

SECOND-ORDER FAST-SLOW STOCHASTIC SYSTEMS*

NHU N. NGUYEN[†] AND GEORGE YIN[‡]

Abstract. This paper focuses on systems of nonlinear second-order stochastic differential equations with multiscales. The motivation for our study stems from mathematical physics and statistical mechanics, for example, Langevin dynamics and stochastic acceleration in a random environment. Our aim is to carry out asymptotic analysis to establish large deviations principles. Our focus is on obtaining the desired results for systems under weaker conditions. When the fast-varying process is a diffusion, neither Lipschitz continuity nor linear growth needs to be assumed. Our approach is based on combinations of the intuition from Smoluchowski–Kramers approximation and the methods initiated in [A. A. Puhalskii, *Ann. Probab.*, 44 (2016), pp. 3111–3186] relying on the concepts of relatively large deviations compactness and the identification of rate functions. When the fast-varying process is under a general setup with no specified structure, the paper establishes the large deviations principle of the underlying system under the assumption on the local large deviations principles of the corresponding first-order system.

Key words. second-order stochastic differential equation, random environment, large deviations, local large deviations, averaging principle

MSC codes. 34E15, 34F05, 60F05, 60F10, 60J60

DOI. 10.1137/23M1567382

1. Introduction. In recent years, much effort has been devoted to analyzing stochastic systems arising from a wide variety of fields. For example, an averaging principle for complex Ginzburg–Landau equations was studied by Gao [17], homogenization in ergodic media was treated in Chen et al. [3], homogenization of stochastic convection-diffusion equations was studied in Bessaih, Efendiev, and Maris [1], mean field limits of particle-based stochastic systems were obtained in Isaacson, Ma, and Spiliopoulos [21], and Freidlin–Wentzell type large deviations results were obtained for multiscale stochastic partial differential equations in Hong, Li, and Liu [20]. One of the salient features in many applications is time-scale separation. For example, diffusions with fast and slow motions were treated in Khasminskii and Yin [24]. Although the first-order stochastic differential equations (SDEs) have been analyzed extensively, properties associated with the second-order SDEs are less well known. In applications, for example, in numerous systems in mathematical physics and statistical mechanics, such equations naturally arise; see, for example, the work of Kesten and Papanicolaou in [22, 23].

Because of this need, this paper is devoted to fully nonlinear second-order stochastic systems. We begin with the study of a class of second-order SDEs

*Received by the editors April 20, 2023; accepted for publication (in revised form) March 29, 2024; published electronically July 18, 2024.

<https://doi.org/10.1137/23M1567382>

Funding: The work of the first author was partially supported by an AMS-Simons travel grant. The work of the second author was partially supported by National Science Foundation grant DMS-2204240.

[†]Department of Mathematics and Applied Mathematical Sciences, University of Rhode Island, Kingston, RI 02881 USA (nhu.nguyen@uri.edu).

[‡]Department of Mathematics, University of Connecticut, Storrs, CT 06269 USA (gyin@uconn.edu).

$$(1.1) \quad \begin{cases} \varepsilon^2 \ddot{X}_t^\varepsilon = F_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) - \lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) \dot{X}_t^\varepsilon, & X_0^\varepsilon = x_0^\varepsilon \in \mathbb{R}^d, \quad \dot{X}_0^\varepsilon = x_1^\varepsilon \in \mathbb{R}^d, \\ \dot{Y}_t^\varepsilon = \frac{1}{\varepsilon} b_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) + \frac{1}{\sqrt{\varepsilon}} \sigma_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) \dot{W}_t, & Y_0^\varepsilon = y_0^\varepsilon \in \mathbb{R}^l, \end{cases}$$

where $\varepsilon > 0$ is a small parameter. Equation (1.1) is a multiscale and fully nonlinear system. In the above, for each $\varepsilon > 0$, $F_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}^d$, $\lambda_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}$, $b_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}^l$, $\sigma_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}^{l \times m}$ are measurable functions of their arguments (t, x, y) , and W_t is an m -dimensional vector-valued standard Brownian motion with $\dot{W}_t$ being its formal derivative. For each $\varepsilon > 0$, the weak solution of (1.1) is defined in the usual way, i.e., there exist a suitable probability space and an adapted Brownian motion W_t such that there are adapted processes $(X^\varepsilon, Y^\varepsilon)$ satisfying the system of stochastic integral equations corresponding to (1.1) almost surely.

Equations given in (1.1) may be considered as a singular perturbation problem with multiple time scales. Intuitively, as $\varepsilon^2 \rightarrow 0$ in the first equation of (1.1), it can be approximated by a first-order equation, whereas Y^ε in the second equation of (1.1) can be viewed as a fast-varying process, as it will be referred to in what follows. Further heuristic reasoning can be found in the beginning of section 2.1.1.

Our main effort is devoted to obtaining asymptotic properties of the underlying systems. Under mild conditions, we establish the large deviations principle (LDP) for the family of coupled processes $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ with μ^ε being the occupation measures of the fast-varying process Y^ε . Neither Lipschitz continuity nor a growth condition of $F^\varepsilon, b^\varepsilon, \sigma^\varepsilon$ is assumed. From the LDP of such couples, we can obtain averaging and LDPs for $\{X^\varepsilon\}_{\varepsilon>0}$. In addition, continuing our investigation, in this paper, we further reveal asymptotic properties without assuming specific structure of the fast process. In lieu of (1.1), we consider

$$(1.2) \quad \varepsilon^2 \ddot{X}_t^\varepsilon = F_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) - \lambda_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) \dot{X}_t^\varepsilon,$$

with ξ_t^ε being a process without a specified structure. We refer to ξ_t^ε as a fast-varying process, as we do for Y_t^ε in the previous paragraph.

Why do we care about the second-order stochastic systems? This is because numerous problems in physics, statistical mechanics, engineering, etc., involve such systems. In fact, in the study of ODEs, we encountered many second-order equations, including Airy's equations, Duffing equations, Liénard equations, Rayleigh's equations, etc. They have been used in a wide variety of applications. Adding stochastic perturbations to these equations leads to second-order SDEs of various kinds.

To further illustrate, consider the motions of a net of particles in a net of random force fields, described by Newton's law as $\ddot{x}_\varepsilon(t) = \tilde{F}_\varepsilon(t, \omega, x_\varepsilon(t), \dot{x}_\varepsilon(t), \chi_\varepsilon(t))$, where $x_\varepsilon(t)$ denotes the location of the particles at time t . The $\tilde{F}_\varepsilon$ denotes the random force fields depending on time t , sample point ω , the particle's locations x_ε , the particle's velocities $\dot{x}_\varepsilon$, and the random environments $\chi_\varepsilon(t)$ interacting with the system. To begin, turbulent diffusions and stochastic accelerations were considered by Kesten and Papanicolaou in [22, 23] under suitable conditions. Here we focus on the motions of particles, in which the Reynolds number (see e.g., [35] for a definition) is very small so that inertial effects are negligible compared to the damping force by assuming that

$$\tilde{F}_\varepsilon(t, \omega, x_\varepsilon(t), \dot{x}_\varepsilon(t), \chi_\varepsilon(t)) = F_\varepsilon(t, x_\varepsilon(t), \chi_\varepsilon(t)) - \frac{\lambda_\varepsilon(t, x_\varepsilon(t), \chi_\varepsilon(t))}{\varepsilon} \dot{x}_\varepsilon(t).$$

Now, by scaling $X_t^\varepsilon := x_\varepsilon(t/\varepsilon)$ and $\xi_t^\varepsilon := \chi_\varepsilon(t/\varepsilon)$, the system can be rewritten as (1.2). One of the examples of $\chi_\varepsilon(t)$ is a diffusion process. In this case, ξ_t^ε is a fast diffusion

process that is fully coupled with the system, which leads to the system of equations in (1.1). Another motivation is from the averaging and large deviations principles for systems of SDEs. System (1.1) can be viewed as the second-order version of the problem considered in [27] and references therein. It should be mentioned that there has been much recent interest in studying stochastic second-order systems in random environments. For example, the work [39] studied stochastic Hamiltonian systems living in random environments with the random environment represented by a random switching process.

Because ε is a small parameter, as ε is getting smaller and smaller, we expect the system to display certain limit behavior, in which the averaging principle plays an important role in studying heterogeneity that often occurs in physics as well as in biology, economics, queuing theory, and game theory, among others; see, e.g., [14]. Typically, analyzing and simulating heterogeneous models is much more challenging than the corresponding homogeneous models, in which the heterogeneous property is replaced by its average value. The averaging principle for a system guarantees the validity of this replacement. On the other hand, the LDPs (see [11, 10]), characterizing quantitatively the rare events, play an important role in many areas with a wide range of applications. To mention just a few, they include equilibrium and nonequilibrium statistical mechanics, multifractals, and thermodynamics of chaotic systems [36]. By establishing the LDPs for system (1.1) and (1.2), we provide an insight into the motions of (small) particles in random force fields, which is heterogeneous. Further, the heterogeneity is allowed to interact with the system. Not only will this work illustrate that averaging of the heterogeneity works in this case, but it will also provide the picture of the dynamics around the averaged system.

From the point of view of development of homogenization and large deviations, much effort has been devoted to studying averaging and LDPs of the first-order differential equations under a random environment (given by diffusion process, switching process, wideband noise, etc.) in the setting of fast-slow systems. Such problems have been addressed in [18, 25, 26, 40, 41, 42, 43] under certain settings in which the fast process is often not fully coupled with the slow system. Very recently, the question for the fully coupled system was addressed in [34]. Some other related studies can be found in [2, 19, 27]. Reference [26] considered systems under wideband noise; [33] studied systems under rough path noise; [7, 8, 9] investigated systems in infinite dimensional settings. In contrast to the systems considered in the aforementioned works with emphases on first-order equations, we consider systems of second-order differential equations of the forms (1.1) and (1.2). From a statistical physics point of view, there were some works treating the stochastic accelerations and the Langevin equations such as [4, 15, 40] for the study of Smoluchowski–Kramers approximation, [6] for the LDPs, [5] for the MDPs (moderate deviations principles) in the absence of the random environment, and [31, 32] for the LDPs of Langevin systems with random environment under certain specific settings. To the best of our knowledge, this paper is one of the first works to address the problem of averaging and large deviations principles for second-order equations in a random environment that are fully coupled. We establish the LDPs under mild and natural conditions.

To establish the desired LDPs for system (1.1), in light of the work of [27, 34] on the first-order SDEs, we first establish the LDP for the family of coupled processes $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$, where μ^ε is defined as a random occupation measure of Y^ε . Then, the LDPs for the families $\{X^\varepsilon\}_{\varepsilon>0}$ and/or $\{\mu^\varepsilon\}_{\varepsilon>0}$ can be handled by some standard projection techniques in the large deviations theory. Without assuming any regularity of F^ε , b^ε , and σ^ε , we could not establish a “good” connection between the solution of

the second-order equations and the corresponding first-order equations. Our approach is based on a combination of the approach of Puhalskii in [34] for the first-order coupled system (namely, obtaining the relatively large deviations compactness and then carefully identifying the rate functions) and the intuition of Smoluchowski–Kramers approximation. To establish LDPs of SDEs, the weak convergence methods initiated by Dupuis and Ellis [12] have been used by many authors (see e.g., [2, 6, 12, 26] and references therein), and they have been shown to be effective in proving the LDPs for many systems. However, using the weak convergence approach for our problem may require stronger assumptions such as Lipschitz continuity of coefficients in the equation of X^ε (as shown in, e.g., [2, Assumption 2.1] and [6, Hypothesis 1]). Such conditions are needed in Budhiraja, Dupuis, and Ganguly [2] because of the need to prove the lower bound [2, eq. (2.13)], in which some uniqueness properties of auxiliary optimal controls are required. The paper [2] studied the first-order SDEs with a fast-varying jump process; the aforementioned difficulty arises in [2] due to the presence of multiple time scales rather than the presence of the jump process. Here, we are dealing with fully nonlinear second-order stochastic systems with multiscales, but we do not use the weak convergence method to avoid requiring the Lipschitz continuity and other growth conditions. In [6], Cerrai and Freidlin considered the second-order SDEs without coupling with another fast-varying process. To establish the desired convergence, Lipschitz continuities of coefficients in the system are necessary. In [13], Feng and Kurtz introduced the HJB equations/viscosity solutions approach. In [13, section 11.6], first-order SDEs are considered, and conditions for the validity of LDPs are derived. However, these conditions rely on the existence of functions possessing certain properties, which are often difficult to verify in terms of the coefficients. Although Feng and Kurtz were able to provide explicit conditions on the coefficients, a key requirement is that $\sigma_t^\varepsilon(x, y)$ be independent of x (see [13, Lemma 11.60, p. 278]). As will be seen later, we do not need the Lipschitz continuity for (1.1), nor do we need $\sigma_t^\varepsilon(x, y)$ to be independent of x as in [13]. In this paper, we manage to establish LDPs of X^ε in multiscale and fully coupled system (1.1) under mild conditions, which is another of our goals.

To establish the desired LDP for the system under the general fast random process (1.2), we have to use a different approach. We assume that the corresponding first-order equation satisfies the local LDP, which will be shown to be verifiable and satisfied in many problems. To prove the LDP, we show that the family of $\{X^\varepsilon\}_{\varepsilon>0}$ is exponentially tight and satisfies the local LDP.

The rest of the paper is arranged as follows. We divide the presentation into two parts. The first part, section 2, is devoted to the second-order systems with a fast-varying diffusion (1.1). Section 2.1 formulates the problem and states the results. The detailed proof of results is provided in section 2.2. The second part of the paper, presented in section 3, substantially extends the results to second-order equations with general fast-varying random processes (1.2). The formulation, conditions, results, and detailed proofs are presented. Finally, section 4 presents two examples to illustrate our formulation and results.

2. Second-order fast-slow systems with fast diffusion.

2.1. Notation, formulation, and results. Throughout the paper, $|\cdot|$ denotes a Euclid norm while $\|\cdot\|$ indicates the operator sup-norm, $\mathcal{C}(\mathcal{X}, \mathcal{Y})$ is the space of continuous functions from $\mathcal{X}$ to $\mathcal{Y}$, and if $\mathcal{Y}$ is a Euclid space, we write $\mathcal{C}(\mathcal{X}, \mathcal{Y})$ as $\mathcal{C}(\mathcal{X})$ for simplicity. Let $\mathcal{M}(\mathbb{R}^l)$ be the set of finite measures on $\mathbb{R}^l$ endowed with the weak topology, let $\mathcal{P}(\mathbb{R}^l)$ be the set of probability densities $m(y)$ on $\mathbb{R}^l$ such that

$m \in \mathbb{W}_{\text{loc}}^{1,1}(\mathbb{R}^d)$ and $\sqrt{m} \in \mathbb{W}^{1,2}(\mathbb{R}^l)$, where $\mathbb{W}^{1,2}(\mathbb{R}^l)$ (resp., $\mathbb{W}_{\text{loc}}^{1,1}(\mathbb{R}^d)$) is the Sobolev space (resp., local Sobolev space) with suitable exponents, and let $\mathcal{C}_0^1(\mathbb{R}^l)$ be the space of continuously differentiable functions with compact supports in $\mathbb{R}^l$. Let $\mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ represent the subset of $\mathcal{C}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ of functions $\mu = (\mu_t, t \in \mathbb{R}_+)$ such that $\mu_t - \mu_s$ is an element of $\mathcal{M}(\mathbb{R}^l)$ for $t \geq s$ and $\mu_t(\mathbb{R}^l) = t$. It is endowed with the subspace topology and is a complete and separable metric space, being closed in $\mathcal{C}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$. We define the random process $\mu^\varepsilon = (\mu_t^\varepsilon, t \in \mathbb{R}_+)$ of the fast process Y^ε by

$$(2.1) \quad \mu_t^\varepsilon(\mathcal{A}) := \int_0^t \mathbf{1}_{\mathcal{A}}(Y_s^\varepsilon) ds \quad \forall \mathcal{A} \in \mathcal{B}(\mathbb{R}^l).$$

Then, μ^ε is a random element of $\mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ and we can regard $(X^\varepsilon, \mu^\varepsilon)$ as a random element of $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$. Note that the elements of $\mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ can be regarded as a σ -finite measure on $\mathbb{R}_+ \times \mathbb{R}^l$. As a result, we use the notation $\mu(dt, dy)$ for $\mu \in \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$. For a symmetric positive definite matrix A and matrix z of suitable dimensions, we define $\|z\|_A := z^\top A z$. Following Puhalskii's notation, $\|z\|_A$ can be either matrices or numbers, depending on the dimension z . We also use ∇_x , ∇_{xx} , div_x to denote the gradient, the Hessian, and the divergence, respectively, with respect to indicated variables. It should be clear from the context.

We will establish the LDP and describe explicitly the rate function for the family $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$. The LDP and the rate function of $\{X^\varepsilon\}_{\varepsilon>0}$ are obtained directly by standard projections in the large deviations theory. To proceed, we recall briefly the basic definitions of the LDP. For further references, see [11, 10, 28].

DEFINITION 2.1. *We say the family of $\{\mathbb{P}^\varepsilon\}_{\varepsilon>0}$ in some metric space $\mathbb{S}$ enjoys the LDP with a rate function $\mathbb{I}$ if the following conditions are satisfied: (1) $\mathbb{I} : \mathbb{S} \rightarrow [0, \infty]$ is inf-compact, that is, the level sets $\{z \in \mathbb{S} : \mathbb{I}(z) \leq L\}$ are compact in $\mathbb{S}$ for any $L > 0$; (2) for any open subset G of $\mathbb{S}$,*

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}^\varepsilon(G) \geq -\mathbb{I}(G) := -\inf_{z \in G} \mathbb{I}(f);$$

and (3) for any closed subset F of $\mathbb{S}$,

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}^\varepsilon(F) \leq -\mathbb{I}(F) := -\inf_{z \in F} \mathbb{I}(f).$$

We say that a family of random elements of $\mathbb{S}$ obeys the LDP if the family of their laws obeys the LDP.

Our main effort in this section is to consider system (1.1) and to establish the LDP for the family of the processes $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ with μ^ε being the empirical process associated with Y^ε as in (2.1), where $(X^\varepsilon, Y^\varepsilon)$ is a solution of the second-order differential equation with random environment given in (1.1). Such a solution is defined as follows.

One can rewrite (1.1) as

$$(2.2) \quad \begin{cases} \dot{X}_t^\varepsilon = p_t^\varepsilon, & X_0^\varepsilon = x_0^\varepsilon \in \mathbb{R}^d, \\ \varepsilon^2 \dot{p}_t^\varepsilon = F_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) - \lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) p_t^\varepsilon, & p_0^\varepsilon = x_1^\varepsilon \in \mathbb{R}^d, \\ \dot{Y}_t^\varepsilon = \frac{1}{\varepsilon} b_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) + \frac{1}{\sqrt{\varepsilon}} \sigma_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) \dot{W}_t, & Y_0^\varepsilon = y_0^\varepsilon \in \mathbb{R}^l. \end{cases}$$

Recall that for each $\varepsilon > 0$, the coefficients $F_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}^d$, $\lambda_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}$, $b_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}^l$, $\sigma_t^\varepsilon(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}^{l \times m}$ are functions

of (t, x, y) ; $x_0^\varepsilon, x_1^\varepsilon \in \mathbb{R}^d, y_0^\varepsilon \in \mathbb{R}^l$ are initial values that can be random. Throughout the paper, we assume that these functions are measurable and locally bounded in (t, x, y) such that the system of equations (2.2) admits a weak solution $(X^\varepsilon, p^\varepsilon, Y^\varepsilon)$ with trajectories in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^l)$ for every initial condition $(x_0^\varepsilon, x_1^\varepsilon, y_0^\varepsilon)$, and then the system of equations (1.1) admits a weak solution $(X^\varepsilon, Y^\varepsilon)$ with trajectories in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d \times \mathbb{R}^l)$. More precisely, we assume that there exist a complete probability space $(\Omega^\varepsilon, \mathcal{F}^\varepsilon, \mathbb{P}^\varepsilon)$ with filtration $\mathcal{F}^\varepsilon = (\mathcal{F}_t^\varepsilon, t \in \mathbb{R}_+)$, a Brownian motion $(W_t, t \in \mathbb{R}_+)$ with respect to $\mathcal{F}^\varepsilon$, and processes $X^\varepsilon = (X_t^\varepsilon, t \in \mathbb{R}_+)$, $p^\varepsilon = (p_t^\varepsilon, t \in \mathbb{R}_+)$, and $Y^\varepsilon = (Y_t^\varepsilon, t \in \mathbb{R}_+)$ that are $\mathcal{F}^\varepsilon$ -adapted and have continuous trajectories satisfying the following equations:

$$\begin{cases} X_t^\varepsilon = x_0^\varepsilon + \int_0^t p_s^\varepsilon ds, \\ p_t^\varepsilon = x_1^\varepsilon + \frac{1}{\varepsilon^2} \int_0^t \left(F_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) - \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) p_s^\varepsilon \right) ds, \\ Y_t^\varepsilon = y_0^\varepsilon + \frac{1}{\varepsilon} \int_0^t b_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^t \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s \end{cases}$$

for all $t \in \mathbb{R}_+$, $\mathbb{P}^\varepsilon$ -a.s. It is noted that it may not guarantee the uniqueness of the solution. (To ensure the uniqueness, one may need to require further that the coefficients are Lipschitz continuous, which we do not assume here.) Next, we need some conditions, which are mild and natural, to establish the LDP for the family of coupled processes $\{X^\varepsilon, \mu^\varepsilon\}_{\varepsilon > 0}$.

Assumption 2.1. Assume that for all $L > 0$ and $t > 0$,

$$(2.3) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{s \in [0, t]} \sup_{y \in \mathbb{R}^l} \sup_{x \in \mathbb{R}^d: |x| \leq L} [|F_s^\varepsilon(x, y)| + |\lambda_s^\varepsilon(x, y)| + |b_s^\varepsilon(x, y)| + \|\Sigma_s^\varepsilon(x, y)\|] < \infty,$$

where $\Sigma_t^\varepsilon(x, y) := \sigma_t^\varepsilon(x, y)[\sigma_t^\varepsilon(x, y)]^\top$,

$$(2.4) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{s \in [0, t]} \sup_{y \in \mathbb{R}^l} \sup_{x \in \mathbb{R}^d} \frac{x^\top F_s^\varepsilon(x, y)}{(1 + |x|^2)\lambda_s^\varepsilon(x, y)} < \infty,$$

$$(2.5) \quad \lim_{M \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \sup_{s \in [0, t]} \sup_{y \in \mathbb{R}^l, |y| \geq M} \sup_{x \in \mathbb{R}^n, |x| \leq L} \frac{[b_s^\varepsilon(x, y)]^\top y}{|y|} < 0,$$

$$(2.6) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{s \in [0, \infty), y \in \mathbb{R}^l, x \in \mathbb{R}^d} [|\nabla_s \lambda_s^\varepsilon(x, y)| + |\nabla_x \lambda_s^\varepsilon(x, y)| + |\nabla_y \lambda_s^\varepsilon(x, y)| + \|\nabla_{yy} \lambda_s^\varepsilon(x, y)\|] < \infty,$$

and

$$(2.7) \quad \liminf_{\varepsilon \rightarrow 0} \inf_{s \in [0, \infty), y \in \mathbb{R}^l, x \in \mathbb{R}^d} \lambda_s^\varepsilon(x, y) > \kappa_0 > 0.$$

Remark 2.1. The condition (2.3) consists of (locally in (t, x) and globally in y) boundedness conditions of $F^\varepsilon, b^\varepsilon$ and Σ^ε . Note that (2.4) is a growth-rate condition, which is milder than linear growth of $\frac{F_s^\varepsilon(x, y)}{\lambda_s^\varepsilon(x, y)}$; e.g., $\frac{F_t^\varepsilon(x, y)}{\lambda_t^\varepsilon(x, y)} = \frac{1}{x}$ satisfies this condition but is not linear growth. Moreover, it does not imply any growth-rate condition for $F_s^\varepsilon(x, y)$. The condition (2.5) is a stability condition, which in fact is needed for the ergodicity of the fast process. It is noted that we do not require any Lipschitz continuity and growth-rate conditions for these coefficients. Lower boundedness and

regularity conditions (2.6) and (2.7) of $\lambda_t^\varepsilon(x, y)$ are natural and often used in the literature of mathematical physics; see, e.g., [5, 6].

Assume that there are “limit” measurable functions $F_t(x, y)$, $\lambda_t(x, y)$, $b_t(x, y)$, and $\sigma_t(x, y)$ of the families of functions $F_t^\varepsilon(x, y)$, $\lambda_t^\varepsilon(x, y)$, $b_t^\varepsilon(x, y)$, $\sigma_t^\varepsilon(x, y)$ as $\varepsilon \rightarrow 0$, respectively, in the sense that for all $t > 0$ and $L > 0$,

$$(2.8) \quad \lim_{\varepsilon \rightarrow 0} \sup_{s \in [0, t]} \sup_{y \in \mathbb{R}^l, |y| \leq L} \sup_{x \in \mathbb{R}^d, |x| \leq L} \left[|F_s^\varepsilon(x, y) - F_s(x, y)| + |\lambda_s^\varepsilon(x, y) - \lambda_s(x, y)| \right. \\ \left. + |b_s^\varepsilon(x, y) - b_s(x, y)| + \|\sigma_s^\varepsilon(x, y) - \sigma_s(x, y)\| \right] = 0.$$

Assumption 2.2. Assume that the “limit” function $b_t(x, y)$ is Lipschitz continuous in y locally uniformly in (t, x) ; $b_t(x, y)$ and $\Sigma_t(x, y) := \sigma_t(x, y)[\sigma_t(x, y)]^\top$ are continuous in x locally uniformly in t and uniformly in y ; $\Sigma_t(x, y)$ is of class C^1 in y , with the first partial derivatives being bounded and Lipschitz continuous in y locally uniformly in (t, x) ; and $\operatorname{div}_y \Sigma_t(x, y)$ is continuous in (x, y) . The matrix $\Sigma_t(x, y)$ is positive definite uniformly in y and locally uniformly in (t, x) . In addition, the “limit” function $F_t(x, y)$ is locally Lipschitz continuous in x locally uniformly in t and uniformly in y . The conditions (2.6) and (2.7) hold for λ_t . Moreover, for all $t > 0$,

$$(2.9) \quad \lim_{|y| \rightarrow \infty} \sup_{s \in [0, t]} \sup_{x \in \mathbb{R}^d} \frac{[b_s(x, y)]^\top y}{|y|^2} < 0.$$

Rate function. Denote by $\mathcal{G}$ the collection of (φ, μ) such that the function $\varphi = (\varphi_t, t \in \mathbb{R}_+) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ is absolutely continuous (with respect to the Lebesgue measure on $\mathbb{R}_+$) and the function $\mu = (\mu_t, t \in \mathbb{R}_+) \in \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$, when considered as a measure on $\mathbb{R}_+ \times \mathbb{R}^l$, is absolutely continuous (with respect to the Lebesgue measure on $\mathbb{R}_+ \times \mathbb{R}^l$), i.e., $\mu(ds, dy) = m_s(y) dy ds$, and for almost all s , $m_s(y)$ (as a function of y) belongs to $\mathcal{P}(\mathbb{R}^l)$.

For $(\varphi, \mu) \in \mathcal{G}$, $\mu(ds, dy) = m_s(y) dy ds$, define

$$\mathbb{I}_1(\varphi, \mu) = \int_0^\infty \left[\sup_{\beta \in \mathbb{R}^d} \beta^\top \left(\dot{\varphi}_s - \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y)}{\lambda_s(\varphi_s, y)} m_s(y) dy \right) \right. \\ \left. + \sup_{h \in C_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \left([\nabla h(y)]^\top \left(\frac{1}{2} \operatorname{div}_y (\Sigma_s(\varphi_s, y) m_s(y)) - b_s(\varphi_s, y) m_s(y) \right) \right. \right. \\ \left. \left. - \frac{1}{2} \|\nabla h(y)\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) \right) dy \right] ds,$$

and define $\mathbb{I}_1(\varphi, \mu) = \infty$ if $(\varphi, \mu) \notin \mathcal{G}$.

DEFINITION 2.2. The family of random variables with distributions $\{\mathbb{P}^\varepsilon\}_{\varepsilon > 0}$ is said to be exponentially tight in the space $\mathbb{S}$ if there exists an increasing sequence of compact sets $(K_L)_{L \geq 1}$ of $\mathbb{S}$ such that $\lim_{L \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}^\varepsilon(K_L) = -\infty$.

THEOREM 2.1. Assume that Assumptions 2.1 and 2.2 hold, that the family of initial values $\{x_0^\varepsilon\}_{\varepsilon > 0}$ obeys the LDP in $\mathbb{R}^d$ with a rate function $\mathbb{I}_0$, that

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon |x_1^\varepsilon| < \infty \quad a.s.,$$

and that the family of initial values $\{y_0^\varepsilon\}_{\varepsilon > 0}$ is exponentially tight in $\mathbb{R}^l$. Then the family $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon > 0}$ obtained from (1.1) obeys the LDP in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ with rate function $\mathbb{I}$ defined as $\mathbb{I}(\varphi, \mu) = \mathbb{I}_0(\varphi_0) + \mathbb{I}_1(\varphi, \mu)$ if $(\varphi, \mu) \in \mathcal{G}$, and $\mathbb{I}(\varphi, \mu) = \infty$ otherwise.

COROLLARY 2.2. *Under the hypotheses of Theorem 2.1, the family $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies the LDP in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ with the rate function $\mathbb{I}_X$ defined by*

$$\mathbb{I}_X(\varphi) = \inf_{\mu \in \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))} \mathbb{I}(\varphi, \mu).$$

As an alternative representation, if function $\varphi = (\varphi_t, t \in \mathbb{R}_+) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}_+$, then

$$(2.10) \quad \begin{aligned} \mathbb{I}_X(\varphi) = & \mathbb{I}_0(\varphi_0) + \int_0^\infty \sup_{\beta \in \mathbb{R}^d} \left[\beta^\top \dot{\varphi}_s - \sup_{m \in \mathcal{P}(\mathbb{R}^l)} \left(\beta^\top \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y)}{\lambda_s(\varphi_s, y)} m(y) dy \right. \right. \\ & + \sup_{h \in \mathcal{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \left([\nabla h(y)]^\top \left(\frac{1}{2} \operatorname{div}_y (\Sigma_s(\varphi_s, y) m(y)) - b_s(\varphi_s, y) m(y) \right) \right. \\ & \left. \left. - \frac{1}{2} \|\nabla h(y)\|_{\Sigma_s(\varphi_s, y)}^2 m(y) \right) dy \right] ds; \end{aligned}$$

otherwise, $\mathbb{I}_X(\varphi) = \infty$.

2.1.1. Zero points of $\mathbb{I}(\varphi, \mu)$, averaging principle of (1.1), and its large deviations analysis. We start with an intuitive discussion on the behavior of (1.1) as $\varepsilon \rightarrow 0$. Intuitively, there are two phases as $\varepsilon \rightarrow 0$. First, ε^2 goes to 0. At this phase, X_t^ε is close to the solution of the following associated first-order equation (or the overdamped equation in the language of statistical physics):

$$(2.11) \quad \begin{cases} 0 = F_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}}) - \lambda_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}}) \dot{\bar{X}}_t^\varepsilon, & \bar{X}_0^\varepsilon = x_0^\varepsilon \in \mathbb{R}^d, \\ \dot{Y}_t^{\varepsilon, \bar{X}} = \frac{1}{\varepsilon} b_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}}) + \frac{1}{\sqrt{\varepsilon}} \sigma_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}}) \dot{W}_t, & Y_0^{\varepsilon, \bar{X}} = y_0^\varepsilon \in \mathbb{R}^l. \end{cases}$$

Next, $Y_t^{\varepsilon, \bar{X}}$ converges to its invariant distribution as $\varepsilon \rightarrow 0$. More precisely, if we let $\tilde{Y}_t^{\bar{X}} := Y_{t\varepsilon}^{\varepsilon, \bar{X}}$, then

$$\begin{cases} \dot{\bar{X}}_t^\varepsilon = \frac{F_t^\varepsilon(\bar{X}_t^\varepsilon, \tilde{Y}_{t/\varepsilon}^{\bar{X}})}{\lambda_t^\varepsilon(\bar{X}_t^\varepsilon, \tilde{Y}_{t/\varepsilon}^{\bar{X}})}, & \bar{X}_0^\varepsilon = x_0^\varepsilon \in \mathbb{R}^d, \\ \dot{\tilde{Y}}_t^{\bar{X}} = b_{t\varepsilon}^\varepsilon(\bar{X}_{t\varepsilon}^\varepsilon, \tilde{Y}_t^{\bar{X}}) + \sqrt{\varepsilon} \sigma_{t\varepsilon}^\varepsilon(\bar{X}_{t\varepsilon}^\varepsilon, \tilde{Y}_t^{\bar{X}}) \dot{\tilde{W}}_t, & \tilde{Y}_0^{\varepsilon, \bar{X}} = y_0^\varepsilon \in \mathbb{R}^l, \end{cases}$$

where $\tilde{W}_t$ is another standard Brownian motion. As a consequence, because $\tilde{Y}_{t/\varepsilon}^{\bar{X}}$ will come to and stay close to its invariant measure as $\varepsilon \rightarrow 0$, $\bar{X}_t^\varepsilon$ will tend to $\bar{X}_t$, the solution of the following differential equation:

$$\dot{\bar{X}}_t = \overline{F/\lambda}_t(\bar{X}_t), \quad \bar{X}_0 = \bar{x}_0,$$

where $\bar{x}_0$ is the limit of x_0^ε , and

$$\overline{F/\lambda}_{t_1}(x) := \int_{\mathbb{R}^l} \frac{F_{t_1}(x, y)}{\lambda_{t_1}(x, y)} \bar{\nu}^{t_1, x}(dy),$$

and for each fixed (t_1, x) , $\bar{\nu}^{t_1, x}(dy)$ is the invariant measure of the following SDE:

$$\dot{\tilde{Y}}_t = b_{t_1}(x, \tilde{Y}_t) + \sigma_{t_1}(x, \tilde{Y}_t) \dot{\tilde{W}}_t.$$

The convergence of X_t^ε to $\bar{X}_t$ as $\varepsilon \rightarrow 0$ forms an averaging principle of (1.1). However, we are interested in not only the convergence of X^ε to $\bar{X}_t$, but also the tail probability of this convergence, i.e., the rate of the convergence of the probability of the event $\{\|X^\varepsilon - \bar{X}\| > \eta\}$ to 0, for any $\eta > 0$. We show that the convergence is exponentially fast. The answer to these questions can be obtained from the LDP for $\{X_t^\varepsilon\}_{\varepsilon>0}$ and explicit representations of the rate function.

To proceed, we apply our results to make the above intuition rigorous. It is shown later that $\mathbb{I}_1(\varphi, \mu) = 0$ provided that a.e.

$$\dot{\varphi}_s = \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y)}{\lambda_s(\varphi_s, y)} m_s(y) dy,$$

$m_s(y)$ satisfies the equation

$$(2.12) \quad \int_{\mathbb{R}^l} \left(\frac{1}{2} \text{tr}(\Sigma_s(\varphi_s, y) [\nabla_{yy} h(y)]) + [\nabla_y h(y)]^\top b_s(\varphi_s, y) \right) m_s(y) dy = 0 \quad \forall h \in C_0^\infty(\mathbb{R}^l),$$

and $\mathbb{I}_0(\varphi_0) = 0$. Alternatively, $m_s(\cdot)$ is the invariant density of the diffusion process with the drift $b_s(\varphi_s, \cdot)$ and the diffusion matrix $\Sigma_s(\varphi_s, \cdot)$. Therefore, as $\varepsilon \rightarrow 0$, the trajectories of $\{X^\varepsilon\}_{\varepsilon>0}$ hover around $\bar{X}$ with exponential tail probability, where $\bar{X}$ is defined as the solution of the ODE

$$(2.13) \quad \dot{\bar{X}}_t = \overline{F/\lambda_t}(\bar{X}_t), \quad \bar{X}_0 = \bar{x}_0,$$

with

$$\overline{F/\lambda_t}(x) := \int_{\mathbb{R}^l} \frac{F_t(x, y)}{\lambda_t(x, y)} \bar{m}_t(y) dy,$$

and $\bar{m}_t(\cdot)$ satisfies (2.12) and $\bar{x}_0$ satisfies $\mathbb{I}_0(\bar{x}_0) = 0$. Letting

$$B_\eta^c(\bar{X}) := \left\{ \varphi \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) : \|\varphi_t - \bar{X}_t\|_{\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)} := \sum_{n=1}^{\infty} \frac{1}{2^n} \left(1 \wedge \sup_{t \leq n} |\varphi_t - \bar{X}_t| \right) \geq \eta \right\},$$

the LDP established in this paper implies that

$$\mathbb{P}^\varepsilon(X^\varepsilon \in B_\eta^c(\bar{X})) \sim e^{-\frac{1}{\varepsilon} \mathbb{I}_X(B_\eta^c(\bar{X}))},$$

where $\mathbb{I}_X(B_\eta^c(\bar{X})) = \inf_{\varphi \in B_\eta^c(\bar{X})} \mathbb{I}_X(\varphi)$. If we assume that $\bar{X}$ is the unique solution of (2.13), it is the unique solution of $\mathbb{I}_X(\varphi) = 0$. As a result, $\mathbb{I}_X(B_\eta^c(\bar{X})) > 0$. Indeed, if $\mathbb{I}_X(B_\eta^c(\bar{X})) = 0$, there exists $\{\varphi_k\}_{k=1}^\infty \subset B_\eta^c(\bar{X})$ such that $\lim_{k \rightarrow \infty} \mathbb{I}_X(\varphi_k) = 0$. Because $\mathbb{I}_X$ is a rate function, there exists a convergent subsequence (still denoted by φ_k) of $\{\varphi_k\}$ with limit denoted by $\bar{\varphi} \in B_\eta^c(\bar{X})$. Since $\mathbb{I}_X$ is lower semicontinuous, $0 \leq \mathbb{I}_X(\bar{\varphi}) = \mathbb{I}_X(\lim_{k \rightarrow \infty} \varphi_k) \leq \lim_{k \rightarrow \infty} \mathbb{I}_X(\varphi_k) = 0$. This leads to $\mathbb{I}_X(\bar{\varphi}) = 0$, which is a contradiction. Because $\mathbb{I}_X(B_\eta^c(\bar{X})) > 0$, $\mathbb{P}(\|X^\varepsilon - \bar{X}\| > \eta) \rightarrow 0$ exponentially fast for any $\eta > 0$.

Remark 2.2. In section 2.1.1, we illustrate that from our LDP result, we can establish the averaging principle of (1.1) with exponential convergence rate in the sense that X^ε converges to $\bar{X}$ (of (2.13)) with exponential tail probability, i.e., for any $\eta > 0$, $\mathbb{P}(\|X^\varepsilon - \bar{X}\| > \eta) \rightarrow 0$ exponentially fast. From a different angle, references [41, 42] treated convergence rate for averaging principles of different problems using certain moments. In this process, just as in treating L_2 or L_p convergence rates in numerical approximation of SDEs, global Lipschitz conditions are needed. For our second-order equations, the Lipschitz continuity need not be assumed. This is an advantage.

2.1.2. Alternative representations of $\mathbb{I}(\varphi, \mu)$. One can write the rate function $\mathbb{I}(\varphi, \mu)$ as

$$\mathbb{I}(\varphi, \mu) = \mathbb{I}_0(\varphi_0) + \int_0^\infty \left[\sup_{\beta \in \mathbb{R}^d} \beta^\top \left(\dot{\varphi}_s - \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y)}{\lambda_s(\varphi_s, y)} \nu_s(dy) \right) + \mathbb{J}_{s, \varphi_s}(\nu_s) \right] ds,$$

where $\nu_s(dy) = m_s(y)dy$ and

$$\begin{aligned} \mathbb{J}_{s, \varphi_s}(\nu_s) := & \sup_{h \in \mathcal{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \left([\nabla h(y)]^\top \left(\frac{1}{2} \operatorname{div}_x (\Sigma_s(\varphi_s, y) m_s(y)) - b_s(\varphi_s, y) m_s(y) \right) \right. \\ & \left. - \frac{1}{2} \|\nabla h(y)\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) \right) dy. \end{aligned}$$

In fact, for each $(s, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, $\mathbb{J}_{s, x}(\nu)$ is the large deviations rate function for the empirical measures

$$\nu_t^{s, x}(dy) = \frac{1}{t} \int_0^t \mathbf{1}_{dx}(\tilde{Y}_r^{s, x}) dr$$

for rate $\varepsilon = 1/t$ as $t \rightarrow \infty$ and

$$\dot{\tilde{Y}}_t^{s, x} = b_s(x, \tilde{Y}_t^{s, x}) + \sigma_s(x, \tilde{Y}_t^{s, x}) d\tilde{W}_t;$$

see [34, section 2, Corollaries 2.2 and 2.3].

Moreover, if $\mathbb{I}(\varphi, \mu)$ is finite, it is necessary that

$$\dot{\varphi}_s = \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y) m_s(y)}{\lambda_s(\varphi_s, y)} dy \quad \text{a.e.,}$$

and in this case, we have $\mathbb{I}(\varphi, \mu) = \mathbb{I}_0(\varphi_0) + \int_0^\infty \mathbb{J}_{s, \varphi_s}(\nu_s) ds$. On the other hand, one can also write the rate function (see [34, section 2, Proposition 2.1])

$$(2.14) \quad \mathbb{I}(\varphi, \mu) = \mathbb{I}_0(\varphi_0) + \frac{1}{2} \int_0^\infty \int_{\mathbb{R}^l} \left\| \frac{\nabla_y m_s(y)}{2m_s(y)} - \mathcal{J}_{s, m_s(\cdot), \varphi_s}(y) \right\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) dy ds,$$

with $\mu(dy, ds) = m_s(y) dy ds$, where for each $s \in \mathbb{R}_+$, each function $m_s(\cdot)$ belongs to $\mathcal{P}(\mathbb{R}^l)$, and $\mathcal{J}_{t, m(\cdot), u}$ is a function defined as follows. Denote by $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_s(\varphi_s, y), m_s(y) dy)$ the Hilbert space of all $\mathbb{R}^d$ -valued functions (of y) in $\mathbb{R}^l$ with norm

$$\|f\|_{\Sigma, m}^2 = \int_{\mathbb{R}^l} \|f(y)\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) dy,$$

and by $\mathbb{L}_{\text{loc}}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$ the space consisting of functions whose products with arbitrary $\mathcal{C}_0^\infty$ -functions belong to $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$. Then $\mathcal{J}_{t, m(\cdot), u}$ is defined as a function of y by

$$\mathcal{J}_{t, m(\cdot), u}(y) = \Pi_{\Sigma_t(x, \cdot), m(\cdot)}(\Sigma_t(x, y)^{-1} (b_t(x, y) - \operatorname{div}_x \Sigma_t(x, y)/2)),$$

where $\Pi_{\Sigma_t(x, \cdot), m(\cdot)}$ maps $\phi(y) \in \mathbb{L}_{\text{loc}}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$ to $\Pi_{\Sigma_t(x, \cdot), m(\cdot)} \phi(y)$, which belongs to $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$ and satisfies that for all $h \in \mathcal{C}_0^\infty(\mathbb{R}^l)$,

$$\int_{\mathbb{R}^l} [\nabla h(y)]^\top \Sigma_t(x, y) \Pi_{\Sigma_t(x, \cdot), m(\cdot)} \phi(y) m(y) dy = \int_{\mathbb{R}^l} [\nabla h(y)]^\top \Sigma_t(x, y) \phi(y) m(y) dy.$$

If $\phi(y) \in L^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y)dy)$, then $\Pi_{\Sigma_t(x, \cdot), m(\cdot)}\phi(y)$ is the orthogonal projection of ϕ onto $L^2_{0,1,2}(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y)dy)$.

In fact, $\mathbb{I}(\varphi, \mu)$ is defined similarly to the rate function of the family of processes $\{(\bar{X}^\varepsilon, \mu^{\varepsilon, \bar{X}})\}_{\varepsilon>0}$, where $\mu_t^{\varepsilon, \bar{X}}(\mathcal{A}) = \int_0^t \mathbf{1}_{\mathcal{A}}(Y_s^{\varepsilon, \bar{X}})ds$, $\mathcal{A} \in \mathcal{B}(\mathbb{R}^l)$, and $(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}})$ is the solution of the following equation:

$$\begin{cases} \dot{\bar{X}}_t^\varepsilon = \frac{F_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}})}{\lambda_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}})}, & \bar{X}_0^\varepsilon = x_0^\varepsilon \in \mathbb{R}^d, \\ \dot{Y}_t^{\varepsilon, \bar{X}} = \frac{1}{\varepsilon} b_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}}) + \frac{1}{\sqrt{\varepsilon}} \sigma_t^\varepsilon(\bar{X}_t^\varepsilon, Y_t^{\varepsilon, \bar{X}}) \dot{W}_t, & Y_0^\varepsilon = y_0^\varepsilon \in \mathbb{R}^l. \end{cases}$$

2.2. Proof of Theorem 2.1. This section is devoted to proving Theorem 2.1. In the proof, we use C to represent a generic positive constant that is independent of ε . The value C may change at different appearances; we will specify which parameters it depends on if it is necessary.

2.2.1. A road map for the development of our analysis and proof. To make the proof more readable, we first provide a road map and then illustrate the details in the following sections. The proof of the LDP of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ is based on the approach of [34], which relies on the properties that if a family of random elements is exponentially tight, then it is sequentially large deviations (LD) relatively compact, i.e., any subsequence contains a further subsequence enjoying the LDP with some rate function. The remaining work is done by carefully identifying the rate functions. Specifically, the details are as follows.

Step 1. The exponential tightness of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ is proved in section 2.2.2 by applying the (extended) Puhalskii criteria. Particularly, dealing with X^ε , we prove (2.16), which shows that $\{X^\varepsilon\}$ cannot be large with exponentially small probability, and (2.17), leading to needed continuity properties. To prove these, a first step is to use Lemma 2.1 to deal with the large factor $\frac{1}{\varepsilon^2}$. Then, taking advantage of the martingale property of stochastic integrals enables us to establish desired estimates. It is similar for μ^ε .

Step 2. After proving exponential tightness of $\{(X^\varepsilon, \mu^\varepsilon)\}$, thanks to Proposition 2.1, $\{(X^\varepsilon, \mu^\varepsilon)\}$ is sequentially LD relatively compact (Definition 2.3). Therefore, the second step is devoted to identifying the LD limit points. More precisely, we let $\hat{\mathbb{I}}$ be LD limit rate functions or LD limit points of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ (i.e., a rate function of some subsequence of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ that obeys the LDP), and we prove that $\hat{\mathbb{I}} = \mathbb{I}$ ($\mathbb{I}$ is the rate function defined in section 2.1). Details for this step are as follows.

Step 2(a). We introduce another characterization of the rate function in section 2.2.3. Precisely, for each step function $\beta(s)$ and each $f(t, x, y)$ real-valued $\mathcal{C}^{1,2,2}(\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l)$ -function with compact support in y locally uniformly in (t, x) , define $\Phi_{t \wedge \tau}^{\beta, f}$ as in (2.39). Then, $\mathbb{I}^*$ is defined as the supremum of $\Phi_{t \wedge \tau}^{\beta, f}$ over β, f , and stopping times τ (see (2.40)). Later, as a byproduct of the study of the regularity of $\mathbb{I}^*$ (in section 2.2.5), it is shown that $\mathbb{I}^* = \mathbb{I}$, thanks to their alternative representations (2.14) and (2.54).

Step 2(b). In section 2.2.4, we prove the lower bound of the LD limit, i.e., $\mathbb{I}^* \leq \hat{\mathbb{I}}$ for any (φ, μ) or $\sup_{(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_+(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))} (\Phi_{t \wedge \tau}^{\beta, f}(\varphi, \mu) - \hat{\mathbb{I}}(\varphi, \mu)) = 0$. To prove this claim, using Lemma 2.2 (or [11, Theorem 2.1.10]), it suffices to show that $\mathbb{E}^\varepsilon \exp\{\frac{1}{\varepsilon} \Phi_{t \wedge \tau}^{\varepsilon, \beta, f}(X^\varepsilon, \mu^\varepsilon)\} = 1$, and $\Phi_{t \wedge \tau}^{\varepsilon, \beta, f}(\varphi, \mu) \rightarrow \Phi_{t \wedge \tau}^{\beta, f}(\varphi, \mu)$ as $\varepsilon \rightarrow 0$ uniformly in compact sets; see Theorem 2.4.

Step 2(c). Section 2.2.5 is devoted to the proof of the upper bound of the LD limit, i.e., $\hat{\mathbb{I}} \leq \mathbb{I}^*$. We first show this claim at sufficiently regular (dense) points; see

Theorem 2.5. Then, it is shown for arbitrary points by an approximation using regular points and applying some continuity properties of rate functions.

Remark 2.3. Although our proof is inspired by the approach of [34], let us briefly highlight some differences in our works as follows. (i) System (1.1) is second-order, while [34] studied first-order equations. We do not assume any Lipschitz continuities for coefficients. Therefore, it is not possible to get a “good estimate” (e.g., exponentially close) between (1.1) and its corresponding overdamped first-order equation. (ii) To identify the rate function in Step 2, when defining $\mathbb{I}^*$, we need to narrow the space taking the supremum to control terms containing the derivative p^ε of X^ε (which is due to considering second-order systems), especially the integral involving p^ε and the diffusion part of the fast process (see (2.45)); see Remark 2.5.

2.2.2. Exponential tightness of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$. In this section, we establish the exponential tightness of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^l) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$. We recall these preliminaries below; see [11, 10, 28] for more details.

DEFINITION 2.3. *The family $\{\mathbb{P}^\varepsilon\}_{\varepsilon>0}$ is said to be sequentially LD relatively compact if any subsequence $\{\mathbb{P}^{\varepsilon_k}\}_{k\geq 1}$ of $\{\mathbb{P}^\varepsilon\}_{\varepsilon>0}$ contains a further subsequence $\{\mathbb{P}^{\varepsilon_{k_j}}\}_{j\geq 1}$ which satisfies the LDP with some LD rate function as $j \rightarrow \infty$. We say that a family of random elements of $\mathbb{S}$ is sequentially LD relatively compact (resp., exponentially tight) if their laws have the indicated property.*

PROPOSITION 2.1 ([34, Theorem 4.1]). *If a family $\{\mathbb{P}^\varepsilon\}_{\varepsilon>0}$ is exponentially tight, then it is sequentially LD relatively compact.*

To start, we introduce the following technical lemma, which will be used often in some calculations in this section.

LEMMA 2.1. *For real-valued continuous function $g(s)$, real-valued continuously differentiable function $u(s)$, and real-valued Itô process $w(s)$, $w(s) > 0 \forall s$ with the quadratic variation denoted by $\langle dw, dw \rangle_s$, we have the following identity:*

$$\begin{aligned}
 (2.15) \quad & \frac{1}{\varepsilon^2} \int_0^t u(s) \int_0^s e^{-\frac{1}{\varepsilon^2} \int_r^s w(r') dr'} g(r) dr ds \\
 &= \int_0^t \frac{u(s)g(s)}{w(s)} ds + \int_0^t \frac{\nabla_s u(s)}{w(s)} \left(\int_0^s e^{-\frac{1}{\varepsilon^2} \int_r^s w(r') dr'} g(r) dr \right) ds \\
 &\quad - \frac{u(t)}{w(t)} \int_0^t e^{-\frac{1}{\varepsilon^2} \int_s^t w(r) dr} g(s) ds - \int_0^t \frac{u(s)}{w(s)^2} \left(\int_0^s e^{-\frac{1}{\varepsilon^2} \int_r^s w(r') dr'} g(r) dr \right) dw(s) \\
 &\quad + \int_0^t \frac{u(s)}{w(s)^3} \left(\int_0^s e^{-\frac{1}{\varepsilon^2} \int_r^s w(r') dr'} g(r) dr \right) \langle dw, dw \rangle_s.
 \end{aligned}$$

Moreover, the identity (2.15) still holds if g is an $\mathbb{R}^d$ -valued function and u can be either $\mathbb{R}$ -valued or $\mathbb{R}^d$ -valued (with the operations corresponding to u and g being understood as the inner product in $\mathbb{R}^d$).

Proof. Using integration by parts for

$$u(s) \int_0^s e^{-\frac{1}{\varepsilon^2} \int_0^r w(r') dr'} g(r) dr \text{ and } \frac{e^{-\frac{1}{\varepsilon^2} \int_0^s w(r) dr}}{w(s)},$$

(2.15) follows from standard calculations. $\square$

THEOREM 2.3. *Suppose that Assumption 2.1 holds, that the family $\{(x_0^\varepsilon, y_0^\varepsilon)\}_{\varepsilon>0}$ is exponentially tight, and that $\limsup_{\varepsilon \rightarrow 0} \varepsilon x_1^\varepsilon < \infty$ a.s. Then $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ obtained from (1.1) is exponentially tight and sequentially LD relatively compact in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^l) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$.*

Since $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^l) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ is a closed subset of $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ and $\mathbb{P}^\varepsilon((X^\varepsilon, \mu^\varepsilon) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))) = 1$, it is sufficient to prove that the family $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ is exponentially tight in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$. To prove this, it suffices to verify $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ satisfying the (extended) Puhalskii criteria (see [28, Theorem 3.1] and [13, Remark 4.2]), namely, for all $\ell, t > 0$,

$$(2.16) \quad \lim_{L \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}^\varepsilon \left(\sup_{s \in [0, t]} |X_s^\varepsilon| > L \right) = -\infty,$$

$$(2.17) \quad \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \sup_{s \in [0, t]} \varepsilon \log \mathbb{P}^\varepsilon \left(\sup_{s \leq s_1 \leq s+\delta} |X_{s_1}^\varepsilon - X_s^\varepsilon| > \ell \right) = -\infty,$$

$$(2.18) \quad \lim_{L \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}^\varepsilon (\mu^\varepsilon([0, t], \{y \in \mathbb{R}^l : |y| > L\}) > \ell) = -\infty,$$

$$(2.19) \quad \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \sup_{s \in [0, t]} \varepsilon \log \mathbb{P}^\varepsilon \left(\sup_{s_1 \in [s, s+\delta]} d(\mu_{s_1}^\varepsilon, \mu_s^\varepsilon) > \ell \right) = -\infty.$$

Remark 2.4. In general, (2.16)–(2.19) only imply the sequentially exponential tightness (i.e., any subsequence is exponentially tight). However, because $(X^\varepsilon, \mu^\varepsilon)$ is continuous in ε in distribution, it is true that $(X^\varepsilon, \mu^\varepsilon)$ is exponentially tight, although in our proof only the sequentially exponential tightness is needed.

From (2.2), by the variation of parameter formula, we obtain

$$(2.20) \quad p_t^\varepsilon = x_1^\varepsilon e^{-A_\varepsilon(t)} + \frac{1}{\varepsilon^2} \int_0^t e^{-A_\varepsilon(t,s)} F_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) ds,$$

where for any $0 \leq s \leq t$, $\varepsilon > 0$, $A_\varepsilon(t, s) := \frac{1}{\varepsilon^2} \int_s^t \lambda_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr$ and $dA_\varepsilon(t) = A_\varepsilon(t, 0)$.

Proof of (2.16). It is readily seen that

$$(2.21) \quad \begin{aligned} \ln(1 + |X_t^\varepsilon|^2) - \ln(1 + |x_0^\varepsilon|^2) &= \int_0^t \frac{2(X_s^\varepsilon)^\top p_s^\varepsilon}{1 + |X_s^\varepsilon|^2} ds \\ &= \int_0^t \frac{2(X_s^\varepsilon)^\top x_1^\varepsilon}{1 + |X_s^\varepsilon|^2} e^{-A_\varepsilon(s)} ds + \frac{1}{\varepsilon^2} \int_0^t \frac{2}{1 + |X_s^\varepsilon|^2} (X_s^\varepsilon)^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds. \end{aligned}$$

Denote $v_t^\varepsilon = \frac{2}{1 + |X_t^\varepsilon|^2} (X_t^\varepsilon)^\top \left(\int_0^t e^{-A_\varepsilon(t,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right)$. We have from (2.21) and Lemma 2.1 that

$$(2.22) \quad \begin{aligned} &\ln(1 + |X_t^\varepsilon|^2) - \ln(1 + |x_0^\varepsilon|^2) \\ &= \int_0^t \frac{2(X_s^\varepsilon)^\top x_1^\varepsilon}{1 + |X_s^\varepsilon|^2} e^{-A_\varepsilon(s)} ds + \int_0^t \frac{2(X_s^\varepsilon)^\top F_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}{(1 + |X_s^\varepsilon|^2) \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)} ds \\ &\quad + \int_0^t \frac{2}{\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)} \left(\frac{p_s^\varepsilon}{1 + |X_s^\varepsilon|^2} - \frac{2(X_s^\varepsilon)^\top [X_s^\varepsilon]^\top p_s^\varepsilon}{(1 + |X_s^\varepsilon|^2)^2} \right)^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds \\ &\quad - \frac{v_t^\varepsilon}{\lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon)} - \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]} d\lambda^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) \\ &\quad + \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^3} \langle d\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon), d\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) \rangle_s ds. \end{aligned}$$

Combining (2.22) and the Itô lemma, one has

$$\begin{aligned}
& \ln(1 + |X_t^\varepsilon|^2) - \ln(1 + |x_0^\varepsilon|^2) \\
&= \int_0^t \frac{2(X_s^\varepsilon)^\top x_1^\varepsilon}{1 + |X_s^\varepsilon|^2} e^{-A_\varepsilon(s)} ds + \int_0^t \frac{2(X_s^\varepsilon)^\top F_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}{(1 + |X_s^\varepsilon|^2) \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)} ds \\
&+ \int_0^t \frac{2}{\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)} \left(\frac{p_s^\varepsilon}{1 + |X_s^\varepsilon|^2} - \frac{2(X_s^\varepsilon [X_s^\varepsilon]^\top p_s^\varepsilon)}{(1 + |X_s^\varepsilon|^2)^2} \right)^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds \\
&- \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \nabla_s \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) ds \\
&- \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} [\nabla_X \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top p_s^\varepsilon ds \\
&- \frac{1}{\varepsilon} \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} [\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top b_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) ds \\
&- \frac{1}{2\varepsilon} \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \|\nabla_{YY} \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2 ds \\
&+ \frac{1}{\varepsilon} \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^3} \|\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2 ds \\
&- \frac{1}{\sqrt{\varepsilon}} \int_0^t \frac{v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} [\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
&=: K_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}} \int_0^t \frac{-v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} [\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s,
\end{aligned}$$

where K_t^ε is the remaining terms in the right-hand side. Therefore, we get

$$(2.23) \quad \frac{1}{\varepsilon} [\ln(1 + |X_t^\varepsilon|^2) - \ln(1 + |x_0^\varepsilon|^2)] = \frac{1}{\varepsilon} \widehat{K}_t^\varepsilon + D_t^\varepsilon,$$

where

$$\begin{aligned}
\widehat{K}_t^\varepsilon &:= \left[K_t^\varepsilon + \frac{1}{\varepsilon^2} \int_0^t \frac{|v_s^\varepsilon|^2}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^4} \|\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2 ds \right] \quad \text{and} \\
D_t^\varepsilon &= \frac{1}{\varepsilon \sqrt{\varepsilon}} \int_0^t \frac{-v_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} [\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
&\quad - \frac{1}{\varepsilon^3} \int_0^t \frac{|v_s^\varepsilon|^2}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^4} \|\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2 ds.
\end{aligned}$$

Let $\zeta_L^\varepsilon = \inf \{t \geq 0 : |X_t^\varepsilon| > L\}$. It is obvious that ζ_L^ε is an $\mathcal{F}^\varepsilon$ -stopping time. Since D_t^ε is a local martingale, we have from (2.23) that

$$(2.24) \quad \mathbb{E}^\varepsilon \exp \left\{ \frac{1}{\varepsilon} \left[\ln(1 + |X_{t \wedge \zeta_L^\varepsilon}^\varepsilon|^2) - \ln(1 + |x_0^\varepsilon|^2) - \widehat{K}_{t \wedge \zeta_L^\varepsilon}^\varepsilon \right] \right\} \leq 1.$$

On the other hand, from (2.20) and Assumption 2.1, and noting $\int_0^t e^{-A_\varepsilon(t,s)} ds \leq \int_0^t e^{-\frac{\kappa_0(t-s)}{\varepsilon^2}} ds \leq \frac{\varepsilon^2}{\kappa_0}$, one obtains that there is a finite constant $C_{t,L}$ depending only on t, L that satisfies, for all sufficiently small ε ,

$$(2.25) \quad |p_{t \wedge \zeta_L^\varepsilon}^\varepsilon| \leq C_{t,L} + x_1^\varepsilon e^{-A_\varepsilon(t)}.$$

Similarly, we have for ε small

$$(2.26) \quad |v_{t \wedge \zeta_L^\varepsilon}^\varepsilon| \leq C_{t,L} \int_0^t e^{-\frac{\kappa_0(t-s)}{\varepsilon^2}} ds \leq \varepsilon^2 C_{t,L}.$$

Combining (2.25), (2.26), the definition of $\widehat{K}_t^\varepsilon$, Assumption 2.1, and $\limsup_{\varepsilon \rightarrow 0} \varepsilon x_1^\varepsilon < \infty$ yields that as ε is small

$$(2.27) \quad |\widehat{K}_{t \wedge \zeta_L^\varepsilon}^\varepsilon| \leq C + \varepsilon C_{t,L},$$

where C is some finite constant depending on neither ε nor t, L . Therefore, from (2.24) and (2.27), we have that for all ε sufficiently small,

$$\mathbb{E}^\varepsilon \exp \left\{ \frac{1}{\varepsilon} \left[\ln(1 + |X_{t \wedge \zeta_L^\varepsilon}^\varepsilon|^2) - \ln(1 + |x_0^\varepsilon|^2) - C - \varepsilon C_{t,L} \right] \right\} \leq 1.$$

Thus, one has for any $t, L, N > 0$,

$$(2.28) \quad \begin{aligned} & \mathbb{P}^\varepsilon \left(\sup_{s \in [0,t]} |X_s^\varepsilon| > L \right) = \mathbb{P}^\varepsilon (|X_{t \wedge \zeta_L^\varepsilon}^\varepsilon| > L) \\ & \leq \mathbb{P}^\varepsilon (|x_0^\varepsilon| > N) + \mathbb{E}^\varepsilon \exp \left\{ \frac{1}{\varepsilon} \left[\ln(1 + |X_{t \wedge \zeta_L^\varepsilon}^\varepsilon|^2) - \ln(1 + L^2) \right] \right\} \mathbf{1}_{\{|x_0^\varepsilon| \leq N\}} \\ & \leq \mathbb{P}^\varepsilon (|x_0^\varepsilon| > N) + \exp \left\{ \frac{1}{\varepsilon} \left[\ln(1 + N^2) + C + \varepsilon C_{t,L} - \ln(1 + L^2) \right] \right\}. \end{aligned}$$

From (2.28) and the logarithm equivalence principle [10, Lemma 1.2.15], we obtain that for all $t, N > 0$,

$$(2.29) \quad \lim_{L \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P} \left(\sup_{s \in [0,t]} |X_s^\varepsilon| > L \right) \leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}^\varepsilon (|x_0^\varepsilon| > N).$$

Because $\{x_0^\varepsilon\}_{\varepsilon > 0}$ is exponentially tight and using (2.29), we obtain (2.16) for any $t > 0$. $\square$

Proof of (2.17). By applying Lemma 2.1 to (2.20), one has

$$(2.30) \quad \begin{aligned} X_t^\varepsilon &= x_0^\varepsilon + \int_0^t p_s^\varepsilon ds = x_0^\varepsilon + \int_0^t x_1^\varepsilon e^{-A_\varepsilon(s)} ds + \frac{1}{\varepsilon^2} \int_0^t \int_0^s e^{-A_\varepsilon(s,r)} F_s^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr ds \\ &= x_0^\varepsilon + x_1^\varepsilon \int_0^t e^{-A_\varepsilon(r)} dr + \int_0^t \frac{F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)}{\lambda_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)} dr \\ &\quad - \frac{1}{\lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon)} \int_0^t e^{-A_\varepsilon(t,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \\ &\quad - \int_0^t \frac{1}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) d\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) \\ &\quad + \int_0^t \frac{1}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^3} \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) \langle d\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon), d\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) \rangle_s. \end{aligned}$$

We obtain from (2.30) and Itô's formula that

$$(2.31) \quad \begin{aligned} X_t^\varepsilon &= x_0^\varepsilon + x_1^\varepsilon \int_0^t e^{-A_\varepsilon(r)} dr + \int_0^t \frac{F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)}{\lambda_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)} dr \\ &\quad - \frac{1}{\lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon)} \int_0^t e^{-A_\varepsilon(t,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \\ &\quad - \int_0^t \frac{\nabla_s \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds \end{aligned}$$

$$\begin{aligned}
& - \int_0^t \frac{[\nabla_X \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top p_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds \\
& - \frac{1}{\varepsilon} \int_0^t \frac{[\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top b_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds \\
& - \frac{1}{2\varepsilon} \int_0^t \frac{\|\nabla_{YY} \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds \\
& + \frac{1}{\varepsilon} \int_0^t \frac{\|\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^3} \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ds \\
& - \frac{1}{\sqrt{\varepsilon}} \int_0^t \frac{\left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ([\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \\
& =: \bar{K}_t^\varepsilon - \bar{D}_t^\varepsilon,
\end{aligned}$$

where

$$\bar{D}_t^\varepsilon := \frac{1}{\sqrt{\varepsilon}} \int_0^t \frac{\left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) ([\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2},$$

and $\bar{K}_t^\varepsilon$ is the remaining terms in the right-hand side of (2.31). By the regularity of λ_t^ε , it is not difficult to see that

$$\begin{aligned}
(2.32) \quad & \left| \frac{1}{\lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon)} \int_0^t e^{-A_\varepsilon(t,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) ds - \frac{1}{\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)} \int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) ds \right| \\
& \leq \left| \frac{e^{-A_\varepsilon(t)}}{\lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon)} - \frac{e^{-A_\varepsilon(s)}}{\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)} \right| \int_0^s e^{A_\varepsilon(r)} |F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)| dr \\
& \quad + \frac{e^{-A_\varepsilon(t)}}{\lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon)} \int_s^t e^{A_\varepsilon(r)} |F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)| dr \\
& \leq C \frac{|t-s|}{\varepsilon^2} \sup_{r \in [s,t]} |\lambda_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)| \int_0^s e^{-\frac{\kappa_0}{\varepsilon^2}(s-r)} |F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)| dr \\
& \quad + C \int_s^t e^{-\frac{\kappa_0}{\varepsilon^2}(t-r)} |F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon)| dr.
\end{aligned}$$

We obtain from the definition of $\bar{K}_t^\varepsilon$, an application of (2.32), and recalling the definition of ζ_L^ε that there is a finite constant C_L depending only on L such that for all small ε ,

$$(2.33) \quad \sup_{s \in [0, T]} \sup_{t \in [s, s+\delta]} \left| \bar{K}_{t \wedge \zeta_L^\varepsilon}^\varepsilon - \bar{K}_{s \wedge \zeta_L^\varepsilon}^\varepsilon \right| \leq C_{T,L} \delta \quad \forall T > 0, 0 < \delta < 1.$$

Now, let $T > 0, \ell > 0$ be fixed, and let $L > 0$ be fixed but otherwise arbitrary. We have that for any small δ satisfying $\delta < 1$, $C_{T,L} \delta < \ell/2$, and small ε ,

$$\begin{aligned}
(2.34) \quad & \mathbb{P}^\varepsilon \left(\sup_{t \in [s, s+\delta]} |X_t^\varepsilon - X_s^\varepsilon| > \ell \right) \\
& \leq \mathbb{P}^\varepsilon (\zeta_L^\varepsilon \leq T + 1) + \mathbb{P}^\varepsilon \left(\sup_{t \in [s, s+\delta]} \left| X_{t \wedge \zeta_L^\varepsilon}^\varepsilon - X_{s \wedge \zeta_L^\varepsilon}^\varepsilon \right| > \ell \right)
\end{aligned}$$

$$\begin{aligned} &\leq \mathbb{P}^\varepsilon(\zeta_L^\varepsilon \leq T+1) + \mathbb{P}^\varepsilon\left(\sup_{t \in [s, s+\delta]} |\overline{D}_{t \wedge \zeta_L^\varepsilon}^\varepsilon - \overline{D}_{s \wedge \zeta_L^\varepsilon}^\varepsilon| > \frac{\ell}{2}\right) \\ &\leq \mathbb{P}^\varepsilon\left(\sup_{t \in [0, T+1]} |X_t^\varepsilon| > L\right) + \sum_{k=1}^d \mathbb{P}^\varepsilon\left(\sup_{t \in [s, s+\delta]} |\overline{D}_{t \wedge \zeta_L^\varepsilon}^{\varepsilon, k} - \overline{D}_{s \wedge \zeta_L^\varepsilon}^{\varepsilon, k}| > \frac{\ell}{2}\right), \end{aligned}$$

where $\overline{D}_t^{\varepsilon, k}$ is the k th component of $\overline{D}_t^\varepsilon$, $k = 1, \dots, d$. It is readily seen that $\{\overline{D}_{t \wedge \zeta_L^\varepsilon}^{\varepsilon, k} - \overline{D}_{s \wedge \zeta_L^\varepsilon}^{\varepsilon, k}\}_{t \geq s}$ is a martingale with the quadratic variations bounded by

$$\frac{C_L}{\varepsilon} \int_{s \wedge \zeta_L^\varepsilon}^{t \wedge \zeta_L^\varepsilon} \int_0^s e^{-\frac{2\kappa_0(s-r)}{\varepsilon^2}} dr ds \leq \varepsilon C_L \delta.$$

By the exponential martingale inequality [29, Theorem 7.4, p. 44], we have

$$\begin{aligned} &\mathbb{P}^\varepsilon\left(\sup_{t \in [s, s+\delta]} |\overline{D}_{t \wedge \zeta_L^\varepsilon}^{\varepsilon, k} - \overline{D}_{s \wedge \zeta_L^\varepsilon}^{\varepsilon, k}| > \frac{\ell}{2}\right) \\ (2.35) \quad &\leq \mathbb{P}^\varepsilon\left(\sup_{t \in [s, s+\delta]} \left|\overline{D}_{t \wedge \zeta_L^\varepsilon}^{\varepsilon, k} - \overline{D}_{s \wedge \zeta_L^\varepsilon}^{\varepsilon, k}\right| > \frac{\ell}{4} + \frac{\ell}{4\varepsilon C_L \delta} \varepsilon C_L \delta\right) \\ &\leq \exp\left\{-\frac{\ell^2}{8\varepsilon C_L \delta}\right\}. \end{aligned}$$

Combining (2.34) and (2.35), the logarithm equivalence principle [10, Lemma 1.2.15] yields that

$$\begin{aligned} &\lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \sup_{s \in [0, T]} \varepsilon \log \mathbb{P}^\varepsilon\left(\sup_{t \in [s, s+\delta]} |X_t^\varepsilon - X_s^\varepsilon| > \ell\right) \\ (2.36) \quad &\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}^\varepsilon\left(\sup_{t \in [0, T+1]} |X_t^\varepsilon| > L\right) \quad \forall L > 0. \end{aligned}$$

Letting $L \rightarrow \infty$ and using (2.16), we obtain (2.17). $\square$

Proof of (2.18) and (2.19). Once we established the exponential tightness of $\{X^\varepsilon\}_{\varepsilon > 0}$, the proof of (2.18) and (2.19) for $\{\mu^\varepsilon\}_{\varepsilon > 0}$, which is in fact the occupation measure of a diffusion, is similar to that of the first-order coupled systems. As a consequence, such proofs can be found in [34, p. 3134]. $\square$

2.2.3. Characterization of rate function. Let $\beta(s) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ be a step function satisfying that there are $0 = t_0 < t_1 < \dots < t_m < \infty$ and $\beta_i \in \mathbb{R}, i = 1, \dots, m$, such that

$$(2.37) \quad \beta(s) = \sum_{i=1}^m \beta_i \mathbf{1}_{[t_{i-1}, t_i)}(s).$$

For $\varphi_s \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ and $\beta(s)$ of the form (2.37), we define

$$(2.38) \quad \int_0^t \beta(s) d\varphi_s := \sum_{i=1}^m \beta_i^\top (\varphi_{t \wedge t_i} - \varphi_{t \wedge t_{i-1}}).$$

Now, for each step function $\beta(s)$, each $f(t, x, y)$ real-valued $\mathcal{C}^{1,2,2}(\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^l)$ -function with compact support in y locally uniformly in (t, x) , and each $(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$, let

$$\begin{aligned}
\Phi_t^{\beta, f}(\varphi, \mu) &:= \int_0^t \beta(s) d\varphi_s - \int_0^t \int_{\mathbb{R}^l} \frac{[\beta(s)]^\top F_s(\varphi_s, y)}{\lambda_s(\varphi_s, y)} \mu(ds, dy) \\
&\quad - \int_0^t \int_{\mathbb{R}^l} [\nabla_y f(s, \varphi_s, y)]^\top b_s(\varphi_s, y) \mu(ds, dy) \\
&\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \text{tr} \left(\Sigma_s(\varphi_s, y) \nabla_{yy} f(s, \varphi_s, y) \right) \mu(ds, dy) \\
&\quad - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \|\nabla_y f(s, \varphi_s, y)\|_{\Sigma_s(\varphi_s, y)}^2 \mu(ds, dy).
\end{aligned}
\tag{2.39}$$

Moreover, let $\tau(\varphi, \mu)$ be a continuous function of $(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ that is also a stopping time relative to the flow $G = (G_t, t \in \mathbb{R}_+)$ on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ of the σ -algebra G_t generated by the mappings $\varphi \rightarrow \varphi_s$ and $\mu \rightarrow \mu_s$ for $s \leq t$. Let us also assume that $\varphi_{t \wedge \tau(\varphi, \mu)}$ is a bounded function of (φ, μ) . It is seen that under Assumption 2.2, $\Phi_t^{\beta, f}(\varphi, \mu)$ is continuous in (φ, μ) .

Next, define

$$\mathbb{I}^*(\varphi, \mu) = \sup_{\beta, f, t, \tau} \Phi_{t \wedge \tau(\varphi, \mu)}^{\beta, f}(\varphi, \mu),
\tag{2.40}$$

where the supremum is taken over $\beta(s)$, $f(s, x, y)$, and $\tau(\varphi, \mu)$ satisfying the requirements as above and over $t \geq 0$. It is seen that $\mathbb{I}^*$ is lower semicontinuous in (φ, μ) .

Now, let $\widehat{\mathbb{I}}$ be LD limit rate functions or LD limit points of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon > 0}$ (i.e., a rate function of some subsequence of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon > 0}$ that obeys the LDP) such that $\widehat{\mathbb{I}}(\varphi, \mu) = \infty$ unless $\varphi_0 = \widehat{x}$, where $\widehat{x}$ is a preselected element of $\mathbb{R}^d$. This restriction will be removed in section 2.2.6. We will identify the rate functions. For any such LD limit point $\widehat{\mathbb{I}}$, we aim to prove $\widehat{\mathbb{I}} = \mathbb{I}^*$ by showing the upper bound $\widehat{\mathbb{I}} \geq \mathbb{I}^*$ and the lower bound $\widehat{\mathbb{I}} \leq \mathbb{I}^*$; see details in section 2.2.4 and section 2.2.5. Moreover, it will be seen that $\mathbb{I}^*(\varphi, \mu) = \mathbb{I}(\varphi, \mu)$ provided $\mathbb{I}^*(\varphi, \mu) < \infty$, $\varphi_0 = \widehat{x}$, and $\mathbb{I}_0(\widehat{x}) = 0$. Throughout this section, the assumptions in Theorem 2.1 are always assumed to be satisfied.

Remark 2.5. Note that $\mathbb{I}^*$ is defined similarly but not identically as in the case of first-order coupled systems in [34], although the solution of (1.1) shares the same rate function with the corresponding first-order system. Compared with [34], $\mathbb{I}^*$ is defined by taking the supremum over smaller space when we did not allow β to be a function of X . This modification has an important role in the proof of the lower bound of the LD limits, i.e., the inequality $\mathbb{I}^* \leq \widehat{\mathbb{I}}$. Otherwise, it would be impossible to control terms containing the derivative p^ε of X^ε , especially the integral involving p^ε and the diffusion part of the fast process (see (2.45)). Meanwhile, it would have led to difficulty in proving the upper bound of the LD limits, i.e., the inequality $\mathbb{I}^* \geq \widehat{\mathbb{I}}$. However, it will be shown that we still can get the upper bound as in the first-order system (in [34]); see the details in section 2.2.5.

2.2.4. Lower bound of large deviations limits. This section is devoted to proving $\mathbb{I}^* \leq \widehat{\mathbb{I}}$. We have the following theorem.

THEOREM 2.4. *Let $\widehat{\mathbb{I}}$ be an LD limit point of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon > 0}$. For any $t > 0$, β, f, τ are as given above,*

$$\sup_{(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))} \left(\Phi_{t \wedge \tau(\varphi, \mu)}^{\beta, f}(\varphi, \mu) - \widehat{\mathbb{I}}(\varphi, \mu) \right) = 0.
\tag{2.41}$$

Then $\mathbb{I}^(\varphi, \mu) \leq \widehat{\mathbb{I}}(\varphi, \mu)$ for all $(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$.*

Proof. For $\beta(s)$ being of the form (2.37) and $f(s, x, y)$ being a function with compact support in y locally uniformly in (t, x) , denote

$$\begin{aligned}
 (2.42) \quad & \Gamma_t^{\varepsilon, \beta}(\varphi, \mu) \\
 &= - \int_0^t \int_{\mathbb{R}^l} [\beta(s)]^\top x_1^\varepsilon e^{-A_\varepsilon^{\varphi, y}(s)} \mu(ds, dy) \\
 &+ \int_0^t \int_{\mathbb{R}^l} \frac{\beta(t)}{\lambda_t^\varepsilon(\varphi_t, y)} e^{-A_\varepsilon^{\varphi, y}(t, s)} F_s^\varepsilon(\varphi_s, y) \mu(ds, dy) \\
 &+ \int_0^t \int_{\mathbb{R}^l} [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon^{\varphi, y}(s, r)} F_r^\varepsilon(\varphi_r, y) dr \right) \frac{\nabla_s \lambda_s^\varepsilon(\varphi_s, y)}{[\lambda_s^\varepsilon(\varphi_s, y)]^2} \mu(ds, dy) \\
 &+ \int_0^t \int_{\mathbb{R}^l} [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon^{\varphi, y}(s, r)} F_r^\varepsilon(\varphi_r, y) dr \right) \frac{[\nabla_X \lambda_s^\varepsilon(\varphi_s, y)]^\top \dot{\varphi}_s}{[\lambda_s^\varepsilon(\varphi_s, y)]^2} \mu(ds, dy) \\
 &+ \frac{1}{\varepsilon} \int_0^t \int_{\mathbb{R}^l} [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon^{\varphi, y}(s, r)} F_r^\varepsilon(\varphi_r, y) dr \right) \frac{[\nabla_Y \lambda_s^\varepsilon(\varphi_s, y)]^\top b_s^\varepsilon(\varphi_s, y)}{[\lambda_s^\varepsilon(\varphi_s, y)]^2} \mu(ds, dy) \\
 &+ \frac{1}{2\varepsilon} \int_0^t \int_{\mathbb{R}^l} [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon^{\varphi, y}(s, r)} F_r^\varepsilon(\varphi_r, y) dr \right) \frac{\|\nabla_{YY} \lambda_s^\varepsilon(\varphi_s, y)\|_{\Sigma_s^\varepsilon(\varphi_s, y)}^2}{[\lambda_s^\varepsilon(\varphi_s, y)]^2} \mu(ds, dy) \\
 &- \frac{1}{\varepsilon} \int_0^t \int_{\mathbb{R}^l} [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon^{\varphi, y}(s, r)} F_r^\varepsilon(\varphi_r, y) dr \right) \frac{\|\nabla_Y \lambda_s^\varepsilon(\varphi_s, y)\|_{\Sigma_s^\varepsilon(\varphi_s, y)}^2}{[\lambda_s^\varepsilon(\varphi_s, y)]^3} \mu(ds, dy) \\
 &- \frac{1}{2\varepsilon^2} \int_0^t \int_{\mathbb{R}^l} \frac{\left| [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon^{\varphi, y}(s, r)} F_r^\varepsilon(\varphi_r, y) dr \right) \right|^2}{[\lambda_s^\varepsilon(\varphi_s, y)]^4} \|\nabla_Y \lambda_s^\varepsilon(\varphi_s, y)\|_{\Sigma_s^\varepsilon(\varphi_s, y)}^2 \mu(ds, dy), \\
 &+ \frac{1}{\varepsilon} \int_0^t \int_{\mathbb{R}^l} \frac{[\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon^{\varphi, y}(s, r)} F_r^\varepsilon(\varphi_r, y) dr \right)}{[\lambda_s^\varepsilon(\varphi_s, y)]^2} \|\nabla_Y \lambda_s^\varepsilon(\varphi_s, y)\|_{\Sigma_s^\varepsilon(\varphi_s, y)}^2 \mu(ds, dy),
 \end{aligned}$$

where $A_\varepsilon^{\varphi, y}(t, s) := \frac{1}{\varepsilon^2} \int_s^t \int_{\mathbb{R}^l} \lambda_r^\varepsilon(\varphi_r, y) \mu(dr, dy)$, and

$$\begin{aligned}
 (2.43) \quad & \Psi_t^{\varepsilon, \beta, f}(\varphi, \mu) := f(t, \varphi_t, Y_t^\varepsilon) - f(0, \varphi_0, y_0^\varepsilon) - \int_0^t \int_{\mathbb{R}^l} \nabla_s f(s, \varphi_s, y) \mu(ds, dy) \\
 & - \int_0^t \int_{\mathbb{R}^l} [\nabla_x f(s, \varphi_s, y)]^\top \dot{\varphi}_s \mu(ds, dy),
 \end{aligned}$$

and

$$\begin{aligned}
 (2.44) \quad & \Phi_t^{\varepsilon, \beta, f}(\varphi, \mu) := \int_0^t \beta(s) d\varphi_s - \int_0^t \int_{\mathbb{R}^l} \frac{[\beta(s)]^\top F_s^\varepsilon(\varphi_s, y)}{\lambda_s^\varepsilon(\varphi_s, y)} \mu(ds, dy) \\
 & - \int_0^t \int_{\mathbb{R}^l} [\nabla_y f(s, \varphi_s, y)]^\top b_s^\varepsilon(\varphi_s, y) \mu(ds, dy) \\
 & - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \text{tr} \left(\Sigma_s^\varepsilon(\varphi_s, y) \nabla_{yy} f(s, \varphi_s, y) \right) \mu(ds, dy) \\
 & - \frac{1}{2} \int_0^t \int_{\mathbb{R}^l} \|\nabla_y f(s, \varphi_s, y)\|_{\Sigma_s^\varepsilon(\varphi_s, y)}^2 \mu(ds, dy).
 \end{aligned}$$

We have from (2.20) and Lemma 2.1 that

$$\begin{aligned}
 (2.45) \quad & \int_0^t [\beta(s)]^\top dX_s^\varepsilon = \int_0^t [\beta(s)]^\top p_s^\varepsilon ds \\
 & = \int_0^t [\beta(s)]^\top x_1^\varepsilon e^{-A_\varepsilon(s)} ds + \frac{1}{\varepsilon^2} \int_0^t [\beta(s)]^\top \int_0^s e^{-A_\varepsilon(s, r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr
 \end{aligned}$$

$$\begin{aligned}
&= \int_0^t [\beta(s)]^\top x_1^\varepsilon e^{-A_\varepsilon(s)} ds + \int_0^t \frac{[\beta(s)]^\top F_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}{\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)} ds \\
&\quad - \frac{\beta(t)}{\lambda_t^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon)} \int_0^t e^{-A_\varepsilon(t,s)} F_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) ds \\
&\quad - \int_0^t [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) \frac{\nabla_s \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} ds \\
&\quad - \int_0^t [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) \frac{[\nabla_X \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top p_s^\varepsilon}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} ds \\
&\quad - \frac{1}{\varepsilon} \int_0^t [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) \frac{[\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top b_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} ds \\
&\quad - \frac{1}{2\varepsilon} \int_0^t [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) \frac{\|\nabla_{YY} \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} ds \\
&\quad + \frac{1}{\varepsilon} \int_0^t [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) \frac{\|\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^3} ds \\
&\quad - \frac{1}{\sqrt{\varepsilon}} \int_0^t \frac{[\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} [\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s.
\end{aligned}$$

Moreover, Itô's formula yields that

(2.46)

$$\begin{aligned}
f(t, X_t^\varepsilon, Y_t^\varepsilon) - f(0, X_0^\varepsilon, Y_0^\varepsilon) &= \int_0^t \nabla_s f(s, X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t [\nabla_X f(s, X_s^\varepsilon, Y_s^\varepsilon)]^\top p_s^\varepsilon ds \\
&\quad + \frac{1}{\varepsilon} \int_0^t [\nabla_Y f(s, X_s^\varepsilon, Y_s^\varepsilon)]^\top b_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} [\nabla_Y f(s, X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
&\quad + \frac{1}{2\varepsilon} \int_0^t \text{tr} \left(\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) \nabla_{YY} f(s, X_s^\varepsilon, Y_s^\varepsilon) \right) ds.
\end{aligned}$$

Combining (2.46), (2.45), and the definitions of $\Phi_t^{\varepsilon, \beta, f}$, $\Gamma_t^{\varepsilon, \beta}$, and $\Psi_t^{\varepsilon, \beta, f}$ in (2.44), (2.42), and (2.43), we obtain that

(2.47)

$$\begin{aligned}
&\frac{1}{\varepsilon} \left[\Phi_t^{\varepsilon, \beta, f}(X^\varepsilon, \mu^\varepsilon) + \Gamma_t^{\varepsilon, \beta}(X^\varepsilon, \mu^\varepsilon) \right] + \Psi_t^{\varepsilon, \beta, f}(X^\varepsilon, \mu^\varepsilon) \\
&= \frac{1}{\sqrt{\varepsilon}} \int_0^t [\nabla_Y f(s, X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
&\quad - \frac{1}{2\varepsilon} \int_0^t \|\nabla_Y f(s, X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2 ds \\
&\quad - \frac{1}{\varepsilon \sqrt{\varepsilon}} \int_0^t \frac{[\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} [\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^\top \sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
&\quad - \frac{1}{2\varepsilon^3} \int_0^t \frac{\left| [\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right) \right|^2}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^4} \|\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2 ds \\
&\quad + \frac{1}{\varepsilon^2} \int_0^t \frac{[\beta(s)]^\top \left(\int_0^s e^{-A_\varepsilon(s,r)} F_r^\varepsilon(X_r^\varepsilon, Y_r^\varepsilon) dr \right)}{[\lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)]^2} \|\nabla_Y \lambda_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)\|_{\Sigma_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon)}^2 ds.
\end{aligned}$$

Since the right-hand side of (2.47) is a local martingale and $\tau(X^\varepsilon, \mu^\varepsilon)$ is a stopping time with respect to $\mathcal{F}^\varepsilon$ due to the measurability of $X_t^\varepsilon, \mu_t^\varepsilon$ with respect to $\mathcal{F}_t^\varepsilon$, we have that

(2.48)

$$\mathbb{E}^\varepsilon \exp \left\{ \frac{1}{\varepsilon} \left[\Phi_{t \wedge \tau(X^\varepsilon, \mu^\varepsilon)}^{\varepsilon, \beta, f}(X^\varepsilon, \mu^\varepsilon) + \Gamma_{t \wedge \tau(X^\varepsilon, \mu^\varepsilon)}^{\varepsilon, \beta}(X^\varepsilon, \mu^\varepsilon) + \varepsilon \Psi_{t \wedge \tau(X^\varepsilon, \mu^\varepsilon)}^{\varepsilon, \beta, f}(X^\varepsilon, \mu^\varepsilon) \right] \right\} = 1.$$

LEMMA 2.2 ([11, Theorem 2.1.10]). Assume that the net $\{\nu_\varepsilon\}_{\varepsilon>0}$ is exponentially tight, and let $\widehat{\mathbb{I}}$ represent an LD limit point of $\{\nu_\varepsilon\}_{\varepsilon>0}$. Let Φ_ε be a net of uniformly bounded real-valued functions on $\mathbb{S}$ such that $\int_{\mathbb{S}} \exp(\frac{1}{\varepsilon} \Phi_\varepsilon(z)) \nu_\varepsilon(dz) = 1$. If Φ_ε converges to Φ uniformly on compact sets (as $\varepsilon \rightarrow 0$) with the function Φ being continuous, then $\sup_{z \in \mathbb{S}} (\Phi(z) - \widehat{\mathbb{I}}(z)) = 0$.

As in the proof of (2.16) and (2.17) in section 2.2.2, it is not difficult to obtain from Assumption 2.1 and the fact that $\varphi_{t \wedge \tau(\varphi, \mu)}$ is a bounded function of (φ, μ) that there is finite a constant C , which is independent of ε such that for all small enough ε , $|\Gamma_{t \wedge \tau(\varphi, \mu)}^{\varepsilon, \beta}(\varphi, \mu)| \leq C\varepsilon$ uniformly over (φ, μ) . Similarly, there is a constant C such that for ε sufficiently small, $|\Psi_{t \wedge \tau(\varphi, \mu)}^{\varepsilon, \beta, f}(\varphi, \mu)| < C$ uniformly over (φ, μ) . As a result, one has

$$(2.49) \quad \Gamma_{t \wedge \tau(\varphi, \mu)}^{\varepsilon, \beta}(\varphi, \mu) + \varepsilon \Psi_{t \wedge \tau(\varphi, \mu)}^{\varepsilon, \beta, f}(\varphi, \mu) \rightarrow 0$$

as $\varepsilon \rightarrow 0$ uniformly in compact sets. Finally, by assumption (2.8), we have

$$(2.50) \quad \Phi_{t \wedge \tau(\varphi, \mu)}^{\varepsilon, \beta, f}(\varphi, \mu) \rightarrow \Phi_{t \wedge \tau(\varphi, \mu)}^{\beta, f}(\varphi, \mu)$$

as $\varepsilon \rightarrow 0$ uniformly in compact sets. Combining (2.48) and (2.49) and then applying Lemma 2.2 yields (2.41). Then, it follows immediately that $\mathbb{I}^*(\varphi, \mu) \leq \widehat{\mathbb{I}}(\varphi, \mu)$ for all $(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$. The proof is complete. $\square$

2.2.5. Upper bound of large deviations limits. Let $\widehat{\mathbb{I}}$ be an LD limit point of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ such that $\widehat{\mathbb{I}}(\varphi, \mu) = \infty$ unless $\varphi_0 = \widehat{x}$, a preselected element of $\mathbb{R}^d$. In this section, we aim to prove that $\widehat{\mathbb{I}}(\varphi, \mu) \leq \mathbb{I}^*(\varphi, \mu)$ for any $(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ such that $\varphi_0 = \widehat{x}$. The completion of the proof will be given later in section 2.2.6. With the results established in sections 2.2.2 and 2.2.4, this part can be done similarly to that of [34, sections 6–8] because the rate function has a similar variational representation. Although our $\mathbb{I}^*$ is defined as the supremum in a smaller space than in [34], we can still prove $\widehat{\mathbb{I}} \leq \mathbb{I}^*$ by a similar argument as in [34]. We will only provide a sketch of the main ideas and highlight the differences, whereas detailed arguments can be found in [34, sections 6–8].

It is obvious that it suffices to consider the case $\mathbb{I}^*(\varphi, \mu) < \infty$. Therefore, we should investigate the regularity of (φ, μ) provided $\mathbb{I}^*(\varphi, \mu) < \infty$ first. It is shown in [34, section 6] that if $(\varphi, \mu) \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \times \mathcal{C}_\uparrow(\mathbb{R}_+, \mathbb{R}^l)$, $\mathbb{I}^*(\varphi, \mu) < \infty$, then $\mu(ds, dy) = m_s(y) dy ds$, φ_s is absolutely continuous (w.r.t. Lebesgue measure on $\mathbb{R}_+$), and $m_s(y)$ is a probability density function in $\mathbb{R}^l$. In this case, $\mathbb{I}^*$ has the following representation:

$$(2.51) \quad \begin{aligned} & \mathbb{I}^*(\varphi, \mu) \\ &= \int_0^\infty \left(\sup_{\beta \in \mathbb{R}^d} \left(\beta^\top \dot{\varphi}_s - \beta^\top \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y) m_s(y)}{\lambda_s(\varphi_s, y)} dy \right) \right. \\ & \quad + \sup_{h \in C_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} [\nabla_y h(y)]^\top \left(\frac{1}{2} \operatorname{div}(\Sigma_s(\varphi_s, y) m_s(y)) - b_s(\varphi_s, y) m_s(y) \right) dy \\ & \quad \left. - \frac{1}{2} \int_{\mathbb{R}^l} \|\nabla_y h(y)\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) dy \right) ds \end{aligned}$$

$$\begin{aligned}
&= \int_0^\infty \left(\sup_{\beta \in \mathbb{R}^d} \left(\beta^\top \dot{\varphi}_s - \beta^\top \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y) m_s(y)}{\lambda_s(\varphi_s, y)} dy \right) \right. \\
&\quad + \sup_{g \in \mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, \Sigma_s(\varphi_s, y), m_s(y) dy)} \int_{\mathbb{R}^l} \left([g(y)]^\top \Sigma_s(\varphi_s, y) \left(\frac{\nabla_y m_s(y)}{2m_s(y)} - \mathcal{J}_{s, m_s(\cdot), \varphi_s}(y) \right) \right. \\
&\quad \left. \left. - \frac{1}{2} \|g(y)\|_{\Sigma_s(\varphi_s, y) m_s(y)} \right) dy \right) ds.
\end{aligned}$$

In the above, $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, \Sigma_s(\varphi_s, y), m_s(y) dy)$ represents the closure of the set of the gradients of functions from $\mathcal{C}_0^\infty(\mathbb{R}^l)$ in $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_s(\varphi_s, y), m_s(y) dy)$, in which $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_s(\varphi_s, y), m_s(y) dy)$ is the Hilbert space of $\mathbb{R}^d$ -valued functions (of y) in $\mathbb{R}^l$ endowed with the norm $\|f\|_{\Sigma_s, m}^2 = \int_{\mathbb{R}^l} \|f(y)\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) dy$, and for each $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, each function $m(\cdot)$ being a probability density function in $\mathbb{R}^l$, $\mathcal{J}_{t, m(\cdot), u}$ is a function of y defined by

$$\mathcal{J}_{t, m(\cdot), u}(y) = \Pi_{\Sigma_t(x, \cdot), m(\cdot)}(\Sigma_t(x, y)^{-1}(b_t(x, y) - \operatorname{div}_x \Sigma_t(x, y)/2)),$$

where $\Pi_{\Sigma_t(x, \cdot), m(\cdot)}$ maps a function $\phi(y) \in \mathbb{L}_{\text{loc}}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$, which is the space consisting of functions whose products with arbitrary $\mathcal{C}_0^\infty$ -functions belong to $\mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$, into

$$\Pi_{\Sigma_t(x, \cdot), m(\cdot)} \phi(y) \in \mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$$

and satisfies that, for all $h \in \mathcal{C}_0^\infty(\mathbb{R}^l)$,

$$\int_{\mathbb{R}^l} [\nabla h(y)]^\top \Sigma_t(x, y) \Pi_{\Sigma_t(x, \cdot), m(\cdot)} \phi(y) m(y) dy = \int_{\mathbb{R}^l} [\nabla h(y)]^\top \Sigma_t(x, y) \phi(y) m(y) dy.$$

If $\phi(y) \in \mathbb{L}^2(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$, then $\Pi_{\Sigma_t(x, \cdot), m(\cdot)} \phi(y)$ is the orthogonal projection of ϕ onto $\mathbb{L}_0^{1,2}(\mathbb{R}^l, \mathbb{R}^l, \Sigma_t(x, y), m(y) dy)$. Moreover, it is readily seen from (2.51) that

$$(2.52) \quad \text{if } \mathbb{I}^*(\varphi, \mu) < \infty, \text{ then } \dot{\varphi}_s = \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y)}{\lambda_s(\varphi_s, y)} m_s(y) dy.$$

In addition, the supremum in the last term in (2.51) is attained at

$$(2.53) \quad \widehat{g}(y) = \frac{\nabla_y m_s(y)}{2m_s(y)} - \mathcal{J}_{s, m_s(\cdot), \varphi_s}(y)$$

so that

$$(2.54) \quad \mathbb{I}^*(\varphi, \mu) = \frac{1}{2} \int_0^\infty \int_{\mathbb{R}^l} \left\| \frac{\nabla_y m_s(y)}{2m_s(y)} - \mathcal{J}_{s, m_s(\cdot), \varphi_s}(y) \right\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) dy ds.$$

This combined with (2.14) also shows the equality of $\mathbb{I}$ and $\widehat{I}^*$.

Now, we proceed to the proof, which contains two main steps.

Step (i): Identify the LD limits at sufficiently regular (dense) points.

THEOREM 2.5. *Assume that the assumptions of Theorem 2.1 hold. Let $\widehat{\mathbb{I}}$ be an LD limit point of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon > 0}$ such that $\widehat{\mathbb{I}}(\varphi, \mu) = \infty$ unless $\varphi_0 = \widehat{x}$. Let $(\widehat{\varphi}, \widehat{\mu}) \in \mathcal{G}$ be such that $\widehat{\varphi}_0 = \widehat{x}$ and $\widehat{\mu}(ds, dy) = \widehat{m}_s(dy) ds$, where $\widehat{m}_s(y)$ has the form*

$$(2.55) \quad \hat{m}_s(y) = M_s \left(\tilde{m}_s(y) \hat{\eta}^2 \left(\frac{|y|}{r} \right) + e^{-a|y|} \left(1 - \hat{\eta}^2 \left(\frac{|y|}{r} \right) \right) \right),$$

where $\tilde{m}_s(y)$ is a probability density in y locally bounded away from zero and belonging to $C^1(\mathbb{R}^l)$ as a function of y with $|\nabla m_s(y)|$ being locally bounded in (s, y) , and $\hat{\eta}(y)$ is a nonincreasing $[0, 1]$ -valued $C_0^1(\mathbb{R}_+)$ -function, with $y \in \mathbb{R}_+$, that equals 1 for $y \in [0, 1]$ and equals 0 for $y \geq 2$; $r > 0$ and $a > 0$, and M_s is the normalizing constant. For given $\tilde{m}_s(y)$, $\hat{\eta}(y)$, and r , there exists $a_0 > 0$ such that for all $a > a_0$, $\mathbb{I}(\varphi, \mu) = \mathbb{I}^*(\hat{\varphi}, \hat{\mu})$.

Technical lemmas. Before proving Theorem 2.5, we first need some technical lemmas. For β, h as defined in section 2.2.3, we denote

$$(2.56) \quad \tau_N^{\beta, h}(\varphi, \mu) = \inf \left\{ t \in \mathbb{R}_+ : \int_0^t \int_{\mathbb{R}^l} \|\nabla_y h_s(\varphi_s, y)\|_{\Sigma_s(\varphi_s, y)} \mu(ds, dy) + \sup_{s \in [0, t]} |\varphi_s| + t \geq N \right\}.$$

Performing integration by parts in (2.39) yields that

$$(2.57) \quad \begin{aligned} \Phi_t^{\beta, h}(\varphi, \mu) := & \int_0^t \left(\beta(s) \dot{\varphi}_s - [\beta(s)]^\top \int_{\mathbb{R}^l} \frac{F_s(\varphi_s, y) m_s(y)}{\lambda_s(\varphi_s, y)} \right. \\ & + \int_{\mathbb{R}^l} [\nabla_y h(s, \varphi_s, y)]^\top \left(\frac{1}{2} \operatorname{div}(\Sigma_s(\varphi_s, y) m_s(y)) - b_s(\varphi_s, y) m_s(y) \right) dy \\ & \left. - \frac{1}{2} \int_{\mathbb{R}^l} \|\nabla_y h(s, \varphi_s, y)\|_{\Sigma_s(\varphi_s, y)}^2 m_s(y) dy \right) ds. \end{aligned}$$

Let

$$(2.58) \quad \theta_N^{\beta, h}(\varphi, \mu) := \Phi_{\tau_N^{\beta, h}}^{\beta, h}(\varphi, \mu),$$

and for each $\delta > 0$, $K_\delta := \{(\varphi, \mu) : \hat{\mathbb{I}}(\varphi, \mu) \leq \delta\}$. The following are some technical lemmas needed for the proof of Theorem 2.5.

LEMMA 2.3 ([34, Lemma 7.1]) (approximation of τ, θ). *Assume the following conditions for the boundedness and the convergence (uniformly in K_δ) of $\{\beta_s^i\}_{i=1}^\infty$, $\{h_s^i(x, y)\}_{i=1}^\infty$ to β_s , $h_s(x, y)$:*

$$(2.59) \quad \int_0^N |\beta_s|^2 ds + \int_0^N \operatorname{ess\,sup}_{(\varphi, \mu \in K_\delta)} \int_{\mathbb{R}^l} |\nabla h_s(\varphi_s, y)| m_s(y) dy ds < \infty,$$

$$(2.60) \quad \begin{aligned} & \lim_{i \rightarrow \infty} \int_0^N |\beta_s - \beta_s^i|^2 ds \\ & + \lim_{i \rightarrow \infty} \sup_{(\varphi, \mu) \in K_\delta} \int_0^N \int_{\mathbb{R}^l} |\nabla h_s(\varphi_s, y) - \nabla h_s^i(\varphi_s, y)|^2 m_s(y) dy ds = 0. \end{aligned}$$

Then we have the convergence

$$\tau_N^{\beta^i, h^i}(\varphi, \mu) \xrightarrow{i \rightarrow \infty} \tau_N^{\beta, h}(\varphi, \mu) \quad \text{and} \quad \theta_N^{\beta^i, h^i}(\varphi, \mu) \xrightarrow{i \rightarrow \infty} \theta_N^{\beta, h}(\varphi, \mu) \quad \text{uniformly in } K_\delta.$$

LEMMA 2.4 ([34, Lemma 7.2]) (localizing supremum). *If $h_s(x, y)$ is measurable and belongs to class $\mathbb{W}_{\text{loc}}^{1,1}$ in y , vanishes when y is outside of some open ball in $\mathbb{R}^l$ locally uniformly in (s, x) , and is such that the derivative $Dh_s(x, y)$ is continuous in (x, y) for almost all $s \in \mathbb{R}_+$ and that $\int_0^N \sup_{x \in \mathbb{R}^d: |x| \leq L} \int_{\mathbb{R}^l} |Dh_s(x, y)|^q dy ds < \infty$ for all $q > 1$ and $L > 0$, then, $\sup_{(\varphi, \mu) \in K_\delta} (\theta_N^{\beta, h}(\varphi, \mu) - \hat{\mathbb{I}}(\varphi, \mu)) = 0$, and the supremum is attained.*

LEMMA 2.5 ([34, Lemma 7.3]) (regularities and growth-rate properties of a certain (dense) class). Assume $m_s(y)$ is an $\mathbb{R}_+$ -valued measurable function and is a probability density in y for almost every s . Assume further that it is bounded away from zero on bounded sets of (s, y) and is in $C^1(\mathbb{R}^l)$, with $|\nabla m_s(y)|$ being locally bounded in (s, y) , and that $m_s(y) = M_s e^{-a|y|}$ ($a > 0$) for all $|y|$ large enough locally uniformly in s . There exists an a_0 such that if $a > a_0$, there is a $w_s(x, \cdot)$ such that $\mathcal{J}_{s, m_s(\cdot), u}(\cdot) = \nabla w_s(x, \cdot)$ and satisfies certain regularity and growth-rate properties [34, (7.13)–(7.15)].

Proof of Theorem 2.5. Let a_0 and then $\hat{w}_s(x, y)$ be defined as in Lemma 2.5 for $\hat{m}_s(y)$. Let $\hat{\beta} = 0$ and $\hat{h}_s(x, y) = \frac{1}{2} \ln \hat{m}_s(y) - \hat{w}_s(x, y)$. Then $\nabla \hat{h}_s(x, y) = \frac{\nabla \hat{m}_s(y)}{2\hat{m}_s(y)} - \nabla \hat{w}_s(x, y)$. We want to apply Lemma 2.4 for $\hat{\beta}, \hat{h}_s(x, y)$. However, $\hat{h}_s(x, y)$ might not have a compact support in y . Hence, in order to use Lemma 2.4, we need to restrict it to a compact set. Therefore, we shall truncate $\hat{h}_s(x, y)$. Let $\eta(t)$ represent an $\mathbb{R}_+$ -valued nonincreasing $C_0^\infty(\mathbb{R}_+)$ -function such that $\eta(t) = 1$ for $0 \leq t \leq 1$ and $\eta(t) = 0$ for $t \geq 2$. Let $\hat{w}_s^i(x, y) = \hat{w}_s(x, y)\eta(\frac{|y|}{i})$ and $\hat{h}_s^i(x, y) = \frac{1}{2}\eta(\frac{|y|}{i}) \ln \hat{m}_s(y) - \hat{w}_s^i(x, y)$. As in [34, Lemma 7.4], we can prove that $\hat{h}_s^i(x, y)$ satisfies the conditions in Lemma 2.4.

Next, given $N \in \mathbb{N}$, let $\tau_N^{0, \hat{h}^i}$ and $\theta_N^{0, \hat{h}^i}$ be defined by the respective equations (2.56) and (2.58) with $\beta = 0$ and $h = \hat{h}_s^i(x, y)$. Since the functions $\beta = 0$ and $\hat{h}_s^i(x, y)$ satisfy the hypothesis in Lemma 2.4, there exists $(\varphi^{N, i}, \mu^{N, i}) \in \mathcal{G}$ such that $\theta_N^{0, \hat{h}^i}(\varphi^{N, i}, \mu^{N, i}) = \hat{\mathbb{I}}(\varphi^{N, i}, \mu^{N, i})$ and $(\varphi^{N, i}, \mu^{N, i}) \in K_{2N+2}$ for all i . In particular, $\mu^{N, i}(ds, dy) = m_s^{N, i}(y) dy ds$, where $m_s^{N, i}(\cdot)$ belongs to $\mathcal{P}(\mathbb{R}^l)$, and the set $\{(\varphi^{N, i}, \mu^{N, i}), i = 1, 2, \dots\}$ is relatively compact. Since $\hat{\mathbb{I}}(\varphi^{N, i}, \mu^{N, i}) \geq \mathbb{I}^*(\varphi^{N, i}, \mu^{N, i})$ and $\theta_N^{0, \hat{h}^i}(\varphi^{N, i}, \mu^{N, i}) \leq \mathbb{I}^*(\varphi^{N, i}, \mu^{N, i})$, one has

$$(2.61) \quad \theta_N^{0, \hat{h}^i}(\varphi^{N, i}, \mu^{N, i}) = \mathbb{I}^*(\varphi^{N, i}, \mu^{N, i}) = \hat{\mathbb{I}}(\varphi^{N, i}, \mu^{N, i}).$$

Extract a convergent subsequence (still denoting the index by i) $\mu^{N, i} \rightarrow \mu^N$ in $\mathcal{C}_+(\mathbb{R}_+, \mathcal{M}(\mathbb{R}^l))$ and $\varphi^{N, i} \rightarrow \varphi^N$ in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$.

Because of (2.61) and (2.57), $\mathbb{I}^*(\varphi^{N, i}, \mu^{N, i})$ obtains supremum at $\hat{h}_s^i(x, y)$ when $s \leq \tau_N^{0, \hat{h}^i}(\varphi^{N, i}, \mu^{N, i})$. Therefore, by using (2.53), we can characterize $m_s^{N, i}$ (noting that $\mu^{N, i}(ds, dy) = m_s^{N, i}(y) dy ds$) and then can show that the convergence of (2.59) and (2.60) in the hypothesis of Lemma 2.3 is satisfied (see [34, (7.46)–(7.48)]). Thus, by Lemma 2.3, we have that $\tau_N^{0, \hat{h}^i}(\varphi^{N, i}, \mu^{N, i}) \rightarrow \tau_N^{0, \hat{h}}(\varphi^N, \mu^N)$ as $i \rightarrow \infty$, and that $m_s^{N, i}(y) \rightarrow \hat{m}_s(y)$ in $\mathbb{L}^1([0, \tau_N^{0, \hat{h}}(\varphi^N, \mu^N)] \times \mathbb{R}^l)$. Therefore, $\mu^N(ds, dy) = \hat{m}_s(y) dy ds$ for almost all $s \leq \tau_N^{0, \hat{h}}(\varphi^N, \mu^N)$.

Using $\dot{\varphi}_s^{N, i} = \int_{\mathbb{R}^l} \frac{F_s(\varphi_s^{N, i}, y) m_s^{N, i}(y)}{\lambda_s(\varphi_s^{N, i}, y)} dy$ due to (2.52) and applying [34, Lemma 6.7], we obtain from the convergence of $\varphi^{N, i} \rightarrow \varphi^N$ in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ and $m_s^{N, i}(y) \rightarrow \hat{m}_s(y)$ in $\mathbb{L}^1([0, \tau_N^{0, \hat{h}}(\varphi^N, \mu^N)] \times \mathbb{R}^l)$ as $i \rightarrow \infty$ that $\dot{\varphi}_s^N = \int_{\mathbb{R}^l} \frac{F_s(\varphi_s^N, y) \hat{m}_s(y)}{\lambda_s(\varphi_s^N, y)} dy$ a.e. for $s \leq \tau_N^{0, \hat{h}}(\varphi^N, \mu^N)$. By the uniqueness, $\varphi_s^N = \hat{\varphi}_s$ for $s \leq \tau_N^{0, \hat{h}}(\varphi^N, \mu^N)$. As a byproduct, $\dot{\varphi}_s^{N, i} \rightarrow \dot{\hat{\varphi}}_s$ as $i \rightarrow \infty$ a.e. on $[0, \tau_N^{0, \hat{h}}(\varphi^N, \mu^N)]$.

We have proved that $\tau_N^{0, \hat{h}}(\varphi^N, \mu^N) = \tau_N^{0, \hat{h}}(\hat{\varphi}, \hat{\mu})$ and $\varphi_s^N = \hat{\varphi}_s$, $\mu_s^N = \hat{\mu}_s$ for $s \leq \tau_N^{0, \hat{h}}(\hat{\varphi}, \hat{\mu})$ so that $\theta_N^{0, \hat{h}}(\varphi^N, \mu^N) = \theta_N^{0, \hat{h}}(\hat{\varphi}, \hat{\mu})$. We can show that $\theta_N^{0, \hat{h}}(\varphi^N, \mu^N) = \lim_{i \rightarrow \infty} \theta_N^{0, \hat{h}^i}(\varphi^{N, i}, \mu^{N, i})$. Therefore, taking the limit in (2.61), we have $\mathbb{I}^*(\varphi^N, \mu^N) \geq \theta_N^{0, \hat{h}}(\varphi^N, \mu^N) \geq \hat{\mathbb{I}}(\varphi^N, \mu^N)$, which together with the fact that $\mathbb{I}^* \leq \hat{\mathbb{I}}$ obtained in the previous section implies that $\mathbb{I}^*(\varphi^N, \mu^N) = \theta_N^{0, \hat{h}}(\varphi^N, \mu^N) = \hat{\mathbb{I}}(\varphi^N, \mu^N)$. Therefore, we have for all $N > 0$,

$$(2.62) \quad \mathbb{I}^*(\hat{\varphi}, \hat{\mu}) \geq \theta_N^{0, \hat{h}}(\hat{\varphi}, \hat{\mu}) = \theta_N^{0, \hat{h}}(\varphi^N, \mu^N) = \hat{\mathbb{I}}(\varphi^N, \mu^N).$$

From (2.62), the fact that $\varphi_s^N = \hat{\varphi}_s$, $\mu^N = \hat{\mu}$ until $\tau_N^{0, \hat{h}}(\hat{\varphi}, \hat{\mu})$, and the fact that $\hat{\mathbb{I}}$ is lower semicontinuous and inf-compact, we obtain $\mathbb{I}^*(\hat{\varphi}, \hat{\mu}) \geq \hat{\mathbb{I}}(\hat{\varphi}, \hat{\mu})$. As a result, we can conclude that $\mathbb{I}^*(\hat{\varphi}, \hat{\mu}) = \hat{\mathbb{I}}(\hat{\varphi}, \hat{\mu})$ for all $(\hat{\varphi}, \hat{\mu})$ satisfying the requirements in Theorem 2.5. $\square$

Step (ii). Approximating the LD limits in arbitrary points by regular points. Let $\hat{\mathbb{I}}$ be an LD limit point of $\{(X^\varepsilon, \mu^\varepsilon)\}_{\varepsilon>0}$ and be such that $\hat{\mathbb{I}}(\varphi, \mu) = \infty$ unless $\varphi_0 = \hat{x}$. In this step, it is proven (see [34, Theorem 8.1]) that if $\mathbb{I}^*(\varphi, \mu) < \infty$, then there exists a sequence $(\varphi^{(k)}, \mu^{(k)})$, whose elements have the properties from Theorem 2.5 such that $(\varphi^{(k)}, \mu^{(k)}) \rightarrow (\varphi, \mu)$ and $\mathbb{I}^*(\varphi^{(k)}, \mu^{(k)}) \rightarrow \mathbb{I}^*(\varphi, \mu)$ as $k \rightarrow \infty$. Therefore, one has $\hat{\mathbb{I}}(\varphi, \mu) \leq \mathbb{I}^*(\varphi, \mu) = \lim_{k \rightarrow \infty} \mathbb{I}^*(\varphi^{(k)}, \mu^{(k)}) = \lim_{k \rightarrow \infty} \hat{\mathbb{I}}(\varphi^{(k)}, \mu^{(k)}) \geq \hat{\mathbb{I}}(\varphi, \mu)$. Hence, we have obtained the desired properties in this section.

2.2.6. Completion of the proof of Theorem 2.1. We will complete the proof of Theorem 2.1 by removing the restriction that $\hat{\mathbb{I}}(\varphi, \mu) = \infty$ unless $\varphi_0 = \hat{x}$ in section 2.2.3, where $\hat{x}$ is a preselected element such that $\mathbb{I}_0(\hat{x}) = 0$. This can be done similarly as in [34, section 9] and will be omitted here.

3. Second-order fast-slow systems with general fast random processes.

In this section, we treat (1.2), in which the fast-varying random process ξ_t^ε is under a more general setup without specified structure. We need the following assumptions. At a glance, the conditions may seem to be abstract. Nevertheless, Remark 3.1 illustrates that these assumptions are mild, verifiable, and natural.

Assumption 3.1. The functions $F_t^\varepsilon(x, y)$ and $\lambda_t^\varepsilon(x, y)$ are Lipschitz continuous in x locally uniformly in t and globally uniformly in y , and $\lambda_t^\varepsilon(x, y)$ is bounded below (uniformly) by a positive constant κ_0 . Either $F_t^\varepsilon(x, y)$ and $\lambda_t^\varepsilon(x, y)$ have linear growth in (t, x) globally in y , i.e., there is a universal constant $\tilde{C}$ such that

$$(3.1) \quad |F_t^\varepsilon(x, y)| + |\lambda_t^\varepsilon(x, y)| \leq \tilde{C}(1 + |t| + |x|),$$

or $F_t^\varepsilon(x, y)$ and $\lambda_t^\varepsilon(x, y)$ have linear growth in x locally in y , i.e., the constant $\tilde{C}$ in (3.1) is uniformly in bounded sets of y and ξ_t^ε is such that for any $T > 0$

$$(3.2) \quad \lim_{L \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P} \left(\sup_{0 \leq t \leq T} |\xi_t^\varepsilon| > L \right) = -\infty.$$

DEFINITION 3.1. A family of stochastic processes $\{X^\varepsilon\}_{\varepsilon>0}$ is said to satisfy the local LDP in $\mathcal{C}([0, 1], \mathbb{R}^d)$ with rate function $\mathbb{J}$ if for any $\varphi \in \mathcal{C}([0, 1], \mathbb{R}^d)$,

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}(X^\varepsilon \in B(\varphi, \delta)) \\ &= \lim_{\delta \rightarrow 0} \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}(X^\varepsilon \in B(\varphi, \delta)) = -\mathbb{J}(\varphi), \end{aligned}$$

where $B(\varphi, \delta)$ is the ball centered at φ with radius δ in $\mathcal{C}([0, 1], \mathbb{R}^d)$. $\mathbb{J}$ is called local rate function.

Assumption 3.2. The family of processes $\{Z^\varepsilon\}_{\varepsilon>0}$ given by

$$(3.3) \quad \dot{Z}_t^\varepsilon = \frac{F_t^\varepsilon(Z_t^\varepsilon, \xi_t^\varepsilon)}{\lambda_t^\varepsilon(Z_t^\varepsilon, \xi_t^\varepsilon)}, \quad Z_0^\varepsilon = x_0^\varepsilon,$$

satisfies the local LDP with a rate function $\mathbb{J}$.

Remark 3.1. Seemingly abstract, Assumption 3.2 is not restrictive. In fact, it is the LDP for the first-order systems, which is relatively well understood now. For example, the condition is verified when ξ_t^ε is a (fast-varying) diffusion process

$$d\xi_t^\varepsilon = \frac{1}{\varepsilon} b_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) dt + \frac{1}{\sqrt{\varepsilon}} \sigma_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) dW_t,$$

where W_t is a standard Brownian motion; see [27, 43, 44]. It is also verified when ξ_t^ε is a (fast-varying) Markovian switching process with generator $Q(t)/\varepsilon$ and $Q(t)$ being a time-inhomogeneous and irreducible generator of a Markov chain, or when ξ_t^ε is a (fast-varying) jump process having jumps at rate $O(\varepsilon^{-1})$; see [2]. Furthermore, the condition is verified when ξ_t^ε has no specific representation but satisfies exponential ergodicity [18]. Note also that condition (3.2) in Assumption 3.1 is essentially an exponential tightness of the fast processes, which is readily verified for ξ_t^ε being diffusion processes or Markovian switching processes. When we deal with general fast processes without any specific formulation, the assumption on tightness (3.2) and Assumption 3.2 are very natural. Without the tightness and ergodicity of the fast processes, it is unlikely one can obtain the averaging and LDPs for a fast-slow system under the setting of general fast processes.

Remark 3.2. We did not assume any regularity and growth-rate conditions of the coefficients of the slow component when dealing with (1.1). However, for the general fast random process, it seems to be impossible to use the same assumptions because we do not require any structure for the fast process. As a result, the assumptions in this section are stronger than that of section 2. In particular, we need the Lipschitz continuity and growth-rate conditions of $F_t^\varepsilon(x, y)$. It is worth noting that we used two totally different approaches for the cases of fast diffusions and general fast-varying processes. If the fast process is a diffusion, thanks to the nice structure of martingales, we can identify the rate function after estimating the exponential moment. Therefore, in the first case, after obtaining the exponential tightness and then relatively LD compactness (see Definition 2.3), our remaining work is to identify the rate functions. In the general case, we use a different approach that relies on the property that exponential tightness and the local LDP imply the full LDP. In this situation, we need to connect directly the solutions of the second-order and the first-order equations.

We are now in a position to present the main theorem. The result is stated next, and the proof is given in the next section.

THEOREM 3.1. *Assume that Assumptions 3.1 and 3.2 hold, that the family $\{x_0^\varepsilon\}_{\varepsilon>0}$ is exponentially tight, and that $\limsup_{\varepsilon \rightarrow 0} \varepsilon |x_1^\varepsilon| < \infty$ a.s. Then, the family $\{X^\varepsilon\}_{\varepsilon>0}$ of (1.2) obeys the LDP in $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ with rate function $\mathbb{J}$.*

3.1. Proof of Theorem 3.1. The proof of this theorem is based on the fact that the exponential tightness and local LDP imply the full LDP. The following result is well known in LD theory; see, e.g., [11, 10, 28].

PROPOSITION 3.1. *The exponential tightness and the local LDP for a family $\{X^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{C}([0, 1], \mathbb{R}^d)$ with local rate function $\mathbb{J}$ imply the full LDP in $\mathcal{C}([0, 1], \mathbb{R}^d)$ for this family with rate function $\mathbb{J}$.*

In what follows, we prove the LDP of $\{X^\varepsilon\}_{\varepsilon>0}$ in $\mathcal{C}([0, 1], \mathbb{R}^d)$. It will be seen that it can be extended to the space $\mathcal{C}([0, T], \mathbb{R}^d)$ endowed with the sup-norm topology for any $T > 0$. As a consequence, the LDP still holds in $\mathcal{C}([0, \infty), \mathbb{R}^d)$, the space of

continuous function on $[0, \infty)$ endowed with the local supremum topology. (This fact follows from the Dawson–Gärtner theorem; see [10, Theorem 4.6.1], which states that it suffices to check the LDPs in $\mathcal{C}([0, T], \mathbb{R}^d)$ for any T in the uniform metric.) We will still use C to represent a generic positive constant that is independent of ε . The value C may change at different appearances; if it is necessary, we will specify which parameters it depends on.

Exponential tightness. We aim to prove (2.16) and (2.17). We have

$$(3.4) \quad X_t^\varepsilon = x_0^\varepsilon + \int_0^t x_1^\varepsilon e^{-A_\varepsilon^\xi(s)} ds + \frac{1}{\varepsilon^2} \int_0^t \int_0^s e^{-A_\varepsilon^\xi(s,r)} F_s^\varepsilon(X_r^\varepsilon, \xi_r^\varepsilon) dr,$$

where for any $0 \leq s \leq t \leq 1, \varepsilon > 0$, $A_\varepsilon^\xi(t, s) := \frac{1}{\varepsilon^2} \int_s^t \lambda_r^\varepsilon(X_r^\varepsilon, \xi_r^\varepsilon) dr$ and $A_\varepsilon^\xi(t) = A_\varepsilon^\xi(t, 0)$. So, we can obtain from some direct calculations and Assumption 3.1 that

$$(3.5) \quad |X_t^\varepsilon| \leq |x_0^\varepsilon| + C\varepsilon^2|x_1^\varepsilon| + C \int_0^t \sup_{r \in [0, s]} |F_r^\varepsilon(X_r^\varepsilon, \xi_r^\varepsilon)| ds,$$

and by noting further that $\int_s^t e^{-\frac{\kappa_0 r}{\varepsilon^2}} dr \leq C\varepsilon^2(1 - e^{-\frac{t-s}{\varepsilon^2}}) \leq C\varepsilon\sqrt{|t-s|}$, we get

$$(3.6) \quad |X_t^\varepsilon - X_s^\varepsilon| \leq C\varepsilon|x_1^\varepsilon|\sqrt{|t-s|} + C|t-s| \sup_{r \in [s, t]} |F_r^\varepsilon(X_r^\varepsilon, \xi_r^\varepsilon)|.$$

If (3.1) in Assumption 3.1 holds, (2.16) follows immediately from (3.5) and Gronwall's inequality on noting that $\limsup_{\varepsilon \rightarrow 0} \varepsilon|x_1^\varepsilon| < \infty$ a.s., $\{x_0^\varepsilon\}_{\varepsilon > 0}$ is exponentially tight; then (2.17) follows from (2.16) and (3.6).

Otherwise, assume that (3.2) holds. Let $\tilde{C}_N$ be constant in (3.1) uniformly in $|y| < N$. We get from (3.5) that $\sup_{t \in [0, 1]} |X_t^\varepsilon| \leq C(\tilde{C}_N + N)e^{\tilde{C}_N}$ provided $\sup_{t \in [0, 1]} |\xi_t^\varepsilon| < N$, $|x_0^\varepsilon| < N$. Therefore, for any $N > 0$, for $L > C(\tilde{C}_N + N)e^{\tilde{C}_N}$ one has

$$\mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |X_t| > L \right) \leq \mathbb{P}^\varepsilon(|x_0^\varepsilon| > N) + \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |\xi_t^\varepsilon| > N \right).$$

Letting $L \rightarrow \infty$ and $N \rightarrow \infty$ and using the logarithm equivalence principle [10, Lemma 1.2.15] and (3.2) in Assumption 3.1, we get (2.16). Thus, we also obtain (2.17).

Local LDP. We note that we do not assume any structure of ξ_t^ε . As a result, we could not use integration by parts (Lemma 2.1) to connect the first-order and second-order systems. Therefore, we will establish a relationship in a local sense.

For each continuous function φ , we introduce the auxiliary processes $X_t^{\varepsilon, \varphi}$, the solution of

$$(3.7) \quad \varepsilon^2 \ddot{X}_t^{\varepsilon, \varphi} = F_t^\varepsilon(\varphi_t, \xi_t^\varepsilon) - \lambda_t^\varepsilon(\varphi_t, \xi_t^\varepsilon) \dot{X}_t^{\varepsilon, \varphi}, \quad X_0^\varepsilon = x_0^\varepsilon, \quad \dot{X}_0^\varepsilon = x_1^\varepsilon,$$

and $Z_t^{\varepsilon, \varphi}$, the solution of

$$(3.8) \quad \dot{Z}_t^{\varepsilon, \varphi} = \frac{F_t^\varepsilon(\varphi_t, \xi_t^\varepsilon)}{\lambda_t^\varepsilon(\varphi_t, \xi_t^\varepsilon)}, \quad Z_0^{\varepsilon, \varphi} = x_0^\varepsilon.$$

We have from (3.7) and the variation of parameters formula that

$$(3.9) \quad X_t^{\varepsilon, \varphi} = x_0^\varepsilon + \int_0^t x_1^\varepsilon e^{-A_{\varepsilon, \varphi}^\xi(s)} ds + \frac{1}{\varepsilon^2} \int_0^t \int_0^s e^{-A_{\varepsilon, \varphi}^\xi(s,r)} F_s^\varepsilon(\varphi_r, \xi_r^\varepsilon) dr,$$

where for any $0 \leq s \leq t \leq 1, \varepsilon > 0$, $A_{\varepsilon, \varphi}^{\varepsilon}(t, s) := \frac{1}{\varepsilon^2} \int_s^t \lambda_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon}) dr$ and $A_{\varepsilon, \varphi}^{\varepsilon}(t) = A_{\varepsilon, \varphi}^{\varepsilon}(t, 0)$. From the fact that $A_{\varepsilon}^{\varepsilon}(s, r), A_{\varepsilon, \varphi}^{\varepsilon}(s, r) \geq \frac{\kappa_0(s-r)}{\varepsilon^2}$ and the property of λ , we obtain that

$$(3.10) \quad \begin{aligned} \left| e^{-A_{\varepsilon}^{\varepsilon}(s, r)} - e^{-A_{\varepsilon, \varphi}^{\varepsilon}(s, r)} \right| &\leq C e^{-\frac{\kappa_0(s-r)}{\varepsilon^2}} \cdot \frac{1}{\varepsilon^2} \int_r^s |X_r^{\varepsilon} - \varphi_r| dr \\ &\leq C e^{-\frac{\kappa_0(s-r)}{\varepsilon^2}} \cdot \frac{s-r}{\varepsilon^2} \cdot \sup_{r \in [0, s]} |X_r^{\varepsilon} - \varphi_r|. \end{aligned}$$

A change of variable leads to

$$(3.11) \quad \int_0^s \exp \left\{ \frac{-\kappa_0(s-r)}{\varepsilon^2} \right\} \cdot \frac{s-r}{\varepsilon^2} ds = \varepsilon^2 \int_0^{\frac{s}{\varepsilon^2}} e^{-\kappa_0 r} r dr \leq C \varepsilon^2.$$

Therefore, we obtain from the Lipschitz property of the coefficients and (3.10) that

$$(3.12) \quad \begin{aligned} &\int_0^s \left| e^{-A_{\varepsilon}^{\varepsilon}(s, r)} F_r^{\varepsilon}(X_r^{\varepsilon}, \xi_r^{\varepsilon}) - e^{-A_{\varepsilon, \varphi}^{\varepsilon}(s, r)} F_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon}) \right| dr \\ &\leq \int_0^s e^{-A_{\varepsilon}^{\varepsilon}(s, r)} |F_r^{\varepsilon}(X_r^{\varepsilon}, \xi_r^{\varepsilon}) - F_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})| dr \\ &\quad + \int_0^s |F_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})| \left| e^{-A_{\varepsilon}^{\varepsilon}(s, r)} - e^{-A_{\varepsilon, \varphi}^{\varepsilon}(s, r)} \right| dr \\ &\leq C \varepsilon^2 \sup_{0 \leq r \leq s} |X_r^{\varepsilon} - \varphi_r| + C \varepsilon^2 \sup_{r \in [0, s]} |F_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})| \sup_{0 \leq r \leq s} |X_r^{\varepsilon} - \varphi_r|. \end{aligned}$$

Combining (3.4) and (3.9), applying (3.12), and noting that $\limsup_{\varepsilon \rightarrow 0} \varepsilon |x_1^{\varepsilon}| < \infty$ leads to

$$(3.13) \quad \sup_{s \in [0, t]} |X_s^{\varepsilon} - X_s^{\varepsilon, \varphi}| \leq C \varepsilon \sup_{s \in [0, t]} |X_s^{\varepsilon} - \varphi_s| + C \sup_{r \in [0, t]} |F_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})| \int_0^t \sup_{r \in [0, s]} |X_r^{\varepsilon} - \varphi_r| ds.$$

From (3.7) and (3.8), we obtain that

$$(3.14) \quad \begin{aligned} |X_t^{\varepsilon, \varphi} - Z_t^{\varepsilon, \varphi}| &= \left| \int_0^t \frac{\varepsilon^2 \dot{X}_r^{\varepsilon, \varphi}}{\lambda_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})} dr \right| \leq C \varepsilon^2 \sup_{s \in [0, t]} |\dot{X}_t^{\varepsilon, \varphi}| \\ &\leq C \varepsilon^2 \left(|x_1^{\varepsilon}| + \sup_{r \in [0, t]} |F_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})| \right). \end{aligned}$$

One also gets from (3.3), (3.8), and the Lipschitz continuity of $F^{\varepsilon}, \lambda^{\varepsilon}$ that

$$(3.15) \quad \sup_{s \in [0, t]} |Z_s^{\varepsilon, \varphi} - Z_s^{\varepsilon}| \leq C \left(\sup_{r \in [0, t]} |F_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})| + \sup_{r \in [0, t]} |\lambda_r^{\varepsilon}(\varphi_r, \xi_r^{\varepsilon})| \right) \int_0^t \sup_{r \in [0, s]} |Z_r^{\varepsilon} - \varphi_r| ds.$$

Now, if (3.1) in Assumption 3.1 holds, combining (3.13), (3.14), and (3.15), we get

$$\begin{aligned} \sup_{s \in [0, t]} |X_s^{\varepsilon} - \varphi_s| &\leq C \varepsilon \sup_{s \in [0, t]} |X_s^{\varepsilon} - \varphi_s| + C \varepsilon \left(1 + \sup_{r \in [0, 1]} |\varphi_r| \right) \\ &\quad + C \left(1 + \sup_{r \in [0, 1]} |\varphi_r| \right) \sup_{r \in [0, 1]} |Z_r^{\varepsilon} - \varphi_r| \\ &\quad + C \left(1 + \sup_{r \in [0, 1]} |\varphi_r| \right) \int_0^t \sup_{r \in [0, s]} |X_r^{\varepsilon} - \varphi_r| ds \end{aligned}$$

and

$$\begin{aligned} \sup_{s \in [0, t]} |Z_s^\varepsilon - \varphi_s| &\leq C\varepsilon \left(1 + \sup_{r \in [0, 1]} |\varphi_r| \right) + C \left(1 + \sup_{r \in [0, 1]} |\varphi_r| \right) \sup_{r \in [0, 1]} |Z_r^\varepsilon - \varphi_r| \\ &\quad + C \left(1 + \sup_{r \in [0, 1]} |\varphi_r| \right) \int_0^t \sup_{r \in [0, s]} |X_r^\varepsilon - \varphi_r| ds. \end{aligned}$$

Thus, for small ε , one has

$$(3.16) \quad \begin{aligned} \sup_{t \in [0, 1]} |X_t^\varepsilon - \varphi_t| &\leq C_{1, \varphi} \varepsilon + C_\varphi \sup_{t \in [0, 1]} |Z_t^\varepsilon - \varphi_t| \quad \text{and} \\ \sup_{t \in [0, 1]} |Z_t^\varepsilon - \varphi_t| &\leq C_{2, \varphi} \varepsilon + C_\varphi \sup_{t \in [0, 1]} |X_t^\varepsilon - \varphi_t|, \end{aligned}$$

for some constants $C_{1, \varphi}$, $C_{2, \varphi}$ depending only on $\sup_{r \in [0, 1]} |\varphi_r|$ and independent of ε . So, for any $\delta > 0$ we have from (3.16) that

$$\begin{aligned} \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |X_t^\varepsilon - \varphi_t| < \delta \right) &\leq \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |Z_t^\varepsilon - \varphi_t| < 2\delta C_{2, \varphi} \right) \quad \forall \varepsilon < \delta / C_{2, \varphi} \quad \text{and} \\ \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |X_t^\varepsilon - \varphi_t| < \delta \right) &\geq \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |Z_t^\varepsilon - \varphi_t| < \frac{\delta}{2C_{1, \varphi}} \right) \quad \forall \varepsilon < \delta / C_{1, \varphi}. \end{aligned}$$

Therefore, the local LDP of $\{X^\varepsilon\}_{\varepsilon > 0}$ follows directly from the local LDP of $\{Z^\varepsilon\}_{\varepsilon > 0}$.

If (3.1) in Assumption 3.1 only holds locally and (3.2) holds, then in the event $\sup_{t \in [0, 1]} |\xi_t^\varepsilon| < N$, we still have (3.16) with $C_{1, \varphi}$, $C_{2, \varphi}$ replaced by $C_{1, \varphi, N}$, $C_{2, \varphi, N}$ depending only on φ, N . So, for any $\delta > 0$ one also has that for all $\varepsilon < \delta / C_{1, \varphi, N} \wedge \delta / C_{2, \varphi, N}$,

$$\begin{aligned} \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |X_t^\varepsilon - \varphi_t| < \delta \right) \\ \leq \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |Z_t^\varepsilon - \varphi_t| < 2\delta C_{2, \varphi, N} \right) + \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |\xi_t^\varepsilon| > N \right) \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |X_t^\varepsilon - \varphi_t| < \delta \right) \\ \geq \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |Z_t^\varepsilon - \varphi_t| < \frac{\delta}{2C_{1, \varphi, N}} \right) - \mathbb{P}^\varepsilon \left(\sup_{t \in [0, 1]} |\xi_t^\varepsilon| > N \right). \end{aligned}$$

By letting $\varepsilon \rightarrow 0$, $N \rightarrow \infty$, and $\delta \rightarrow 0$ and using the logarithm equivalence principle [10, Lemma 1.2.15] and (3.2) in Assumption 3.1, we obtain the local LDP for $\{X^\varepsilon\}_{\varepsilon > 0}$. Therefore, the proof of Theorem 3.1 is complete.

4. Examples. In this section, we consider some examples drawn from physics to illustrate our formulation and results in these cases.

4.1. Stochastic acceleration with small-mass particles. Stochastic acceleration considers motions of a net of particles in a net of random force fields, which is described by Newton's law as $\ddot{x}_\varepsilon(t) = \tilde{F}_\varepsilon(t, \omega, x_\varepsilon(t), \dot{x}_\varepsilon(t), \chi_\varepsilon(t))$, where $\tilde{F}_\varepsilon$ denotes

the random force fields. Such models were considered by Kesten and Papanicolaou in [22, 23] and references therein. Note that small and large are relative terms. Here we focus on small-mass particles, by which we mean that the Reynolds number is small (see, e.g., [35] for a definition) so that inertial effects are negligible compared to the damping force, or the ratio “inertial effects/damping force” is parameterized by $\varepsilon \ll 1$. Therefore, random force field $\tilde{F}_\varepsilon$ can be written as $\tilde{F}_\varepsilon = F_\varepsilon - \frac{\lambda_\varepsilon}{\varepsilon}$, and the motion is described by $\ddot{x}_\varepsilon(t) = F_\varepsilon(t, x_\varepsilon(t), \chi_\varepsilon(t)) - \frac{\lambda_\varepsilon(t, x_\varepsilon(t), \chi_\varepsilon(t))}{\varepsilon} \dot{x}_\varepsilon(t)$. Now, by scaling $X_t^\varepsilon := x_\varepsilon(t/\varepsilon)$ and $\xi_t^\varepsilon := \chi_\varepsilon(t/\varepsilon)$, the system can be rewritten as (1.2), i.e.,

$$(4.1) \quad \varepsilon^2 \ddot{X}_t^\varepsilon = F_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) - \lambda_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) \dot{X}_t^\varepsilon, \quad X_0^\varepsilon = x_0 \in \mathbb{R}^d, \quad \dot{X}_0^\varepsilon = x_1 \in \mathbb{R}^d.$$

The above illustrates how the fast-varying process (or fast-varying random environment) $\xi_t^\varepsilon = \chi_\varepsilon(t/\varepsilon)$ comes in. To further demonstrate, we consider some common models of the fast-varying processes for random environment ξ_t^ε and illustrate our results.

4.1.1. Fast-varying diffusion. Consider the case that fast-varying process ξ_t^ε is modeled as a (fast) diffusion, which is common in modeling stochastic processes in physical phenomena, i.e.,

$$(4.2) \quad \dot{\xi}_t^\varepsilon = \frac{1}{\varepsilon} b_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) + \frac{1}{\sqrt{\varepsilon}} \sigma_t^\varepsilon(X_t^\varepsilon, \xi_t^\varepsilon) \dot{W}_t, \quad \xi_0^\varepsilon = \xi_0 \in \mathbb{R}^l.$$

In this situation, stochastic acceleration (4.1)–(4.2) become coupled second-order SDEs (1.1). The LDP for stochastic acceleration in this case is established in Theorem 2.1 with rate function given in variational form (2.10) or representation (2.14). It is important to note that we establish LDP for stochastic acceleration assuming neither Lipschitz continuity nor linear growth rate of F^ε ; see Assumptions 2.1 and 2.2 and Remarks 2.1.

THEOREM 4.1. *Under Assumptions 2.1 and 2.2 and (2.8), stochastic acceleration under fast-varying diffusion environment (4.1)–(4.2) obeys LDP with the rate function given in (2.10).*

4.1.2. Fast-varying jumps. Consider the case that ξ_t is a jump process taking finite values in $\mathcal{M} = \{1, \dots, |\mathcal{M}|\}$, where $|\mathcal{M}|$ denotes the cardinality of the set $\mathcal{M}$. Similarly to [2], the evolution of the jump fast component is constructed through a jump intensity function $c(x, y) = c_y(x) : \mathbb{R}^d \times \mathcal{M} \rightarrow [0, \infty)$ and a transition probability function $r(x, y, y') = r_{yy'}(x) : \mathbb{R}^d \times \mathcal{M} \times \mathcal{M} \rightarrow [0, 1]$, both of which are coupled with X^ε . We describe the construction of jump processes ξ_t^ε as follows.

Assume that for all $(x, y) \in \mathbb{R}^d \times \mathcal{M}$, $\sum_{y' \in \mathcal{M}} r_{yy'}(x) = 1, r_{yy}(x) = 0$. Let $\zeta = \sup_{(x, y) \in \mathbb{R}^d \times \mathcal{M}} c_y(x) + 1$, $E_{yy'}(x) = [0, c_y(x)r_{yy'}(x)]$ for all $(x, y, y') \in \mathbb{R}^d \times \mathcal{M} \times \mathcal{M}$, $y \neq y'$, and $\mathbb{T} =: \{(y, y') \in \mathcal{M} \times \mathcal{M} : r_{yy'}(x) > 0 \text{ for some } x \in \mathbb{R}^d\}$. For $(i, j) \in \mathbb{T}$, let $\bar{N}_{ij}$ be a Poisson random measure on $[0, \zeta] \times [0, T] \times \mathbb{R}_+$ with intensity measure $\mu_\zeta \otimes \mu_T \otimes \mu_\infty$, where μ_T and μ_∞ denote the Lebesgue measures on $[0, T]$ and $\mathbb{R}_+$, respectively, such that for $t \in [0, T]$, $\bar{N}_{ij}(A \times [0, t] \times B) - t\mu_\zeta(A)\mu_\infty(B)$ is an $\mathcal{F}_t$ -martingale for all $A \in \mathcal{B}[0, \zeta]$ and $B \in \mathcal{B}(\mathbb{R}_+)$ with $\mu_\infty(B) < 1$. Then, we define $N_{ij}^{\varepsilon^{-1}}(dr \times dt) = \bar{N}_{ij}(dr \times dt \times [0, \varepsilon^{-1}])$, which is a Poisson random measure on $[0, \zeta] \times [0, T]$ with intensity $\varepsilon^{-1}\mu_\zeta \otimes \mu_T$. The processes $(N_{ij}^{\varepsilon^{-1}})_{(i, j) \in \mathbb{T}}$ are taken to be mutually independent. We will assume that for $0 \leq s \leq t \leq T$, $\{N_{ij}^{\varepsilon^{-1}}(A \times (s, t] \times B) : A \in \mathcal{B}[0, \zeta], B \in \mathcal{B}(\mathbb{R}_+), (i, j) \in \mathbb{T}\}$ is independent of $\mathcal{F}_s$. Now, we consider the following stochastic acceleration with fast-varying jumps:

$$(4.3) \quad \begin{cases} \varepsilon^2 \ddot{X}_t^\varepsilon = F^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) - \lambda^\varepsilon(X_t^\varepsilon, Y_t^\varepsilon) \dot{X}_t^\varepsilon, \\ dY_t^\varepsilon = \sum_{(i,j) \in \mathbb{T}} \int_{r \in [0, \zeta]} (j - i) \mathbf{1}_{\{Y^\varepsilon(t-) = i\}} \mathbf{1}_{E_{ij}(X_t^\varepsilon)}(r) N_{ij}^{\varepsilon^{-1}}(dr \times dt), \\ X_0^\varepsilon = x_0 \in \mathbb{R}^d, \quad \dot{X}_0^\varepsilon = x_1 \in \mathbb{R}^d, \quad Y_0^\varepsilon = y_0 \in \mathcal{M}. \end{cases}$$

According to [2], we make following assumption for the jump process.

Assumption 4.1. The function c is bounded, and there exists a finite constant $C > 0$ such that for all $y, y' \in \mathcal{M}$ and $x_1, x_2 \in \mathbb{R}^d$,

$$|c_y(x_1) - c_y(x_2)| + |r_{yy'}(x_1) - r_{yy'}(x_2)| \leq C|x_1 - x_2|.$$

Moreover,

$$\inf_{x \in \mathbb{R}^d} \min_{y, z \in \mathcal{M}} \sum_{n=1}^{|\mathcal{M}|} r_{yz}^n(x) > 0, \quad \inf_{x \in \mathbb{R}^d} \min_{y \in \mathcal{M}} c_y(x) > 0, \quad \inf_{x \in \mathbb{R}^d} \min_{(y, y') \in \mathbb{T}} r_{yy'}(x) > 0.$$

The rate function for the LDP of (4.3) is constructed as follows. For $\psi = (\psi(j))_{j \in \mathcal{M}}$, with $\psi_j : [0, \zeta] \rightarrow \mathbb{R}_+$ being a measurable map for every j , define

$$\Phi_{ij}^\psi(x) = \begin{cases} \int_{E_{ij}(x)} \psi_j(z) \mu_\zeta dz & \text{if } i \neq j, \\ -\sum_{y: y \neq j} \Phi_{jy}^\psi(x) & \text{if } i = j, \end{cases}$$

and $\mathcal{R} = \{v = (v_{ij})_{(i,j) \in \mathbb{T}}, v_{ij} : [0, 1] \times [0, \zeta] \rightarrow \mathbb{R}_+ \text{ is measurable for all } (i, j) \in \mathbb{T}\}$. For $\varphi \in C([0, 1], \mathbb{R}^d)$, let $\mathcal{V}(\varphi)$ be the collection of all

$$(u = (u_i), v = (v_{ih}), \pi = (\pi_i)) \in \mathbb{M}([0, 1] : \mathbb{R}^d)^{|\mathcal{M}|} \times \mathcal{R} \times \mathbb{M}([0, 1] : \mathcal{P}(\mathcal{M})),$$

where $\mathbb{M}([0, 1] : \mathcal{P}(\mathcal{M}))$, $\mathbb{M}([0, 1] : \mathbb{R}^d)$ denote the space of measurable maps from $[0, 1]$ to $\mathcal{P}(\mathcal{M})$ and from $[0, 1]$ to $\mathbb{R}^d$, respectively, with $\mathcal{P}(\mathcal{M})$ being the space of probability measures on $\mathcal{M}$ equipped with the topology of weak convergence], such that $\int_0^1 \|u_i(s)\|^2 \pi_i(s) ds < \infty$ for each $i \in \mathcal{M}$, and

$$\varphi_t = x_0 + \sum_{j \in \mathcal{M}} \int_0^t \frac{F(\varphi_s, j)}{\lambda(\varphi_s, j)} \pi_j(s) ds; \quad \sum_{j \in \mathcal{M}} \pi_j(s) \Phi_{ji}^{v_{j \cdot}(s, \cdot)}(\varphi_s) = 0, \text{ a.e. } s \in [0, 1], \forall i \in \mathcal{M}.$$

Combining Theorem 3.1 and [2] yields the following result.

THEOREM 4.2. *Assume Assumptions 3.1 and 4.1 hold. Then the family of processes $\{X^\varepsilon\}_{\varepsilon > 0}$ in stochastic acceleration system with fast-varying jump (4.3) satisfies the LDP with the rate function $\mathbb{I}$ given by*

$$(4.4) \quad \mathbb{I}(\varphi) = \inf_{(u, v, \pi) \in \mathcal{V}(\varphi)} \left\{ \sum_{i \in \mathcal{M}} \frac{1}{2} \int_0^1 \|u_i(s)\|^2 \pi_i(s) ds + \sum_{(i,j) \in \mathbb{T}} \int_{[0, \zeta] \times [0, 1]} \ell(v_{ij}(s, z)) \pi_i(s) \mu_\zeta(dz) ds \right\},$$

where $\ell(x) = x \ln x - x + 1$.

4.2. Liénard equation with relaxation oscillations. The Liénard equations, named after physicist Alfred-Marie Liénard, have been extensively studied in the literature of ODEs. During the development of radio and vacuum tubes, the Liénard equations were used to model oscillating circuits. These equations were also used in

mechanical systems in physics and engineering. In the exploration of radio and vacuum tube technologies, much attention was devoted to the study of Liénard equations and such equations with relaxation oscillations. A notable important equation is

$$(4.5) \quad \frac{1}{\nu^2} \ddot{x}^\nu(t) = g(x^\nu(t)) - \kappa \dot{x}^\nu(t),$$

where $\nu \gg 1$ is a large number, κ is a positive constant, and g is a function. Equation (4.5) has been studied in detail in [30] with the motivation from the familiar van der Pol equation [37]. Its variations can also be found in [38] and references therein.

Now, we consider the case that the environment is perturbed by random factors so that the function g and coefficient κ depend on a random process, which varies very quickly. Such a fast-slow setting is natural as multiscale systems arise in many problems in various fields. For example, many processes (e.g., signals, cellular processes) are inherently multiscale in nature with reactions occurring at varying speeds. As a result, we consider the following Liénard equation with relaxation oscillations in a fast-varying random environment:

$$(4.6) \quad \frac{1}{\nu^2} \ddot{x}^\nu(t) = g(x^\nu(t), \xi^\nu(t)) - \kappa(\xi^\nu(t)) \dot{x}^\nu(t),$$

where $\xi^\nu(t)$ is a (fast-varying) random process, which interacts with $x^\nu(t)$. In particular, the time-scale separation comes from applications; see, for example, [30] and references therein. Using our results, we can establish LDP for the family of solutions $\{x^\nu(\cdot)\}_{\nu \gg 1}$ of (4.6). (i) If $\xi^\nu(t)$ has the form of a (fast) diffusion, LDP of $\{x^\nu(\cdot)\}_{\nu \gg 1}$ is established by Theorem 2.1 without any assumption about Lipschitz continuity of g . (ii) If $\xi^\nu(t)$ is a (fast) jump process, LDP of $\{x^\nu(\cdot)\}_{\nu \gg 1}$ can be obtained as in section 4.1.2 (Theorem 4.2). For brevity, we only state the results without the verbatim derivations.

THEOREM 4.3.

- (i) Assume $d\xi^\nu(t) = \nu b_t^\nu(x^\nu(t), \xi^\nu(t)) + \sqrt{\nu} \sigma_t^\nu(x_t^\nu, \xi^\nu(t)) dW(t)$, $\xi_0^\nu = \xi_0 \in \mathbb{R}^l$. Under Assumptions 2.1 and 2.2 and (2.8), $\{x^\nu(\cdot)\}_{\nu \gg 1}$ satisfies LDP with the rate function

$$(4.7) \quad \begin{aligned} \mathbb{I}_X(\varphi) = & \mathbb{I}_0(\varphi_0) + \int_0^\infty \sup_{\beta \in \mathbb{R}^d} \left[\beta^\top \dot{\varphi}_s - \sup_{m \in \mathcal{P}(\mathbb{R}^l)} \left(\beta^\top \int_{\mathbb{R}^l} \frac{g(\varphi_s, y)}{\kappa(y)} m(y) dy \right. \right. \\ & + \sup_{h \in \mathcal{C}_0^1(\mathbb{R}^l)} \int_{\mathbb{R}^l} \left([\nabla h(y)]^\top \left(\frac{1}{2} \operatorname{div}_y (\Sigma_s(\varphi_s, y) m(y)) - b_s(\varphi_s, y) m(y) \right) \right. \\ & \left. \left. - \frac{1}{2} \|\nabla h(y)\|_{\Sigma_s(\varphi_s, y)}^2 m(y) \right) dy \right] ds \end{aligned}$$

if φ is absolutely continuous; otherwise, $\mathbb{I}_X(\varphi) = \infty$.

- (ii) Assume $d\xi^\nu(t) = \sum_{(i,j) \in \mathbb{T}} \int_{r \in [0, \zeta]} (j - i) \mathbf{1}_{\{\xi^\nu(t-) = i\}} \mathbf{1}_{E_{ij}(x_t^\nu)}(r) N_{ij}^\nu(dr \times dt)$, where N_{ij}^ν is a Poisson random measure with (fast) intensity rate $O(\nu)$ constructed precisely as in section 4.1.2. Under Assumptions 3.1 and 4.1, $\{x^\nu\}_{\nu \gg 1}$ satisfies the LDP with the rate function $\mathbb{I}$ given by

$$(4.8) \quad \begin{aligned} \mathbb{I}(\varphi) = & \inf_{(u, v, \pi) \in \mathcal{V}(\varphi)} \left\{ \sum_{i \in \mathcal{M}} \frac{1}{2} \int_0^1 \|u_i(s)\|^2 \pi_i(s) ds \right. \\ & \left. + \sum_{(i,j) \in \mathbb{T}} \int_{[0, \zeta] \times [0, 1]} \ell(v_{ij}(s, z)) \pi_i(s) \mu_\zeta(dz) ds \right\}, \end{aligned}$$

where $\ell(x) = x \ln x - x + 1$, and $\mathcal{V}(\varphi)$ is the collection of all $(u = (u_i), v = (v_{ih}), \pi = (\pi_i))$ such that $\int_0^1 \|u_i(s)\|^2 \pi_i(s) ds < \infty$ for each $i \in \mathcal{M}$, and $\varphi_t = x_0 + \sum_{j \in \mathcal{M}} \int_0^t \frac{g(\varphi_s, j)}{\kappa(j)} \pi_j(s) ds$, and $\sum_{j \in \mathcal{M}} \pi_j(s) \Phi_{ji}^{v_j \cdot (s, \cdot)}(\varphi_s) = 0, a.e. s \in [0, 1], \forall i \in \mathcal{M}$.

REFERENCES

- [1] H. BESSAIH, Y. EFENDIEV, AND R. F. MARIS, *Stochastic homogenization of a convection-diffusion equation*, SIAM J. Math. Anal., 53 (2021), pp. 2718–2745, <https://doi.org/10.1137/19M1302776>.
- [2] A. BUDHIRAJA, P. DUPUIS, AND A. GANGULY, *Large deviations for small noise diffusions in a fast Markovian environment*, Electron. J. Probab., 23 (2018), 112.
- [3] X. CHEN, Z.-Q. CHEN, T. KUMAGAI, AND J. WANG, *Homogenization of symmetric stable-like processes in stationary ergodic media*, SIAM J. Math. Anal., 53 (2021), pp. 2957–3001, <https://doi.org/10.1137/20M1326726>.
- [4] Z. CHEN AND M. I. FREIDLIN, *Smoluchowski–Kramers approximation and exit problems*, Stoch. Dyn., 5 (2005), pp. 569–585.
- [5] L. CHENG, R. LI, AND W. LIU, *Moderate deviations for the Langevin equation with strong damping*, J. Stat. Phys., 170 (2018), pp. 845–861.
- [6] S. CERRAI AND M. FREIDLIN, *Large deviations for the Langevin equation with strong damping*, J. Stat. Phys., 161 (2015), pp. 859–875.
- [7] S. CERRAI AND N. GLATT-HOLTZ, *On the convergence of stationary solutions in the Smoluchowski–Kramers approximation of infinite dimensional systems*, J. Funct. Anal., 278 (2020), 108421.
- [8] S. CERRAI AND G. XI, *A Smoluchowski–Kramers approximation for an infinite dimensional system with state-dependent damping*, Ann. Probab., 50 (2022), pp. 874–904.
- [9] F. CIPRIANO AND T. COSTA, *A large deviations principle for stochastic flows of viscous fluids*, J. Differential Equations, 264 (2018), pp. 5070–5108.
- [10] A. DEMBO AND O. ZEITOUNI, *Large Deviations Techniques and Their Applications*, 2nd ed., Jones and Bartlett, Boston, 1998.
- [11] J. D. DEUSCHEL AND D. W. STROOCK, *Large Deviations*, Academic Press, San Diego, 1989.
- [12] P. DUPUIS AND R. ELLIS, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, 2011.
- [13] J. FENG AND T. KURTZ, *Large Deviations for Stochastic Processes*, American Mathematical Society, Providence, RI, 2006.
- [14] G. FIBICH, A. GAVIOUS, AND E. SOLAN, *Averaging principle for second-order approximation of heterogeneous models with homogeneous models*, PNAS, 109 (2012), pp. 19545–19550.
- [15] M. I. FREIDLIN, *Some remarks on the Smoluchowski–Kramers approximation*, J. Stat. Phys., 117 (2004), pp. 617–634.
- [16] M. I. FREIDLIN AND A. D. WENTZELL, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 1984.
- [17] P. GAO, *Averaging principle for complex Ginzburg–Landau equation perturbed by mixing random forces*, SIAM J. Math. Anal., 53 (2021), pp. 32–61, <https://doi.org/10.1137/20M1325836>.
- [18] A. GUILLIN, *Averaging principle of SDE with small diffusion: Moderate deviations*, Ann. Probab., 31 (2003), pp. 413–443.
- [19] Q. HE AND G. YIN, *Large deviations for multi-scale Markovian switching systems with a small diffusion*, Asymptot. Anal., 87 (2014), pp. 123–145.
- [20] W. HONG, S. LI, AND W. LIU, *Freidlin–Wentzell type large deviation principle for multiscale locally monotone SPDEs*, SIAM J. Math. Anal., 53 (2021), pp. 6517–6561, <https://doi.org/10.1137/21M1404612>.
- [21] S. A. ISAACSON, J. MA, AND K. SPILIOPOULOS, *Mean field limits of particle-based stochastic reaction-diffusion models*, SIAM J. Math. Anal., 54 (2022), pp. 453–511, <https://doi.org/10.1137/20M1365600>.
- [22] H. KESTEN AND G. C. PAPANICOLAOU, *A limit theorem for turbulent diffusion*, Comm. Math. Phys., 65 (1979), pp. 97–128.
- [23] H. KESTEN AND G. C. PAPANICOLAOU, *A limit theorem for stochastic acceleration*, Comm. Math. Phys., 78 (1980), pp. 19–63.

- [24] R. Z. KHASMINSKII AND G. YIN, *On averaging principles: An asymptotic expansion approach*, SIAM J. Math. Anal., 35 (2004), pp. 1534–1560, <https://doi.org/10.1137/S0036141002403973>.
- [25] Y. KIFER, *Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging*, Mem. Amer. Math. Soc., 201 (2009), no. 944.
- [26] H. J. KUSHNER AND P. DUPUIS, *Large deviations estimates for systems with small (wideband) noise effects and applications to stochastic systems and communication theory*, in Probability Theory and Applications, De Gruyter, 2020, pp. 659–662.
- [27] R. S. LIPTSER, *Large deviations for two scaled diffusions*, Probab. Theory Related Fields, 106 (1996), pp. 71–104.
- [28] R. S. LIPTSER AND A. A. PUHALSKII, *Limit theorems on large deviations for semimartingales*, Stochastics Stochastics Rep., 38 (1992), pp. 201–249.
- [29] X. MAO, *Stochastic Differential Equations and Their Applications*, Horwood, Chichester, 1997.
- [30] K. NARITA, *Asymptotic behavior of solutions of SDE for relaxation oscillations*, SIAM J. Math. Anal., 24 (1993), pp. 172–199, <https://doi.org/10.1137/0524011>.
- [31] N. NGUYEN AND G. YIN, *A class of Langevin equations with Markov switching involving strong damping and fast switching*, J. Math. Phys., 61 (2020), 063301.
- [32] N. NGUYEN AND G. YIN, *Large deviations principles for Langevin equations in random environment and applications*, J. Math. Phys., 62 (2021), 083301.
- [33] B. PEI, Y. INAHAMA, AND Y. XU, *Averaging principle for fast-slow system driven by mixed fractional Brownian rough path*, J. Differential Equations, 301 (2021), pp. 202–235.
- [34] A. A. PUHALSKII, *On large deviations of coupled diffusions with time scale separation*, Ann. Probab., 44 (2016), pp. 3111–3186.
- [35] E. M. PURCELL, *Life at low Reynolds number*, Amer. J. Phys., 45 (1977), pp. 3–11.
- [36] H. TOUCHETTE, *The large deviation approach to statistical mechanics*, Phys. Rep., 478 (2009), pp. 1–69.
- [37] B. VAN DER POL, *Über Relaxations Schwingungen*, Jahrb. Drahtl. Telegr. Teleph., 28 (1927), pp. 178–184.
- [38] F. XI AND G. YIN, *Stochastic Liénard equations with state-dependent switching*, Acta Math. Appl. Sin. Engl. Ser., 31 (2015), pp. 893–908.
- [39] F. XI, C. ZHU, AND F. WU, *On strong Feller property, exponential ergodicity and large deviations principle for stochastic damping Hamiltonian systems with state-dependent switching*, J. Differential Equations, 286 (2021), pp. 856–891.
- [40] L. XIE AND L. YANG, *The Smoluchowski-Kramers limits of stochastic differential equations with irregular coefficients*, Stochastic Process. Appl., 150 (2022), pp. 91–115.
- [41] J. XU, *An averaging principle for slow-fast fractional stochastic parabolic equations on unbounded domains*, Stochastic Process. Appl., 150 (2022), pp. 358–396.
- [42] J. XU, Q. LIAN, AND J. WU, *A strong averaging principle rate for two-time-scale coupled forward-backward stochastic differential equations driven by fractional Brownian motion*, Appl. Math. Optim., 88 (2023), 32.
- [43] A. YU VERETENNIKOV, *On large deviations in the averaging principle for SDEs with a full dependence*, Ann. Probab., 27 (1999), pp. 284–296.
- [44] A. YU VERETENNIKOV, *On large deviations for SDEs with small diffusion and averaging*, Stochastic Process. Appl., 89 (2000), pp. 69–79.