, we have

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p} \right) = 0, \quad \forall p \geq 1,$$

where  $\bar{X}_t$  is the solution of the corresponding averaged equation.

**4.2. Stochastic  $p$ -Laplace equation.** Now we consider the stochastic  $p$ -Laplace equation ( $p \geq 2$ ). We choose the Gelfand triple for the slow equation

$$V_1 := H_0^{1,p}(\Lambda) \subseteq H_1 := L^2(\Lambda) \subseteq V_1^* := (H_0^{1,p}(\Lambda))^*$$

and the Gelfand triple for the fast equation

$$V_2 := H_0^{1,2}(\Lambda) \subseteq H_2 := L^2(\Lambda) \subseteq V_2^* := (H_0^{1,2}(\Lambda))^*$$

and let  $A : V_1 \rightarrow V_1^*$  be

$$A(u) := \operatorname{div}(|\nabla u|^{p-2} \nabla u), \quad u \in V_1.$$

More precisely, given  $u \in V_1$ , we define

$$v_1^* \langle A(u), v \rangle_{V_1} := - \int_{\Lambda} |\nabla u(\xi)|^{p-2} \langle \nabla u(\xi), \nabla v(\xi) \rangle d\xi, \quad v \in V_1.$$

Here  $A$  is called the  $p$ -Laplace operator, also denoted by  $\Delta_p$ . Note that  $\Delta_2 = \Delta$ .

Now, we consider the slow-fast stochastic  $p$ -Laplace-heat equations

$$\begin{cases} dX_t^\varepsilon = \left[ \operatorname{div}(|\nabla X_t^\varepsilon|^{p-2} \nabla X_t^\varepsilon) + F(X_t^\varepsilon, Y_t^\varepsilon) \right] dt + G_1(X_t^\varepsilon) dW_t^1, \\ dY_t^\varepsilon = \frac{1}{\varepsilon} [\Delta Y_t^\varepsilon + B_2(X_t^\varepsilon, Y_t^\varepsilon)] dt + \frac{1}{\sqrt{\varepsilon}} G_2(X_t^\varepsilon, Y_t^\varepsilon) dW_t^2, \\ X_0^\varepsilon = x \in H_1, Y_0^\varepsilon = y \in H_2, \end{cases} \quad (4.8)$$

where the coefficients  $F, G_1, B_2$  and  $G_2$  satisfy conditions (4.2)-(4.7).

It is well known that the  $p$ -Laplace operator satisfies the monotonicity and coercivity properties (see, e.g., [21, Example 4.1.9]). So it is easy to check that all the conditions **A1-A4** and **B1-B4** hold. Hence, by Theorem 2.4, we have

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p} \right) = 0, \quad \forall p \geq 1,$$

where  $\bar{X}_t$  is the solution of the corresponding averaged equation.

Note that in the above two examples both the porous medium and the  $p$ -Laplace operators are globally monotone. But our result also applies to many merely locally monotone operators. For illustration, we will consider the stochastic Burgers and stochastic 2D Navier-Stokes equation below (see [21, Section 4.1 and 5.1] for a number of other examples).

**4.3. Stochastic Burgers equation.** Now we consider the stochastic Burgers equation. Taking  $\Lambda = (0, 1) \subset \mathbb{R}$ , we choose the Gelfand triple for the slow equation

$$V_1 := H_0^{1,2}(\Lambda) \subseteq H_1 := L^2(\Lambda) \subseteq V_1^* := (H_0^{1,2}(\Lambda))^*$$

and the Gelfand triple for the fast equation

$$V_2 := H_0^{1,2}(\Lambda) \subseteq H_2 := L^2(\Lambda) \subseteq V_2^* := (H_0^{1,2}(\Lambda))^*$$

and let  $A : V_1 \rightarrow V_1^*$  be

$$A(u) := \Delta u + u \nabla u, \quad u \in V_1.$$

Now, we consider the slow-fast stochastic Burgers-heat equations

$$\begin{cases} dX_t^\varepsilon = [\Delta X_t^\varepsilon + X_t^\varepsilon \nabla X_t^\varepsilon + F(X_t^\varepsilon, Y_t^\varepsilon)] dt + G_1(X_t^\varepsilon) dW_t^1, \\ dY_t^\varepsilon = \frac{1}{\varepsilon} [\Delta Y_t^\varepsilon + B_2(X_t^\varepsilon, Y_t^\varepsilon)] dt + \frac{1}{\sqrt{\varepsilon}} G_2(X_t^\varepsilon, Y_t^\varepsilon) dW_t^2, \\ X_0^\varepsilon = x \in H_1, Y_0^\varepsilon = y \in H_2, \end{cases} \quad (4.9)$$

where the coefficients  $F, G_1, B_2$  and  $G_2$  satisfy conditions (4.2)-(4.7).

It is well known that the operator  $A$  satisfies the monotonicity and coercivity properties (see, e.g., [21, Lemma 5.1.6]). So it is easy to check that all the conditions **A1-A4** and **B1-B4** hold. Hence, by Theorem 2.4, we have

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p} \right) = 0, \quad \forall p \geq 1,$$

where  $\bar{X}_t$  is the solution of the corresponding averaged equation.

**4.4. Stochastic 2D Navier-Stokes equation.** Now we consider the stochastic 2D Navier-Stokes equation. Let  $\Lambda$  be a bounded domain in  $\mathbb{R}^2$  with smooth boundary, denote

$$V_1 = \{v \in H_0^{1,2}(\Lambda; \mathbb{R}^2) : \nabla \cdot v = 0 \text{ a.s. in } \Lambda\}, \quad \|v\|_{V_1}^2 := \int_{\Lambda} |\nabla v(\xi)|^2 d\xi,$$

and let  $H_1$  be the closure of  $V_1$  in  $L^2(\Lambda; \mathbb{R}^2)$ . We choose the Gelfand triple for the slow equation

$$V_1 \subseteq H_1 \subseteq V_1^*$$

and the Gelfand triple for the fast equation

$$V_2 := H_0^{1,2}(\Lambda) \subseteq H_2 := L^2(\Lambda) \subseteq V_2^* := (H_0^{1,2}(\Lambda))^*$$

and let  $A : V_1 \rightarrow V_1^*$  be

$$A(u) := P_H \Delta u - P_H [(u \cdot \nabla)u], \quad u \in V_1,$$

where  $P_H$  is the Helmholtz-Leray projection and  $u \cdot \nabla = \sum_{i=1}^2 u^i \partial_i$  with  $u = (u^1, u^2)$ .

Now, we consider the slow-fast stochastic 2D Navier-Stokes-heat equation

$$\begin{cases} dX_t^\varepsilon = [A(X_t^\varepsilon) + F(X_t^\varepsilon, Y_t^\varepsilon)] dt + G_1(X_t^\varepsilon) dW_t^1, \\ dY_t^\varepsilon = \frac{1}{\varepsilon} [\Delta Y_t^\varepsilon + B_2(X_t^\varepsilon, Y_t^\varepsilon)] dt + \frac{1}{\sqrt{\varepsilon}} G_2(X_t^\varepsilon, Y_t^\varepsilon) dW_t^2, \\ X_0^\varepsilon = x \in H_1, Y_0^\varepsilon = y \in H_2, \end{cases} \quad (4.10)$$

where the coefficients  $F, G_1, B_2$  and  $G_2$  satisfy conditions (4.2)-(4.7).

It is well known that the operator  $A$  satisfies the monotonicity and coercivity properties (see, e.g., [21, Example 5.1.10]). So it is easy to check that all the conditions **A1-A4** and **B1-B4** hold. Hence, by Theorem 2.4, we have

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left( \sup_{t \in [0, T]} \|X_t^\varepsilon - \bar{X}_t\|_{H_1}^{2p} \right) = 0, \quad \forall p \geq 1,$$

where  $\bar{X}_t$  is the solution of the corresponding averaged equation.

## 5. APPENDIX

At the end of this section, we give the proof of Theorem 2.3 based on the techniques used in [21, Theorem 5.1.3].

**Proof of Theorem 2.3:** Let  $\mathcal{H} := H_1 \times H_2$  be the product Hilbert space. For any  $\phi = (\phi_1, \phi_2), \varphi = (\varphi_1, \varphi_2) \in \mathcal{H}$ , we denote the scalar product and the induced norm by

$$\langle \phi, \varphi \rangle_{\mathcal{H}} = \langle \phi_1, \varphi_1 \rangle_{H_1} + \langle \phi_2, \varphi_2 \rangle_{H_2}, \quad \|\phi\|_{\mathcal{H}} = \sqrt{\langle \phi, \phi \rangle_{\mathcal{H}}} = \sqrt{\|\phi_1\|_{H_1}^2 + \|\phi_2\|_{H_2}^2}.$$

Similarly, we also define  $\mathcal{V} := V_1 \times V_2$ . Then  $\mathcal{V}$  is a reflexive Banach space with the following norm:

$$\|\phi\|_{\mathcal{V}} = \sqrt{\|\phi_1\|_{V_1}^2 + \|\phi_2\|_{V_2}^2}.$$

Now we rewrite the system (1.1) for  $Z_t^\varepsilon = (X_t^\varepsilon, Y_t^\varepsilon)$  as

$$dZ_t^\varepsilon = \tilde{A}(Z_t^\varepsilon) dt + G(Z_t^\varepsilon) dW_t, \quad Z_0^\varepsilon = (x, y) \in \mathcal{H}, \quad (5.1)$$

where  $W_t := (W_t^1, W_t^2)$ , which is a  $U_1 \times U_2$ -valued cylindrical-Wiener process and

$$\tilde{A}(Z_t^\varepsilon) = \left( A(X_t^\varepsilon) + F(X_t^\varepsilon, Y_t^\varepsilon), \frac{1}{\varepsilon} B(X_t^\varepsilon, Y_t^\varepsilon) \right),$$

$$G(Z_t^\varepsilon) = \left( G_1(X_t^\varepsilon), \frac{1}{\sqrt{\varepsilon}} G_2(X_t^\varepsilon, Y_t^\varepsilon) \right).$$

Moreover,  $G$  is an operator from  $\mathcal{V}$  to  $L_2(\mathcal{U}, \mathcal{H})$ , where  $\mathcal{U} := U_1 \times U_2$  and  $L_2(\mathcal{U}, \mathcal{H})$  is the space of Hilbert-Schmidt operators from  $\mathcal{U}$  to  $\mathcal{H}$ . The norm in  $L_2(\mathcal{U}, \mathcal{H})$  is defined by

$$\|G(z)\|_{L_2(\mathcal{U}, \mathcal{H})} = \sqrt{\|G_1(x)\|_{L_2(U_1, H_1)}^2 + \frac{1}{\varepsilon} \|G_2(x, y)\|_{L_2(U_2, H_2)}^2}, \quad z = (x, y) \in \mathcal{V}.$$

Let  $\mathcal{V}^*$  be the dual space of  $\mathcal{V}$  and we consider the following Gelfand triple  $\mathcal{V} \subset \mathcal{H} \equiv \mathcal{H}^* \subset \mathcal{V}^*$ . It is easy to see that the following mappings

$$\tilde{A} : \mathcal{V} \rightarrow \mathcal{V}^*, \quad G : \mathcal{V} \rightarrow L_2(\mathcal{U}, \mathcal{H})$$

are well defined. To complete the proof, we only check whether the new coefficients in equation (5.1) satisfy the local monotonicity, coercivity and growth properties by [21, Theorem 5.1.3].

Indeed, for any  $w_1 = (u_1, v_1), w_2 = (u_2, v_2) \in \mathcal{V}$ , by conditions **A2** and **B2**, we have

$$\begin{aligned} & 2_{\mathcal{V}^*} \langle \tilde{A}(w_1) - \tilde{A}(w_2), w_1 - w_2 \rangle_{\mathcal{V}} + \|G(w_1) - G(w_2)\|_{L_2(\mathcal{U}, \mathcal{H})}^2 \\ &= 2_{V_1^*} \langle A(u_1) - A(u_2), u_1 - u_2 \rangle_{V_1} + \langle F(u_1, v_1) - F(u_2, v_2), u_1 - u_2 \rangle_{H_1} \\ & \quad + \frac{2}{\varepsilon} {}_{V_2^*} \langle B(u_1, v_1) - B(u_2, v_2), v_1 - v_2 \rangle_{V_2} + \|G_1(u_1) - G_1(u_2)\|_{L_2(U_1, H_1)}^2 \\ & \quad + \frac{1}{\varepsilon} \|G_2(u_1, v_1) - G_2(u_2, v_2)\|_{L_2(U_2, H_2)}^2 \\ & \leq C \left[ (1 + \|u_2\|_{V_1}^\alpha)(1 + \|u_2\|_{H_1}^\beta) \right] \|u_1 - u_2\|_{H_1}^2 - \frac{\gamma}{2\varepsilon} \|v_1 - v_2\|_{H_2}^2 \\ & \quad + C_\varepsilon \|u_1 - u_2\|_{H_1}^2 + C \|u_1 - u_2\|_{H_1} \|v_1 - v_2\|_{H_2} \\ & \leq C_\varepsilon \left[ (1 + \|w_2\|_{\mathcal{V}}^\alpha)(1 + \|w_2\|_{\mathcal{H}}^\beta) \right] \|w_1 - w_2\|_{\mathcal{H}}^2, \end{aligned}$$

which implies that the local monotonicity condition holds.

For any  $w = (u, v) \in \mathcal{V}$ , there exist constants  $C_\varepsilon > 0$  and  $C > 0$  such that

$$\begin{aligned} 2_{\mathcal{V}^*} \langle \tilde{A}(w), w \rangle_{\mathcal{V}} + \|G(w)\|_{L_2(\mathcal{U}, \mathcal{H})}^2 &= 2_{V_1^*} \langle A(u), u \rangle_{V_1} + \frac{2}{\varepsilon} {}_{V_2^*} \langle B(u, v), v \rangle_{V_2} + \langle F(u, v), u \rangle_{H_1} \\ & \quad + \|G_1(u)\|_{L_2(U_1, H_1)}^2 + \frac{1}{\varepsilon} \|G_2(u, v)\|_{L_2(U_2, H_2)}^2 \\ & \leq C \|u\|_{H_1}^2 - \theta \|u\|_{V_1}^\alpha + C \\ & \quad + \frac{1}{\varepsilon} \left[ C(1 + \|u\|_{H_1}^2 + \|v\|_{H_2}^2) - \eta \|v\|_{V_2}^\kappa \right] \\ & \leq C_\varepsilon (1 + \|w\|_{\mathcal{H}}^2) - C_\varepsilon (\|u\|_{V_1}^\alpha + \|v\|_{V_2}^\kappa) \end{aligned}$$

and

$$\begin{aligned} \|\tilde{A}(w)\|_{\mathcal{V}^*} &\leq \|A(u)\|_{V_1^*} + \frac{1}{\varepsilon} \|B(u, v)\|_{V_2^*} + \|F(u, v)\|_{H_1} \\ &\leq C(1 + \|u\|_{V_1}^{\alpha-1})(1 + \|u\|_{H_1}^{\frac{\beta(\alpha-1)}{\alpha}}) + \frac{C}{\varepsilon} (1 + \|u\|_{H_1} + \|v\|_{H_2} + \|u\|_{H_1}^{\frac{2(\kappa-1)}{\kappa}}) \\ & \quad + \frac{C}{\varepsilon} (1 + \|v\|_{V_2}^{\kappa-1}) \\ &\leq C_\varepsilon (1 + \|u\|_{V_1}^{\alpha-1} + \|v\|_{V_2}^{\kappa-1})(1 + \|w\|_{\mathcal{H}}^{\tilde{\beta}}), \end{aligned}$$

for some  $\tilde{\beta} > 0$ , which implies that the coercivity and growth conditions hold.  $\square$

**Acknowledgement.** The authors would like to thank the referees for some valuable comments. This work is supported by the National Natural Science Foundation of China (No.11931004,12171208,11831014,12271219,12090011), the QingLan Project of Jiangsu Province and the PAPD Project of Jiangsu Higher Education Institutions. Financial support of the DFG through CRC 1283 is also gratefully acknowledged.

## REFERENCES

- [1] R. Bertram, J.E. Rubin, *Multi-timescale systems and fast-slow analysis*, *Math. Biosci.* 287 (2017) 105-121.
- [2] N.N. Bogoliubov, Y.A. Mitropolsky, *Asymptotic methods in the theory of Non-linear Oscillations*, Gordon and Breach Science Publishers, New York (1961).
- [3] C.E. Bréhier, *Strong and weak orders in averaging for SPDEs*, *Stochastic Process. Appl.* 122 (2012) 2553-2593.
- [4] S. Cerrai, *A Khasminskii type averaging principle for stochastic reaction-diffusion equations*, *Ann. Appl. Probab.* 19 (2009) 899-948.
- [5] S. Cerrai, *Averaging principle for systems of reaction-diffusion equations with polynomial nonlinearities perturbed by multiplicative noise*, *SIAM J. Math. Anal.* 43 (2011) 2482-2518.
- [6] S. Cerrai, M. Freidlin, *Averaging principle for stochastic reaction-diffusion equations*, *Probab.Theory Related Fields* 144 (2009) 137-177.
- [7] S. Cerrai, A. Lunardi, *Averaging principle for nonautonomous slow-fast systems of stochastic reaction-diffusion equations: the almost periodic case*, *SIAM J. Math. Anal.* 49 (2017) 2843-2884.
- [8] I. Chueshov, A. Millet, *Stochastic 2D hydrodynamical type systems: well posedness and large deviations*, *Appl. Math. Optim.* 61, 379-420 (2010)
- [9] Z. Dong, X. Sun, H. Xiao, J. Zhai, *Averaging principle for one dimensional stochastic Burgers equation*, *J. Differential Equations* 265 (2018) 4749-4797.
- [10] W. E, B. Engquist, *Multiscale modeling and computations*, *Notice of AMS*, 50 (2003) 1062-1070.
- [11] W. E, D. Liu, E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, *Comm. Pure Appl. Math* 58 (2005) 1544-1585.
- [12] H. Fu, J. Liu, *Strong convergence in stochastic averaging principle for two time-scales stochastic partial differential equations*, *J. Math. Anal. Appl.* 384 (2011) 70-86.
- [13] H. Fu, L. Wan, J. Liu, *Strong convergence in averaging principle for stochastic hyperbolic-parabolic equations with two time-scales*, *Stochastic Process. Appl.* 125 (2015) 3255-3279.
- [14] P. Gao, *Averaging principle for stochastic Kuramoto-Sivashinsky equation with a fast oscillation*, *Discrete Contin. Dyn. Syst.-A* 38 (2018), 5649-5684.
- [15] P. Gao, *Averaging principle for the higher order nonlinear Schrödinger equation with a random fast oscillation*, *J. Stat. Phys.* 171 (2018), 897-926.
- [16] P. Gao, *Averaging Principle for Multiscale Stochastic Klein-Gordon-Heat System*, *J. Nonlinear Sci.* 29 (4) (2019), 1701-1759.
- [17] E. Harvey, V. Kirk, M. Wechselberger, J. Sneyd, *Multiple timescales, mixed mode oscillations and canards in models of intracellular calcium dynamics*, *J. Nonlinear Sci.* 21 (2011) 639-683.
- [18] R.Z. Khasminskii, *On the principle of averaging the Itô stochastic differential equations*, *Kibernetika* 4 (1968), 260-279.
- [19] S. Li, X. Sun, Y. Xie, Y. Zhao, *Averaging principle for two dimensional stochastic Navier-Stokes equations*, <https://arxiv.org/abs/1810.02282>.
- [20] W. Liu, M. Röckner, X. Sun, Y. Xie, *Averaging principle for slow-fast stochastic differential equations with time dependent locally Lipschitz coefficients*, *J. Differential Equations* 268 (6) (2020), 2910-2948.
- [21] W. Liu, M. Röckner, *Stochastic Partial Differential Equations: An Introduction*, Universitext, Springer, 2015.
- [22] E.A. Mastny, E.L. Haseltine, J.B. Rawlings, *Two classes of quasi-steady-state model reductions for stochastic kinetics*, *J. Chem. Phys* 127 (2007) 094106.
- [23] B. Øksendal, *Stochastic differential equations. An introduction with application*, Universitext. Springer-Verlag, Berlin, 1995.

- [24] W. Wang, A.J. Roberts, *Average and deviation for slow-fast stochastic partial differential equations*, *J. Differential Equations* 253 (2012) 1265-1286.
- [25] W. Wang, A.J. Roberts, J. Duan, *Large deviations and approximations for slow-fast stochastic reaction-diffusion equations*, *J. Differential Equations* 253 (2012) 3501-3522.

*Current address*, Liu, W.: School of Mathematics and Statistics/RIMS, Jiangsu Normal University, Xuzhou, 221116, China

*E-mail address*: weiliu@jsnu.edu.cn

*Current address*, Röckner, M.: Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany, and Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), Beijing, 100190, China

*E-mail address*: roeckner@math.uni-bielefeld.de

*Current address*, Sun, X.: School of Mathematics and Statistics/RIMS, Jiangsu Normal University, Xuzhou, 221116, China

*E-mail address*: xbsun@jsnu.edu.cn

*Current address*, Xie, Y.: School of Mathematics and Statistics/RIMS, Jiangsu Normal University, Xuzhou, 221116, China

*E-mail address*: ycxie@jsnu.edu.cn