Zuse Institute Berlin

Takustrasse 7
D-14195 Berlin-Dahlem
Germany

WEI ZHANG[1], CARSTEN HARTMANN[1,3], AND CHRISTOF SCHÜTTE[1,2]

[1]*Institute for Mathematics, Freie Universität Berlin, Germany*

[2]*Zuse Institute Berlin, Germany*

[3]*Institute of Mathematics, Brandenburgische Technische Universität Cottbus-Senftenberg, Germany*

# Effective Dynamics Along Given Reaction Coordinates, and Reaction Rate Theory

ZIB Report 16-35 (June 2016)

Zuse Institute Berlin
Takustrasse 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

# EFFECTIVE DYNAMICS ALONG GIVEN REACTION COORDINATES, AND REACTION RATE THEORY

WEI ZHANG [1] , CARSTEN HARTMANN [1 3] , AND CHRISTOF SCHÜTTE [1 2]

**Abstract.** In molecular dynamics and related fields one considers dynamical descriptions of complex systems in full (atomic) detail. In order to reduce the overwhelming complexity of realistic systems (high dimension, large timescale spread, limited computational resources) the projection of the full dynamics onto some reaction coordinates is examined in order to extract statistical information like free energies or reaction rates. In this context, the effective dynamics that is induced by the full dynamics on the reaction coordinate space has attracted considerable attention in the literature. In this article, we contribute to this discussion: We first show that if we start with an ergodic diffusion processes whose invariant measure is unique then these properties are inherited by the effective dynamics. Then, we give equations for the effective dynamics, discuss whether the dominant timescales and reaction rates inferred from the effective dynamics are accurate approximations of such quantities for the full dynamics, and compare our findings to results from approaches like Zwanzig-Mori, averaging, or homogenization. Finally, by discussing the algorithmic realization of the effective dynamics, we demonstrate that recent algorithmic techniques like the "equation-free" approach and the "heterogeneous multiscale method" can be seen as special cases of our approach.

**Key words.** Ergodic diffusion, reaction coordinate, effective dynamics, model reduction, equation-free, heterogeneous multiscale method.

**AMS subject classifications.** 58J65, 62P35

**1. Introduction.** Diffusion processes are continuous-time Markov processes on a continuous state space, typically $\mathbb{R}^n$, and can be described by stochastic differential equations (SDE). A diffusion process is said to be ergodic if it has a unique invariant (stationary) measure and the ergodic theorem is satisfied, i.e. its stationary statistics can be realized by an infinitely long trajectory [65, 14, 58].

In the past few decades, ergodic diffusion processes have been extensively investigated and widely applied to complex systems arising from physics, biology and chemistry [26, 50, 87, 67]. In realistic applications, e.g. in molecular dynamics or materials science, the systems of interest are high-dimensional and exhibit essential phenomena on vastly different time scales [67, 19, 64, 32]. Mostly, numerical simulation of the process of interest is performed in order to extract particular quantities of interest, e.g., thermodynamic quantities like free energy, or dynamics-related quantities, such as reaction or transition rates. In the latter case, computing good approximations of dynamics-related quantities is often infeasible even on the fastest computers, since the underlying processes (reactions, transitions) are rare events and the inherent timescale are too long to allow for the extraction of reliable statistics from long simulation runs. In order to circumvent this obstacle one selects some *reaction coordinates* or *collective variables* of the system and then tries to compute the desired quantities by projection of the high-dimensional full dynamics onto the low-dimensional reaction coordinate space [32, 35, 51, 57, 64], leading to what we call an *effective dynamics*. There are various *model reduction* approaches that yield an

[1]Institute of Mathematics, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany

[2]Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany

[3]Institute of Mathematics, Brandenburgische Technische Universität Cottbus-Senftenberg, Platz der Deutschen Einheit 1, 03046 Cottbus, Germany

[4]Email : chartman@mi.fu-berlin.de, schuette@zib.de, wei.zhang@fu-berlin.de

effective dynamics starting from the full dynamics, and this article is about putting several of these approaches into a common perspective.

Our key questions are the following:

Q1. Do the most prominent model reduction approaches lead to the same type of effective dynamics?

Q2. Are structural properties the full (high-dimensional) dynamics like ergodicity / existence of an invariant measure or reversibility inherited by the (low-dimensional) effective dynamics?

Q3. When dynamical properties like reaction/transition rates or inherent timescales are computed based on the effective dynamics, how well do they approximate the associated quantities of the full dynamics?

Q4. How are the most prominent approaches (e.g., the equation-free approach [42, 44, 43], the heterogeneous multiscale method (HMM) [20, 21, 1, 77], or other approaches [19, 52]) for performing numerical simulation of the effective dynamics related? Is there a unifying framework for studying them in one setting?

Each of the above questions has been answered to some extent by researchers from different disciplines.

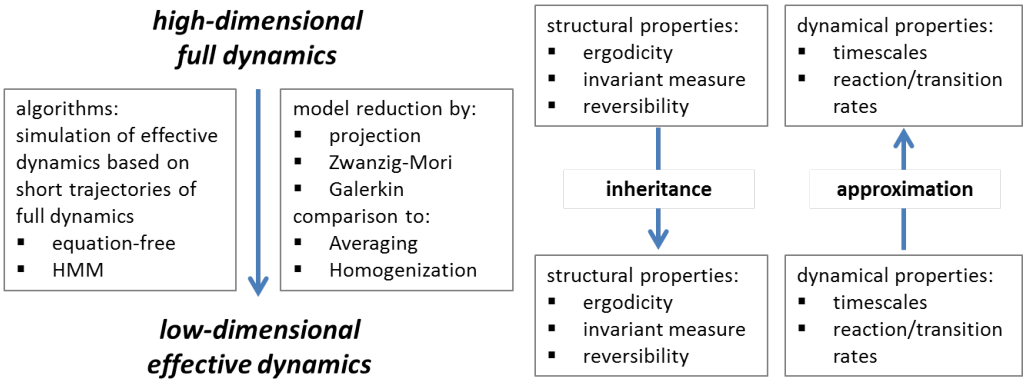


Fig. 1.1: Outline of the article.

Here we discuss them based on a joint background, providing many links to the related literature. However, developing this background requires an extended mathematical discussion of structural and dynamical properties of ergodic diffusion processes. Based on this discussion we will see that in most cases all well-known model reduction techniques yield the same kind of effective dynamics (Q1), that many structural properties are inherited by the effective dynamics (Q2) and that the most prominent simulation techniques for the effective dynamics blend well into the joint background developed herein. The answers provided to questions (Q3) contain completely new insights into the approximation quality of timescales and reaction rates computed via the effective dynamics. Figure 1.1 summarizes these different aspects in a single diagram.

This paper is organized as follows. In Section 2, we discuss various properties of general ergodic diffusion processes. In Section 3, we focus on the model reduction of ergodic diffusion processes and study the preservation of various properties by the effective dynamics. Numerical

methods for simulating the effective dynamics are discussed and compared with each other in Section 3.4. Finally, we conclude our work with further discussions in Section 4.

**2. Ergodic diffusion processes.** In this section, we will introduce some notation related to and study some common properties of diffusion processes that are ergodic with a unique invariant measure.

Consider the diffusion process that satisfies the stochastic differential equation (SDE)

$$\begin{aligned} dx(s) &= b(x(s))ds + \sqrt{2\beta^{-1}}\sigma(x(s))dw(s)\,, \quad s \geq 0\,, \\ x(0) &= x\,, \end{aligned} \tag{2.1}$$

where $x(s) \in \mathbb{R}^n$, parameter $\beta > 0$ is the inverse of system's temperature, and $w(s)$ is a $k$-dimensional Brownian motion. We assume that both, the drift vector $b : \mathbb{R}^n \to \mathbb{R}^n$, and the diffusion matrix $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ are smooth functions. The infinitesimal generator of the dynamics (2.1) is given by

$$\mathcal{L} = \sum_{i=1}^{n} b_i \frac{\partial}{\partial x_i} + \frac{1}{\beta}\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\,, \tag{2.2}$$

where the matrix $a(x) = (\sigma\sigma^T)(x) \in \mathbb{R}^{n\times n}$, $\forall x \in \mathbb{R}^n$. Subsequently we always assume

$$\eta^T a(x)\eta = \sum_{i,j=1}^{n} a_{ij}(x)\eta_i\eta_j \geq c_1|\eta|^2\,, \quad \forall\, \eta\,, x \in \mathbb{R}^n\,, \tag{2.3}$$

for some constant $c_1 > 0$. Here, $b_i, a_{ij}$ denote the components of vector $b$ and matrix $a$, respectively. Furthermore, let $u(s,\cdot)$ be the probability density function (with respect to Lebesgue measure) of system (2.1) at time $s \geq 0$, i.e.

$$\mathbb{P}(x(s) \in A \mid x(0) \sim u_0) = \int_A u(s,x)\,dx, \quad \forall\, A \subset \mathbb{R}^n\,,\, A \text{ is measurable}\,,$$

where $x(0) \sim u_0$ means that the initial probability distribution at time $s = 0$ has a density $u_0$ with respect to the Lebesgue measure $dx$. Then $u$ satisfies the Kolmogorov forward equation (also called Fokker-Planck equation)

$$\frac{\partial u}{\partial s} = -\sum_{i=1}^{n}\frac{\partial(b_i u)}{\partial x_i} + \frac{1}{\beta}\sum_{i,j=1}^{n}\frac{\partial^2(a_{ij}u)}{\partial x_i \partial x_j}\,, \tag{2.4}$$

with initial probability density $u(0,\cdot) = u_0$. We also assume that the process given by (2.1) is ergodic and has a unique invariant measure $\mu$ whose probability density is $\rho > 0$, satisfying

$$d\mu(dx) = \rho(x)\,dx\,, \quad \text{with} \quad \int_{\mathbb{R}^n}\rho(x)dx = 1\,. \tag{2.5}$$

The invariant density $\rho : \mathbb{R}^n \to \mathbb{R}^+$ (set of positive real numbers) satisfies the PDE

$$-\sum_{i=1}^{n}\frac{\partial(b_i\rho)}{\partial x_i} + \frac{1}{\beta}\sum_{i,j=1}^{n}\frac{\partial^2(a_{ij}\rho)}{\partial x_i\partial x_j} = 0\,, \quad \forall\, x \in \mathbb{R}^n\,, \tag{2.6}$$

i.e. the right hand side of the Fokker-Planck equation (2.4) vanishes when we take $u \equiv \rho$.

Given a probability density function $u : \mathbb{R}^n \to \mathbb{R}_0^+$ with $\int_{\mathbb{R}^n} u\,dx = 1$, we also introduce the relative entropy with respect to $\rho$, which is defined as

$$\mathrm{Ent}(u) = \int_{\mathbb{R}^n} u\ln(\frac{u}{\rho})\,dx\,. \tag{2.7}$$

Applying Jensen's inequality, we can verify that $\mathrm{Ent}(u) \geq 0$.

**2.1. Probability velocity, generator decomposition.** In this part we introduce the *probability velocity* of diffusion process given by (2.1), that is useful when investigating system's various properties [68, 86, 30, 78, 25]; some related work can also be found in [68] based on the forward operator.

Specifically, given a probability density function $u > 0$ satisfying $\int_{\mathbb{R}^n} u(x)\,dx = 1$, we define the probability velocity vector $J(u)$ whose $i$th component is given by

$$J_i(u) = b_i - \frac{1}{\beta u}\sum_{j=1}^{n}\frac{\partial(a_{ij}u)}{\partial x_j}\,, \quad 1 \le i \le n\,. \tag{2.8}$$

The use of the terminology *probability velocity* is motivated by the fact that equation (2.4) can be written as a continuity equation

$$\frac{\partial u}{\partial s} = -\mathrm{div}(J(u)u)\,, \tag{2.9}$$

and the quantity $J(u)u$ is usually called *probability current* or *probability flux* in the literatures [86, 30, 25] (notice that the meaning of the notation $J$ used in these citations is different from ours). Especially for $u \equiv \rho$, the equilibrium probability density, we define $J^{st} = J(\rho)$ and call it *stationary probability velocity*. Using (2.8), we can directly verify that

$$J_i(u) = J_i^{st} - \frac{1}{\beta}\sum_{j=1}^{n} a_{ij}\frac{\partial}{\partial x_j}\big(\ln\frac{u}{\rho}\big)\,, \tag{2.10}$$

and PDE (2.6) is equivalent to the equation

$$-\mathrm{div}(J^{st}\rho) = 0\,. \tag{2.11}$$

We will also frequently use the Hilbert space $H = L^2(\mathbb{R}^n, \mu)$ with inner product

$$\langle f, g\rangle_\mu = \int_{\mathbb{R}^n} f\,g\rho\,dx\,, \quad f, g \in H\,, \tag{2.12}$$

and the norm $|f|_\mu = \sqrt{\langle f, f\rangle_\mu}$, $\forall f \in H$. Using integration by parts formula, it is direct to check that $\forall f, g \in D(L) \subset H$, we have

$$\begin{aligned}
&\langle \mathcal{L}f, g\rangle_\mu = \langle f, \mathcal{L}g\rangle_\mu - 2\langle f, J^{st}\cdot\nabla g\rangle_\mu = \langle J^{st}\cdot\nabla f, g\rangle_\mu - \frac{1}{\beta}\langle a\nabla f, \nabla g\rangle_\mu\,,\\
&\langle J^{st}\cdot\nabla f, g\rangle_\mu + \langle J^{st}\cdot\nabla g, f\rangle_\mu = 0\,.
\end{aligned} \tag{2.13}$$

From the first equation in (2.13), we deduce that $\mathcal{L}^* = \mathcal{L} - 2J^{st}\cdot\nabla$, where $\mathcal{L}^*$ denotes the dual operator of $\mathcal{L}$ in $H$. Especially, taking $f = g \in D(\mathcal{L})$ (the domain of $\mathcal{L}$) in (2.13), we obtain

$$\begin{aligned}
&\langle J^{st}\cdot\nabla f, f\rangle_\mu = 0\,,\\
&\langle \mathcal{L}f, f\rangle_\mu = -\frac{1}{\beta}\langle a\nabla f, \nabla f\rangle_\mu\,.
\end{aligned} \tag{2.14}$$

Furthermore, it is direct to verify that the following conditions are equivalent to each other.

$$\begin{aligned}
&J^{st} \equiv 0\,,\\
\Longleftrightarrow\quad &b_i - \frac{1}{\beta\rho}\sum_{j=1}^{n}\frac{\partial(a_{ij}\rho)}{\partial x_j} \equiv 0\,, \quad 1 \le i \le n\,,\\
\Longleftrightarrow\quad &\mathcal{L} = \mathcal{L}^*\,, \quad \langle \mathcal{L}f, g\rangle_\mu = \langle f, \mathcal{L}g\rangle_\mu = -\frac{1}{\beta}\langle a\nabla f, \nabla g\rangle_\mu\,, \quad \forall f, g \in D(L)\,.
\end{aligned} \tag{2.15}$$

This also allows to check the reversibility of the diffusion process: Although reversibility may have different meanings in different communities, roughly speaking, a diffusion process is said to be reversible if its statistical properties are unchanged when we reverse time's arrow [70, 61, 6]. Any of the conditions in (2.15) is equivalent to the process being reversible (see below).

In the general case (whether reversible or not), using equation (2.8) for the stationary probability velocity $J^{st}$, we can recast (2.1) as

$$dx_i(s) = J_i^{st} ds + \frac{1}{\beta\rho} \sum_{j=1}^{n} \frac{\partial(a_{ij}\rho)}{\partial x_j} ds + \sqrt{2\beta^{-1}} \sum_{j=1}^{d} \sigma_{ij}\, dw_j(s)\,, \quad 1 \le i \le n\,, \tag{2.16}$$

and the generator as $\mathcal{L} = \mathcal{L}_a + \mathcal{L}_s$ where

$$\mathcal{L}_a = \frac{1}{2}(\mathcal{L} - \mathcal{L}^*) = J^{st} \cdot \nabla\,, \qquad \mathcal{L}_s = \frac{1}{2}(\mathcal{L} + \mathcal{L}^*) = \frac{1}{\beta\rho} \nabla(\rho a \nabla \cdot)\,. \tag{2.17}$$

The above relations can be obtained directly from (2.13) as well; the decomposition formula for the forward operator was obtained in [68].

For the dual operator $\mathcal{L}^*$, we have $\mathcal{L}^* = \mathcal{L} - 2J^{st} \cdot \nabla = -\mathcal{L}_a + \mathcal{L}_s$, and therefore the corresponding (time reversed) dynamics is

$$d\overleftarrow{x}_i(s) = -J_i^{st} ds + \frac{1}{\beta\rho} \sum_{j=1}^{n} \frac{\partial(a_{ij}\rho)}{\partial x_j} ds + \sqrt{2\beta^{-1}} \sum_{j=1}^{d} \sigma_{ij}\, dw_j(s)\,, \quad 1 \le i \le n\,, \tag{2.18}$$

which is obtained by changing the direction of the probability velocity $J^{st}$ in the original dynamics (2.16). It is straightforward to verify that the time reversed dynamics (2.18) has the same invariant measure $\mu$ as the original dynamics and both dynamics coincide with each other if and only if they are reversible, i.e. when any of the conditions in (2.15) are satisfied.

Finally we turn our attention to the stationary flow governed by

$$\dot{\phi}_s = J^{st}(\phi_s)\,, \qquad \phi_s : \mathbb{R}^n \to \mathbb{R}^n\,, \tag{2.19}$$

with $\phi_0 = \mathrm{id}$. Assuming that (2.19) has a unique solution, it holds that $\phi_{t+s} = \phi_t \circ \phi_s$, $\forall s, t \ge 0$. Let $(\phi_s)_{\#}\nu$ denote the transportation of some probability measure $\nu$ under the map $\phi_s$, being defined by $\big((\phi_s)_{\#}\nu\big)(A) = \nu(\phi_s^{-1}(A))$ for any measurable set $A \subset \mathbb{R}^n$ (see Chapter 5 of [3] and Chapter 1 of [83] for more details). We state the following result, the proof of which is given in Appendix A:

PROPOSITION 1. *Let $J^{st}$ be the stationary probability velocity of (2.1), and let $\phi_s$ be the stationary flow map defined by (2.19). Then*

1. *$(\phi_s)_{\#}\mu = \mu$, $\forall s \ge 0$, where $\mu$ is the unique invariant measure of dynamics (2.1).*
2. *Given a probability measure $\mu_v$ whose probability density function $v$ satisfies $\int v(x)\,dx = 1$. Let $d\big((\phi_s)_{\#}\mu_v\big) = v_s\,dx$. Then $\mathrm{Ent}(v_s) = \mathrm{Ent}(v)$ $\forall s \ge 0$.*

These results show that both the invariant measure $\mu$ and the relative entropy (of any probability density function) are invariant under the action of $\phi_t$. Results similar to Proposition 1 were also obtained in [68] using a different argument.

**2.2. Convergence to equilibrium.** In the following we discuss the process' speed of converge to equilibrium. Let $T_t$ be the semigroup defined by the infinitesimal generator $\mathcal{L}$ on the Hilbert space $H$. It is known that for function $f \in H$, we have $(T_t f)(x) = \mathbf{E}(f(x(t)) \mid x(0) = x)$,

where $x(t)$ satisfies SDE (2.1). Given $f \in H$, it is often interesting to study the decay of the quantity $\mathrm{Var}_\mu(T_t f) = |T_t f - \mathbf{E}_\mu f|_\mu^2$, where $\mathbf{E}_\mu$, and $\mathrm{Var}_\mu$ denote the mathematical expectation and the variance under probability measure $\mu$, respectively (notice that $\mathbf{E}_\mu T_t f = \mathbf{E}_\mu f$). Without loss of generality, we can assume $\mathbf{E}_\mu f = 0$ and compute

$$\frac{\partial}{\partial t}|T_t f|_\mu^2 = 2\langle \mathcal{L} T_t f, T_t f\rangle_\mu = 2\langle \mathcal{L}_s T_t f, T_t f\rangle_\mu = -\frac{2}{\beta}\langle a\nabla T_t f, \nabla T_t f\rangle_\mu \le 0\,, \tag{2.20}$$

where we have used relations (2.14), (2.17) and the assumption (2.3) for matrix $a$. The above inequality implies that $\mathrm{Var}_\mu(T_t f)$ is nonincreasing as $t \to +\infty$. Furthermore if the Poincaré inequality

$$|f|_\mu^2 \le \frac{1}{\lambda}\langle a\nabla f, \nabla f\rangle_\mu\,, \tag{2.21}$$

holds for some fixed $\lambda > 0$, and for all $f \in H$ such that $\mathbf{E}_\mu f = 0$ and $f$ is $\mathcal{C}^1$ differentiable, then we can obtain

$$\frac{\partial}{\partial t}|T_t f|_\mu^2 \le -\frac{2\lambda}{\beta}|T_t f|_\mu^2 \quad \Longrightarrow \quad |T_t f|_\mu^2 \le e^{-\frac{2\lambda}{\beta}t}|f|_\mu^2\,, \quad \forall t > 0\,.$$

The last inequality can be shown to hold for all $f \in H$ by considering an approximation sequence $f^{(i)} \to f$ in $H$, where $f^{(i)} \in H$ and is $\mathcal{C}^1$ differentiable. Therefore we conclude that $\mathrm{Var}_\mu(T_t f) \searrow 0$, for all $f \in H$ as $t \to +\infty$ and the decay is exponentially fast.

The trend towards equilibrium can also be characterized by considering the entropy decay for the solution of the Fokker-Planck equation (2.9). We will only record the following simple result; for a proof see Appendix A and Remark 1 for discussions.

Proposition 2. *Suppose the condition (2.3) is satisfied by the dynamics (2.1) which is ergodic with a unique invariant measure $\mu$. Let $u_t$ be the solution of Fokker-Planck equation (2.9) or, equivalently, (2.4). Then* $\mathrm{Ent}(u_t)$ *is non-increasing as $t \to +\infty$.*

Remark 1. *Much deeper and stronger results (mostly for reversible system) related to the decay of relative entropy can be found in [82, 83, 3] and references therein. Here it might be interesting to point out that Proposition 2 holds for general ergodic processes (2.1) which can be either reversible or irreversible. Furthermore, the decay rate of the relative entropy only depends on the reversible part of the infinitesimal generator.*

Now we turn to the characterization of the time scales in dynamical system (2.1). In the study of molecular dynamics, climate, chemical reactions et al., knowing the systems' different time scales as well as the dynamical behavior of the process at given time scales is an important issue in order to understand the systems. See [67, 84, 73, 57, 45, 64] and references therein.

Here we assume the dynamics is reversible and therefore its infinitesimal generator $\mathcal{L}$ is a symmetric (essentially self-adjoint) operator. We also assume that the essential spectrum of $\mathcal{L}$ is empty [85], therefore its spectrum consists of a set of isolated eigenvalues [16, Theorem 4.1.5]. It is known that the time scales of the dynamics are determined by the eigenvalue problem (notice the minus sign)

$$-\mathcal{L} f = \lambda f\,, \tag{2.22}$$

which holds for some eigenfunction $f \in D(\mathcal{L}) \subset H$ and eigenvalue $\lambda$. Since $\mathcal{L} = \mathcal{L}^*$ is an essentially self-adjoint operator, we know $\lambda \in \mathbb{R}$. It is clear that $\lambda = 0$, $f \equiv 1$ is a solution of

(2.22). Otherwise, we can assume $|f|_\mu = 1$ and $\nabla f \not\equiv 0$. Multiplying both sides of (2.22) by $f$ and integrating with respect to $\mu$ gives

$$\lambda = -\langle \mathcal{L}f, f\rangle_\mu = \frac{1}{\beta}\langle a\nabla f, \nabla f\rangle_\mu \geq \frac{c_1}{\beta}|\nabla f|_\mu^2 > 0\,, \tag{2.23}$$

where we have used the assumption (2.3) for matrix $a$. Therefore, we can assume that there are eigenvalues $\lambda_i \in \mathbb{R}$, $i \geq 0$, with $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_k \leq \cdots$, and eigenfunctions $\varphi_i \in D(\mathcal{L}) \subset H$, s.t. $-\mathcal{L}\varphi_i = \lambda_i\varphi_i$.

For two eigenfunctions $\varphi_i$, $\varphi_j$ corresponding to $\lambda_i \neq \lambda_j$, we have

$$\lambda_i\langle \varphi_i, \varphi_j\rangle_\mu = -\langle \mathcal{L}\varphi_i, \varphi_j\rangle_\mu = -\langle \varphi_i, \mathcal{L}\varphi_j\rangle_\mu = \lambda_j\langle \varphi_i, \varphi_j\rangle_\mu\,,$$

which implies $\langle \varphi_i, \varphi_j\rangle_\mu = 0$. When $\lambda_i = \lambda_j$, we can also choose eigenfunctions $\varphi_i, \varphi_j$ such that $\langle \varphi_i, \varphi_j\rangle_\mu = 0$. Thus, we can assume without loss of generality that eigenfunctions $\varphi_i$ form an orthonormal basis of $H$. And for all $f \in D(\mathcal{L})$ with $\mathbf{E}_\mu f = 0$, using the orthogonal expansion $f = \sum_{i=1}^{+\infty}\langle f, \varphi_i\rangle_\mu\,\varphi_i$, we can verify that Poincaré inequality (2.21) holds for constant $\lambda = \lambda_1$.

For an eigenfunction $\varphi_i$ with $i > 0$, we have $\mathbf{E}_\mu\varphi_i = \langle \varphi_i, \varphi_0\rangle_\mu = 0$ and

$$\frac{\partial}{\partial t}|T_t\varphi_i|_\mu^2 = 2\langle \mathcal{L}T_t\varphi_i, T_t\varphi_i\rangle_\mu = -2\lambda_i|T_t\varphi_i|_\mu^2\,,$$

which implies $\mathrm{Var}_\mu(T_t\varphi_i) = |T_t\varphi_i|_\mu^2 = e^{-2\lambda_i t}|\varphi_i|_\mu^2 = e^{-2\lambda_i t}$, where $\lambda_i > 0$. Therefore we can conclude that $\mathrm{Var}_\mu(T_t\varphi_i)$ for an eigenfunction $\varphi_i$, $i > 0$ decays on time scales of order $(2\lambda_i)^{-1}$.

**2.3. Connection with diffusions on Riemannian manifolds.** Consider the process (2.1) with the unique invariant measure $\mu$ in (2.5) whose probability density is $\rho$. By defining the potential $V = -\beta^{-1}\ln\rho$, we get from equation (2.8) that

$$b_i = J_i^{st} + \frac{e^{\beta V}}{\beta}\sum_{j=1}^{n}\frac{\partial(a_{ij}e^{-\beta V})}{\partial x_j}\,, \quad 1 \leq i \leq n\,, \tag{2.24}$$

and thus (2.1) can be rewritten as

$$dx(s) = \big[J^{st}(x(s)) - a(x(s))\nabla V(x(s)) + \frac{1}{\beta}\nabla\cdot a(x(s))\big]ds + \sqrt{2\beta^{-1}}\sigma(x(s))dw(s)\,, \tag{2.25}$$

with $x(0) = x$. In the last equation, $\nabla\cdot a$ is a vector in $\mathbb{R}^n$ with components $(\nabla\cdot a)_i = \sum_{j=1}^{n}\frac{\partial a_{ij}}{\partial x_j}$, $1 \leq i \leq n$, also see (2.16). The decomposition of the infinitesimal generator in (2.17) now reads

$$\mathcal{L} = J^{st}\cdot\nabla + \frac{e^{\beta V}}{\beta}\nabla\Big(e^{-\beta V}a\nabla\cdot\Big)\,. \tag{2.26}$$

Notice that (2.25) reduces to the standard overdamped dynamics when $J^{st} = 0$ and $a = \sigma = \mathrm{id} \in \mathbb{R}^{n\times n}$.

We may also think of the space $\mathbb{R}^n$ as a Riemannian manifold $M$ and consider the dynamics (2.25) as a diffusion process on $M$. In this case, we can simply take the identity map $\mathrm{id} : M = \mathbb{R}^n \to \mathbb{R}^n$ to obtain a local coordinate. Specifically, from (2.3) we know that the matrix $a = \sigma\sigma^T$ is positive definite, therefore we can define the metric tensor $g(x) = a^{-1}(x)$, $x \in \mathbb{R}^n$ with the induced inner product $\langle v_1, v_2\rangle = v_1^T g v_2$ for $\forall v_1, v_2 \in \mathbb{R}^n$. We have $(g^{-1}(x))_{ij} = g^{ij}(x) = a_{ij}(x)$,

$1 \le i, j \le n$. Also define $G(x) = \det g(x)$ and consider the nature volume measure $\mu_g(dx) = \sqrt{G(x)}\, dx$ on $M$. Under this (global) coordinate chart, we have [39, 66]

$$\begin{aligned} \nabla_g f &= \sum_{i,j=1}^{n} g^{ij} \frac{\partial f}{\partial x_i} \frac{\partial}{\partial x_j}\,, \\ \Delta_g f &= \frac{1}{\sqrt{G}} \sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\Big(\sqrt{G} g^{ij} \frac{\partial f}{\partial x_j}\Big)\,, \end{aligned} \tag{2.27}$$

where $f \in \mathcal{C}^\infty(M)$, $\nabla_g$, $\mathrm{div}_g$ are the gradient, divergence operator respectively, and $\Delta_g = \mathrm{div}_g \nabla_g$ is the Laplace-Beltrami operator on the manifold $M$. Using (2.26), we can write the infinitesimal generator $\mathcal{L}$ as a 2nd order differential operator on $M$ as

$$\begin{aligned} \mathcal{L} f &= J^{st} \cdot \nabla f + \frac{e^{\beta V}}{\beta} \sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\Big(\frac{e^{-\beta V}}{\sqrt{G}} \sqrt{G} g^{ij} \frac{\partial f}{\partial x_j}\Big) \\ &= J^{st} \cdot (g g^{-1} \nabla f) + \frac{1}{\beta} \sum_{i,j=1}^{n} \Big[\frac{\partial}{\partial x_i}\Big(\ln \frac{e^{-\beta V}}{\sqrt{G}}\Big) g^{ij} \frac{\partial f}{\partial x_j} + \frac{1}{\sqrt{G}} \frac{\partial}{\partial x_i}\Big(\sqrt{G} g^{ij} \frac{\partial f}{\partial x_j}\Big)\Big] \\ &= \langle J^{st}, \nabla_g f\rangle - \langle \nabla_g\big(V + \frac{1}{2\beta} \ln G\big), \nabla_g f\rangle + \frac{1}{\beta} \Delta_g f \\ &= \Big[J^{st} - \nabla_g\big(V + \frac{1}{2\beta} \ln G\big) + \frac{1}{\beta} \Delta_g\Big] f\,. \end{aligned}$$

In the last equation above, $J^{st}$ is considered as a vector field on $M$. Accordingly, (2.25) can be written as an SDE on the Riemannian manifold $M$

$$dx(s) = \big[J^{st} - \nabla_g (V + \frac{1}{2\beta} \ln G)\big] ds + \sqrt{2\beta^{-1}} dW(s)\,, \tag{2.28}$$

where $W(s)$ is the Brownian motion on manifold $M$ [39]. Further notice that the stationary probability velocity satisfies

$$-\mathrm{div}_g\big(J^{st} e^{-\beta(V + \frac{1}{2\beta} \ln G)}\big) = 0\,, \tag{2.29}$$

which should be compared to (2.11). And dynamics (2.28) has a unique invariant measure $\mu$ which is related to the measure $\mu_g$ by $d\mu = e^{-\beta(V + \frac{1}{2\beta} \ln G)}\, d\mu_g$.

Especially in the reversible case, our derivations above show that if dynamics (2.1) is both reversible and ergodic, then we can always rewrite it as a gradient system (2.28) evolving on a manifold. This point of view may be conceptually useful. For instance, consider the study of rare events in the zero temperature limit. We can conclude that the most probable transition path of system (2.1) connecting two metastable states is actually the minimal energy path (on the manifold) and the string method (on the manifold) can be applied to compute the path [23, 22, 81, 54]. This point of view will be further exploited in Section 3 when discussing model reduction.

**3. Model reduction of ergodic diffusions.** In this section we consider the model reduction of dynamics (2.1) to a low dimensional effective dynamics [51]. Various properties of the effective dynamics and the relations with the original dynamics will be studied. To this end, we assume that there is a $\mathcal{C}^2$ function $\xi : \mathbb{R}^n \to \mathbb{R}^m$, $1 \le m < n$, which describes the reaction coordinate of the system.

**3.1. Projection operator.** We start by defining a projection operator on the Hilbert space $H$. Given the reaction coordinate function $\xi : \mathbb{R}^n \to \mathbb{R}^m$, we define $\Sigma_z = \{x \in \mathbb{R}^n \mid \xi(x) = z\}$ and assume it is nonempty for all $z \in \mathbb{R}^m$. For any function $f \in H$ and $x \in \mathbb{R}^n$, let

$$\begin{aligned} \mathrm{P}f(x) =& \int_{\Sigma_z} f(x')\, d\mu_z(x') \\ =& \frac{1}{Q(z)} \int_{\Sigma_z} f(x')\rho(x') \Big[\det\Big(\nabla\xi^T(x')\nabla\xi(x')\Big)\Big]^{-\frac{1}{2}} d\sigma_z(x')\,, \end{aligned} \tag{3.1}$$

where $z = \xi(x) \in \mathbb{R}^m$, and $\sigma_z$ denotes the surface measure on $\Sigma_z$. $\mu_z$ is the probability measure on $\Sigma_z$ whose probability density with respect to $\sigma_z$ is $\frac{\rho}{Q(z)}\big[\det(\nabla\xi^T\nabla\xi)\big]^{-\frac{1}{2}}$. The Jacobian matrix $\nabla\xi(x)$ of the map $\xi : \mathbb{R}^n \to \mathbb{R}^m$ is an $n \times m$ matrix. $Q : \mathbb{R}^m \to \mathbb{R}$ is the normalization constant defined as

$$Q(z) = \int_{\Sigma_z} \rho\big[\det(\nabla\xi^T\nabla\xi)\big]^{-\frac{1}{2}} d\sigma_z\,. \tag{3.2}$$

We assume

$$\begin{aligned} &0 < Q(z) < +\infty, \quad \forall z \in \mathbb{R}^m\,, \\ &\mathrm{P}(|\nabla\xi\eta|^2) = \mathrm{P}\big(\eta^T\nabla\xi^T\nabla\xi\eta\big) \geq c_2|\eta|^2\,, \quad \forall \eta \in \mathbb{R}^m,\, z \in \mathbb{R}^m\,, \end{aligned} \tag{3.3}$$

holds for some constant $c_2 > 0$. Notice that we have $\mathrm{P}^2 = \mathrm{P}$ and the function $\mathrm{P}f : \mathbb{R}^n \to \mathbb{R}$ actually only depends on $z = \xi(x) \in \mathbb{R}^m$.

The definition of the projection P in (3.1) is motivated by the fact that, using the co-area formula [48, 28], for any suitable function $\widetilde{h} : \mathbb{R}^m \to \mathbb{R}$ we have

$$\int_{\mathbb{R}^n} f(x)\,\widetilde{h}(\xi(x))\,\rho(x)\,dx = \int_{\mathbb{R}^m} \mathrm{P}f(\xi^{-1}(z))\,\widetilde{h}(z)\,Q(z)\,dz\,. \tag{3.4}$$

The term $\mathrm{P}f(\xi^{-1}(z))$ on the right hand side is well defined since $\mathrm{P}f$ is constant on $\xi^{-1}(z)$. Formally taking $\widetilde{h} = \delta(\cdot - z)$, we see that definition (3.1) is equivalent to the perhaps more familiar expression

$$\begin{aligned} \mathrm{P}f(x) =& \frac{1}{Q(z)} \int_{\mathbb{R}^n} \rho(x')f(x')\delta\big(\xi(x') - z\big)\,dx' \\ =& \mathbf{E}_\mu\big(f(x') \mid \xi(x') = z\big)\,, \end{aligned} \tag{3.5}$$

where $z = \xi(x)$. Also see [13, 51] for related discussions.

Taking $f = \widetilde{h} = 1$ in (3.4), we can obtain that $\int_{\mathbb{R}^m} Q(z)dz = 1$. Therefore we define the probability measure $\nu$ on $\mathbb{R}^m$ with $\nu(dz) = Q(z)\,dz$ for $z \in \mathbb{R}^m$ and consider the Hilbert space $\widetilde{H} = L^2(\mathbb{R}^m, \nu)$ as the space of functions on the reaction coordinate space $\mathbb{R}^m$. $\widetilde{H}$ induces another function space on the full state space, namely

$$H_0 = \big\{f \,|\, f \in H, f = \widetilde{f} \circ \xi, \text{ for some } \widetilde{f} \in \widetilde{H}\big\} \subset H\,. \tag{3.6}$$

It can be verified that $H_0$ forms a linear subspace of $H$ and $H_0 = \mathrm{Im}(\mathrm{P})$, i.e. the image of P defined in (3.1). The relation between $\widetilde{H}$ and $H_0$ can be formalized by introducing the lift map $\iota : \widetilde{H} \to H_0$ by setting $\iota(\widetilde{f}) = \widetilde{f} \circ \xi \in H_0$ for $\widetilde{f} \in \widetilde{H}$, i.e., $\iota(\widetilde{f})(x) = \widetilde{f}(\xi(x))$. Then using equality

(3.4) and $\mathrm{P}(\widetilde{f}\circ\xi)=\widetilde{f}\circ\xi$, we have

$$\begin{aligned}\langle \iota(\widetilde{f}),\iota(\widetilde{h})\rangle_\mu &= \int_{\mathbb{R}^n}\widetilde{f}\big(\xi(x)\big)\widetilde{h}\big(\xi(x)\big)\,\rho(x)\,dx\\ &= \int_{\mathbb{R}^m}\widetilde{f}(z)\,\widetilde{h}(z)\,Q(z)\,dz=\langle\widetilde{f},\widetilde{h}\,\rangle_\nu\,,\end{aligned}\tag{3.7}$$

i.e. $\iota$ is an isomorphism between spaces $\widetilde{H}$ and $H_0$. To simplify notations, we will use the same notation P to denote both, the projection in (3.1) as well as the mapping $\iota^{-1}\circ\mathrm{P}: H\to\widetilde{H}$.

Given some twice differentiable function $\widetilde{f}\in\widetilde{H}$, we define the operator $\widetilde{\mathcal{L}}$ by

$$\widetilde{\mathcal{L}}\widetilde{f}=\sum_{l=1}^m \mathrm{P}(\mathcal{L}\xi_l)\frac{\partial\widetilde{f}}{\partial z_l}+\frac{1}{\beta}\sum_{k,l=1}^m \mathrm{P}\Big(\sum_{i,j=1}^n a_{ij}\frac{\partial\xi_l}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)\frac{\partial^2\widetilde{f}}{\partial z_l\partial z_k}\,,\tag{3.8}$$

where $\xi_l$ is the $l$th component of the reaction coordinate map $\xi$, $1\le l\le m$. The following result is useful for deriving the effective dynamics (for a proof see Appendix C):

Proposition 3. *For $f=\widetilde{f}\circ\xi\in D(\mathcal{L})\cap H_0$ where $\widetilde{f}\in D(\widetilde{\mathcal{L}})$, we have $\mathrm{P}\mathcal{L}f=\widetilde{\mathcal{L}}\widetilde{f}$.*

In other words: $\widetilde{\mathcal{L}}$ is the generator of the full dynamics after projection onto the reaction coordinate function space.

**3.2. Deriving the effective dynamics.** In this subsection, based on the discussions in Subsection 3.1, especially the result of Proposition 3, we consider different approaches to the derivation of the effective dynamics.

**3.2.1. Ito's formula.** The most straightforward way to derive the effective dynamics may be by applying Ito's formula. To this end we use ideas from [51]: Given the reaction coordinate function $\xi$, we consider the quantity $\xi(x(s))\in\mathbb{R}^m$, where $x(s)\in\mathbb{R}^n$ satisfies dynamics (2.1). By Ito's formula and the definition of $\mathcal{L}$ in (2.2) , we have

$$\begin{aligned}d\xi_l(x(s)) =&\sum_{i=1}^n\frac{\partial\xi_l}{\partial x_i}b_i\,ds+\frac{1}{\beta}\sum_{i,j=1}^n\frac{\partial^2\xi_l}{\partial x_i\partial x_j}a_{ij}\,ds+\sqrt{2\beta^{-1}}\sum_{i=1}^n\sum_{j=1}^d\frac{\partial\xi_l}{\partial x_i}\sigma_{ij}\,dw_j(s)\,,\\ =&(\mathcal{L}\xi_l)(x(s))ds+\sqrt{2\beta^{-1}}\sum_{i=1}^n\sum_{j=1}^d\frac{\partial\xi_l(x(s))}{\partial x_i}\sigma_{ij}(x(s))\,dw_j(s)\,,\end{aligned}\tag{3.9}$$

for $1\le l\le m$. Notice that the right hand side of (3.9) not only depends on the value of the reaction coordinate $\xi(x(s))$ but also on the state of the original dynamics $x(s)$. To obtain a dynamics in the closed form (of the reaction coordinate), we replace the coefficients in (3.9) with their projections onto the space $\widetilde{H}$ defined in Subsection 3.1. That is, we consider the effective dynamics (following the terminology of [51])

$$dz(s)=\widetilde{b}(z(s))\,ds+\sqrt{2\beta^{-1}}\widetilde{\sigma}(z(s))\,dw(s)\,,\tag{3.10}$$

where $z(s)\in\mathbb{R}^m$ and coefficients $\widetilde{b}:\mathbb{R}^m\to\mathbb{R}^m$, $\widetilde{\sigma}:\mathbb{R}^m\to\mathbb{R}^{m\times m}$, with

$$\begin{aligned}\widetilde{b}_l(z)=&\mathrm{P}(\mathcal{L}\xi_l)(z)\,,\quad 1\le l\le m\,,\\ \widetilde{a}_{lk}(z)=&(\widetilde{\sigma}\widetilde{\sigma}^T)_{lk}(z)=\mathrm{P}\Big(\sum_{i,j=1}^n a_{ij}\frac{\partial\xi_l}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)(z)\,,\quad 1\le l,k\le m\,,\end{aligned}\tag{3.11}$$

for $\forall z \in \mathbb{R}^m$. Different from the setting in Section 2 where the noise intensity matrix $\sigma$ is allowed to have different numbers of rows and columns, here we can simply assume that $\widetilde{\sigma}$ is an $m \times m$ square matrix. Notice that derivatives of the reaction coordinate $\xi$ are involved in (3.11). On the other hand, recalling the definitions of the infinitesimal generator $\mathcal{L}$ and the quadratic variation and using (3.9) we directly obtain the following alternative formula for $\widetilde{b}$ and $\widetilde{a}$

$$\begin{aligned} \widetilde{b}_l(z) &= \lim_{s\to 0+} \mathbf{E}\Big(\frac{\xi_l(x(s)) - z_l}{s} \;\Big|\; x(0) \sim \mu_z\Big), \quad 1 \le l \le m, \\ \widetilde{a}_{lk}(z) &= \frac{\beta}{2} \lim_{s\to 0+} \mathbf{E}\Big(\frac{(\xi_l(x(s)) - z_l)(\xi_k(x(s)) - z_k)}{s} \;\Big|\; x(0) \sim \mu_z\Big), \quad 1 \le l,k \le m, \end{aligned} \tag{3.12}$$

where the conditional expectations are with respect to the ensemble of trajectories of the full dynamics (2.1) with the initial distribution $\mu_z$ on $\Sigma_z$ (see Subsection 3.1 for definitions). Numerical algorithms for simulating the effective dynamics (3.10) based on formulas (3.11) and (3.12) will be discussed in Subsection 3.4. Applying conditions (2.3) and (3.3), we also obtain

$$\eta^T \widetilde{a}(z)\eta = \mathrm{P}\big(\eta^T \nabla\xi^T a \nabla\xi\eta\big) \ge c_1 \mathrm{P}(|\nabla\xi\eta|^2) \ge c_1 c_2 |\eta|^2, \quad \forall z, \eta \in \mathbb{R}^m. \tag{3.13}$$

Using the form of the coefficient $\widetilde{b}$ and $\widetilde{a}$ from (3.11), it is clear that the operator $\widetilde{\mathcal{L}}$ defined by (3.8) can alternatively be written as

$$\widetilde{\mathcal{L}} = \sum_{l=1}^{m} \widetilde{b}_l \frac{\partial}{\partial z_l} + \frac{1}{\beta} \sum_{k,l=1}^{m} \widetilde{a}_{lk} \frac{\partial^2}{\partial z_l \partial z_k}, \tag{3.14}$$

which is exactly the form of the infinitesimal generator of dynamics (3.10). That is, the projection $\widetilde{\mathcal{L}}$ of the full generator to the reduced space $\widetilde{H}$ is identical to the generator of the effective dynamics (3.10).

Remark 2. *In general, the process $z(s)$ resulting from the effective dynamics (3.10) is not the same as the full process seen from the reaction coordinate space, $\xi(x(s))$, but just an approximation of it. In [51], the associated approximation error was studied for the case that (2.1) is a gradient system and $m = 1$, i.e. in the case that the reaction coordinate $\xi$ is a scalar function.*

Remark 3. *It is well known that the coefficients $b$ and $a$ in (2.1) satisfy the Kramers Moyal expansion [70]*

$$\begin{aligned} b_l(x) &= \lim_{s\to 0+} \mathbf{E}\Big(\frac{x(s) - x_l}{s} \;\Big|\; x(0) = x\Big), \quad 1 \le l \le n, \\ a_{lk}(x) &= \frac{\beta}{2} \lim_{s\to 0+} \mathbf{E}\Big(\frac{(x_l(s) - x_l)(x_k(s) - x_k)}{s} \;\Big|\; x(0) = x\Big), \quad 1 \le l,k \le n, \end{aligned} \tag{3.15}$$

*where $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ and the expectations are with respect to the ensemble of trajectories under dynamics (2.1). (3.12) shows that similar expressions are satisfied by the effective coefficients $\widetilde{b}$ and $\widetilde{a}$. In several papers related to the equation-free approach [47, 27, 36, 75, 49], the authors assumed the existence of Fokker-Planck equation of the coarse variables and compute the coefficients using formula (3.12). In connection with our arguments, it is thus clear that the coarse dynamics in the equation-free approach is actually the effective dynamics (3.10).*

**3.2.2. Mori-Zwanzig formalism.** In this part, we give a reasoning to motivate the effective dynamics (3.10) using the Mori-Zwanzig formalism. The Mori-Zwanzig formalism is an

operator approach to study the coarse-graining of dynamical systems [59, 88, 33]. Regarding the use of this formalism, it is helpful to point out that recent publications in this field (see [38, 41, 8, 9, 7, 15, 55] and references therein) mainly focused on the coarse graining of deterministic dynamical systems, especially Hamiltonian systems related to molecular dynamics, while we consider the model reduction of the ergodic stochastic process (2.1).

Recall that $\mathcal{L}$ in (2.2) is the infinitesimal generator of dynamics (2.1), P is the projection operator defined in Subsection 3.1, and the semigroup $T_t = e^{t\mathcal{L}} : H \to H$ is introduced in Section 2. Following the Mori-Zwanzig approach, for an arbitrary given function $f \in D(\mathcal{L}) \subset H$, we can directly verify the identity

$$e^{t\mathcal{L}} f = e^{t\mathcal{L}} \mathrm{P} f + \int_0^t e^{(t-s)\mathcal{L}} \mathrm{P}\mathcal{L} f_s \, ds + f_t \,, \qquad \forall t \geq 0 \,, \tag{3.16}$$

where $\mathrm{P}^\perp = \mathrm{id} - \mathrm{P}$ and $f_t = e^{t\mathrm{P}^\perp \mathcal{L}} \mathrm{P}^\perp f$. Similarly, we obtain

$$\frac{d}{dt} e^{t\mathcal{L}} f = e^{t\mathcal{L}} \mathcal{L} f = e^{t\mathcal{L}} \mathrm{P}\mathcal{L} f + \int_0^t e^{(t-s)\mathcal{L}} \mathrm{P}\mathcal{L} r_s \, ds + r_t \,, \qquad \forall t \geq 0 \,, \tag{3.17}$$

where $r_t = e^{t\mathrm{P}^\perp \mathcal{L}} \mathrm{P}^\perp \mathcal{L} f$.

We apply the operator P on both sides of (3.17) and notice that the third term on the right hand side disappears since $\mathrm{P} r_t \equiv 0$. Then we follow a similar reasoning as in [8, 9, 7, 38] to drop the memory (integral) term in (3.17). As a result, for function $f = \widetilde{f} \circ \xi \in H_0$, Proposition 3 entails that

$$\frac{d}{dt} \mathbf{E}\Big( \widetilde{f}\big(\xi(x(t))\big) \;\Big|\; x(0) \sim \mu_z \Big) \approx \mathbf{E}\Big( \widetilde{\mathcal{L}} \widetilde{f}\big(\xi(x(t))\big) \;\Big|\; x(0) \sim \mu_z \Big) \,, \tag{3.18}$$

where the expectation has the same meaning as in (3.12). Since our aim is to find a Markov process $z(t) \in \mathbb{R}^m$ which can approximate $\xi(x(t))$, we assume the equality will hold in (3.18) if the expectation is taken with respect to the law of $z(t)$, i.e.

$$\frac{d}{dt} \mathbf{E}\big( \widetilde{f}(z(t)) \mid z(0) = z \big) = \mathbf{E}\big( \widetilde{\mathcal{L}} \widetilde{f}(z(t)) \mid z(0) = z \big) \,, \quad \forall \widetilde{f} \in \widetilde{H} \,. \tag{3.19}$$

Then Dynkin's formula (Chapter 7 of [60]) suggests that $z(t) \in \mathbb{R}^m$ satisfies the SDE whose infinitesimal generator is given by $\widetilde{\mathcal{L}}$. From (3.14), we know this SDE coincides with (3.10) and $z(t)$ is the effective dynamics.

**3.2.3. Galerkin method.** In this part, we show that the effective dynamics (3.10) can also be motivated by applying the Galerkin method [11]. Recall that (see Subsection 2.2) the implicit time-scales of the full dynamics (2.1) are encoded in the eigenvalue problem

$$-\mathcal{L}\phi = \lambda \phi \,, \tag{3.20}$$

where $\mathcal{L}$ is the infinitesimal generator in (2.2). For simplicity, we assume the system is reversible and therefore all the eigenvalues are real. (3.20) can be written in weak form as

$$-\langle \mathcal{L}\phi, \psi \rangle_\mu = \lambda \langle \phi, \psi \rangle_\mu \,, \qquad \forall \psi \in H \,. \tag{3.21}$$

Now suppose we solve (3.21) using a Galerkin discretization with basis functions in the subspace $H_0$. Specifically, suppose $k$ basis functions $\phi_1, \phi_2, \cdots, \phi_k \in H_0$ are given with $\phi_i =$

$\widetilde{\phi}_i \circ \xi$, $\widetilde{\phi}_i \in \widetilde{H}$, $1 \leq i \leq k$ and consider the linear subspace spanned by these $k$ functions, $\mathrm{span}\{\phi_1, \phi_2, \cdots, \phi_k\} \subset H_0 \subset H$. The Galerkin discretization method amounts to finding an approximation of the solution $\phi$ to (3.21) in this linear subspace. Let $\sum_{i=1}^{k} \alpha_i \phi_i$ denote this approximate solution with (presently) unknown coefficients $\alpha_i \in \mathbb{R}$. Substituting it into (3.21) and using test functions $\psi = \phi_j$, $1 \leq j \leq k$ , we obtain

$$-\sum_{i=1}^{k} \alpha_i \langle \mathcal{L}\phi_i, \phi_j \rangle_\mu = \lambda \sum_{i=1}^{k} \alpha_i \langle \phi_i, \phi_j \rangle_\mu = \lambda \sum_{i=1}^{k} \alpha_i \langle \widetilde{\phi}_i, \widetilde{\phi}_j \rangle_\nu \,, \tag{3.22}$$

where the second equality follows from relation (3.7). Equation (3.22) is a finite-dimensional eigenvalue problem with the vector of the unknown coefficients $\alpha_i$ as its eigenvectors. Since $\phi_i, \phi_j \in H_0$, using the co-area formula (3.4) and Proposition 3, it follows that

$$\begin{aligned} &\langle \mathcal{L}\phi_i, \phi_j \rangle_\mu \\ =&\int_{\mathbb{R}^n} \mathcal{L}\phi_i(x)\, \widetilde{\phi}_j(\xi(x))\, \rho(x)\, dx \\ =&\int_{\mathbb{R}^m} (\mathrm{P}\mathcal{L}\phi_i)(z)\, \widetilde{\phi}_j(z)\, Q(z)\, dz \\ =&\langle \widetilde{\mathcal{L}}\widetilde{\phi}_i, \widetilde{\phi}_j \rangle_\nu \,, \end{aligned} \tag{3.23}$$

where the operator $\widetilde{\mathcal{L}}$ is defined in (3.8). Therefore, equation (3.22) takes the form

$$-\sum_{i=1}^{k} \alpha_i \langle \widetilde{\mathcal{L}}\widetilde{\phi}_i, \widetilde{\phi}_j \rangle_\nu = \lambda \sum_{i=1}^{k} \alpha_i \langle \widetilde{\phi}_i, \widetilde{\phi}_j \rangle_\nu \,. \tag{3.24}$$

Now notice in addition that, on the other hand, the same equation (3.24) can be obtained if we consider the Galerkin discretization of the eigenvalue problem

$$-\widetilde{\mathcal{L}}\widetilde{\phi} = \lambda \widetilde{\phi} \tag{3.25}$$

on the subspace $\mathrm{span}\{\widetilde{\phi}_1, \widetilde{\phi}_2, \cdots, \widetilde{\phi}_k\} \subset \widetilde{H}$. However, from (3.14), we know that equation (3.25) is the corresponding eigenvalue problem for the effective dynamics (3.10).

Thus we have shown that the Galerkin discretization of the eigenvalue problem of the generator of the full dynamics (2.1) is identical to the Galerkin discretization of the eigenvalue problem of the projected generator (which is the generator of the effective dynamics (3.10). This also suggests to approximate the original full dynamics using the effective dynamics (3.10). The exact relation between the eigenvalues of the full dynamics and those of the effective dynamics will be further studied in Subsection 3.3.

**3.2.4. Averaging and homogenization.** Here we discuss two special cases when $x = (z, y)$ and the dynamics is either

$$\begin{aligned} dz(s) =& f(z(s), y(s))ds + \sigma_1(z(s), y(s))dw^1(s) \\ dy(s) =& \frac{1}{\epsilon} g(z(s), y(s))ds + \frac{1}{\sqrt{\epsilon}} \sigma_2(z(s), y(s))dw^2(s) \,, \end{aligned} \tag{3.26}$$

or

$$
\begin{aligned}
dz(s) =& \frac{1}{\epsilon} f_0(z(s), y(s))ds + f(z(s), y(s))ds + \sigma_1(z(s), y(s))dw^1(s) \\
dy(s) =& \frac{1}{\epsilon^2} g(z(s), y(s))ds + \frac{1}{\epsilon}\sigma_2(z(s), y(s))dw^2(s)\,,
\end{aligned}
\tag{3.27}
$$

where $x = (z, y) \in \mathbb{R}^m \times \mathbb{R}^{n-m} = \mathbb{R}^n$, and the time scale separation between $z$ and $y$ is explicitly characterized by the parameter $\epsilon \ll 1$. Functions $f, f_0 : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^{n-m}$, $\sigma_1 : \mathbb{R}^n \to \mathbb{R}^{m\times d_1}$, $\sigma_2 : \mathbb{R}^n \to \mathbb{R}^{(n-m)\times d_2}$ are assumed to be smooth. Dynamics (3.26) and (3.27) can be obtained from (2.1) by choosing

$$
b = \begin{pmatrix} f \\ \frac{1}{\epsilon} g \end{pmatrix},\ \sigma = \sqrt{\frac{\beta}{2}} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \frac{1}{\sqrt{\epsilon}}\sigma_2 \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} \frac{1}{\epsilon} f_0 + f \\ \frac{1}{\epsilon^2} g \end{pmatrix},\ \sigma = \sqrt{\frac{\beta}{2}} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \frac{1}{\epsilon}\sigma_2 \end{pmatrix},
\tag{3.28}
$$

respectively. They are referred to as Averaging and Homogenization systems. They are considered as model problems for multiscale methods and have been extensively investigated [64]. As in [64], we assume the dynamics

$$
dy(s) = g(z, y(s))ds + \sigma_2(z, y(s))dw^2(s) \tag{3.29}
$$

is ergodic and has a unique invariant measure whose probability density is $\rho_z(\cdot)$ for all fixed $z \in \mathbb{R}^m$. In the Homogenization case, we assume the validity of the so-called centering condition for $f_0$ in (3.27) :

$$
\int_{\mathbb{R}^{n-m}} f_0(z, y)\rho_z(y)\, dy = 0\,, \quad \forall z \in \mathbb{R}^m\,. \tag{3.30}
$$

Now, we consider the reaction coordinate $\xi : \mathbb{R}^n \to \mathbb{R}^m$ given by $\xi(x) = \xi(z, y) = z$ and the measure $d\mu_z(y) = \rho_z(y)dy$.

Then, on the one hand, the abstract form the SDE (3.9) for $\xi(x(s)) = z(s)$ boils down to the first equation of (3.26) in the Averaging case, and to (3.27) in the Homogenization case, respectively.

On the other hand, for our effective dynamics of the form (3.10),

$$
dz(s) = \widetilde{f}(z(s))ds + \widetilde{\sigma}(z(s))dw_s\,, \tag{3.31}
$$

direct computations, using the above ergodicity and centering assumptions for both cases, result in the following coefficients:

$$
\widetilde{f}(z) = \int_{\mathbb{R}^{n-m}} f(z, y)\rho_z(y)dy\,, \quad \widetilde{\sigma}(z)\widetilde{\sigma}(z)^T = \int_{\mathbb{R}^{n-m}} \sigma_1(z, y)\sigma_1(z, y)^T\rho_z(y)dy\,, \tag{3.32}
$$

for all $z \in \mathbb{R}^m$.

In [64] the limit dynamics of (3.26) and (3.27) for $\epsilon \to 0$ is studied. Thus, we can compare the effective dynamics to the limit dynamics in both cases: In the Averaging case, the effective dynamics (3.31) obtained using conditional expectation is the same as the limit dynamics given in [64]. In the Homogenization case, however, the effective dynamics (3.31) is different from the limit dynamics. Roughly speaking, although the mean contribution of the term $f_0$ in (3.27) vanishes due to the centering condition (3.30), one needs to take into account its fluctuation in order to obtain the correct limit dynamics. This requires to solve the "cell problem" and

to add correction terms to (3.32), see [64] for details. Such corrections are ignored by the model reduction approach using conditional expectation considered in our present work. The underlying differences between Averaging and Homogenization systems have also been observed when studying the parameter estimation problems for multiscale diffusions. We refer to [63, 62] for more detailed discussions.

**3.3. Properties of the effective dynamics.** In this subsection we consider some properties of the effective dynamics (3.10).

**3.3.1. Ergodicity, stationary probability velocity.** We first consider the stationary probability velocity and the invariant measure of the effective dynamics. See Section 2 for related definitions. Notice that the invariant measure has been studied in [51] when the original dynamics (2.1) is a gradient system.

PROPOSITION 4. *Assume that conditions ()2.3) and (3.3) hold. The effective dynamics (3.10) has a unique invariant measure $\nu$ whose probability density with respect to Lebesgue measure is $Q$ (Subsection 3.1). Define $\widetilde{J}^{st} \in \mathbb{R}^m$ whose components are given by*

$$\widetilde{J}^{st}_k(z) = \mathrm{P}(J^{st} \cdot \nabla \xi_k)\,, \quad 1 \le k \le m\,, \tag{3.33}$$

*for $z \in \mathbb{R}^m$ where $J^{st}$ is the stationary probability velocity of dynamics (2.1). Then we have*

$$-\mathrm{div}(\widetilde{J}^{st}(z)\, Q(z)) = 0\,, \tag{3.34}$$

*i.e. $\widetilde{J}^{st}$ is the stationary probability velocity of the effective dynamics (3.10).*

*Proof.* For any smooth function $\widetilde{\phi} : \mathbb{R}^m \to \mathbb{R}$, using integration by parts, chain rule and the co-area formula (3.4), we have

$$\begin{aligned}
&-\int_{\mathbb{R}^m} \mathrm{div}\big(\widetilde{J}^{st}(z)\, Q(z)\big)\, \widetilde{\phi}(z)\, dz \\
&= \int_{\mathbb{R}^m} \widetilde{J}^{st}(z) \cdot \nabla \widetilde{\phi}(z)\, Q(z)\, dz = \sum_{k=1}^m \int_{\mathbb{R}^m} \mathrm{P}(J^{st} \cdot \nabla \xi_k)(z) \frac{\partial \widetilde{\phi}}{\partial z_k}(z)\, Q(z)\, dz \\
&= \int_{\mathbb{R}^m} \mathrm{P}\big(J^{st} \cdot \nabla(\widetilde{\phi} \circ \xi)\big)(z)\, Q(z)\, dz = \int_{\mathbb{R}^n} J^{st}(x) \cdot \nabla(\widetilde{\phi} \circ \xi)(x)\, \rho(x)\, dx \\
&= -\int_{\mathbb{R}^n} \mathrm{div}(J^{st}\rho)\, \widetilde{\phi} \circ \xi\, dx = 0\,,
\end{aligned}$$

which implies (3.34). Using the identities (4.6) and (3.11), we can obtain

$$\widetilde{b}_l(z) = \widetilde{J}^{st}_l(z) + \frac{1}{\beta Q(z)} \sum_{k=1}^m \frac{\partial}{\partial z_k}\Big(\widetilde{a}_{lk}(z)\, Q(z)\Big)\,, \quad 1 \le l \le m\,. \tag{3.35}$$

It can be readily checked that

$$\begin{aligned}
&-\sum_{l=1}^m \frac{\partial}{\partial z_l}\big(\widetilde{b}_l(z) Q(z)\big) + \frac{1}{\beta} \sum_{l,k=1}^m \frac{\partial^2\big(\widetilde{a}_{lk}(z)\, Q(z)\big)}{\partial z_l \partial z_k} \\
&= -\sum_{l=1}^m \frac{\partial}{\partial z_l}\Big[\widetilde{J}^{st}_l(z)\, Q(z) + \frac{1}{\beta} \sum_{k=1}^m \frac{\partial}{\partial z_k}\Big(\widetilde{a}_{lk}(z)\, Q(z)\Big)\Big] + \frac{1}{\beta} \sum_{l,k=1}^m \frac{\partial^2(\widetilde{a}_{lk}(z)\, Q(z))}{\partial z_l \partial z_k} \\
&= -\,\mathrm{div}\big(\widetilde{J}^{st}(z)\, Q(z)\big) = 0\,.
\end{aligned}$$

Furthermore, conditions (2.3) and (3.3) imply that (3.13) holds for matrix $\widetilde{a}$. Comparing this with equation (2.6) and the discussions in Section 2, this implies that the effective dynamics (3.10) has a unique invariant measure $\nu$ whose probability density is $Q$. □

Since the effective dynamics is ergodic with the invariant measure $\nu$, we know it satisfies all the properties discussed in Section 2. For example, similarly to (2.16) and (2.17), we can rewrite the dynamics as

$$dz_l(s) = \widetilde{J}_l^{st} ds + \frac{1}{\beta Q} \sum_{k=1}^{m} \frac{\partial(\widetilde{a}_{lk} Q)}{\partial z_k} ds + \sqrt{2\beta^{-1}} \sum_{k=1}^{m} \widetilde{\sigma}_{lk}\, dw_k(s)\,, \quad 1 \le l \le m\,, \;\; s \ge 0\,, \tag{3.36}$$

and the infinitesimal generator $\widetilde{\mathcal{L}}$ in (3.8) allows the decomposition

$$\widetilde{\mathcal{L}} = \widetilde{\mathcal{L}}_a + \widetilde{\mathcal{L}}_s = \widetilde{J}^{st} \cdot \nabla + \frac{1}{\beta Q} \nabla (Q \widetilde{a} \nabla \cdot)\,. \tag{3.37}$$

Especially, Theorem 4 implies that when the original dynamics (2.1) is reversible, the effective dynamics (3.10) is also reversible.

We can define the metric tensor $\widetilde{g} = \widetilde{a}^{-1}$ and let $\widetilde{G} = \det \widetilde{g}$. The free energy is defined as $F(z) = -\beta^{-1} \ln Q(z)$. Repeating the argument from Subsection 2.3, we can rewrite the effective dynamics as

$$dz(s) = \big[\widetilde{J}^{st} - \nabla_{\widetilde{g}} (F + \frac{1}{2\beta} \ln \widetilde{G})\big]\, ds + \sqrt{2\beta^{-1}} dW(s)\,, \tag{3.38}$$

where $\nabla_{\widetilde{g}}$ is the gradient operator, $W(s)$ is the Brownian motion on the $m$-dimensional manifold $M$. It is known that both the free energy and the diffusion coefficients are relevant when studying model reduction of diffusion processes [51]. Here we look at it in a slightly different way : the effective dynamics can be considered as driven by the mean force and the "standard" Brownian motion if we take a manifold point of view (Subsection 2.3).

**3.3.2. Time scales.** In this part, we consider the relations between the time scales of the full dynamics (2.1) and the effective dynamics (3.10). As before, we assume that the full dynamics (2.1) is reversible, then from the discussions above we know that the effective dynamics is reversible as well. In addition we assume that the essential spectra of both $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ are empty, and therefore both spectra consist of isolated eigenvalues (see Subsection 2.2 for discussions). Let the functions $\varphi_i \in H$ be the orthonormal eigenfunctions of operator $-\mathcal{L}$, i.e. $-\mathcal{L}\varphi_i = \lambda_i \varphi_i$ , with eigenvalues

$$0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots\,, \tag{3.39}$$

and $\varphi_0 \equiv 1$. Similarly, let $\widetilde{\varphi}_i \in \widetilde{H}$ be the orthonormal eigenfunctions of operator $-\widetilde{\mathcal{L}}$ corresponding to eigenvalues $\widetilde{\lambda}_i$, where

$$0 = \widetilde{\lambda}_0 < \widetilde{\lambda}_1 \le \widetilde{\lambda}_2 \le \cdots\,. \tag{3.40}$$

In order to get a relation between the $\lambda_i$ and the $\widetilde{\lambda}_i$ we recall the following well known theorem [4, 76].

THEOREM 3.1 (Min-Max). *Let $\lambda_i$ and $\widetilde{\lambda}_i$ be the eigenvalues in (3.39), (3.40), corresponding to operator $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ respectively. For $i \ge 0$, we have*

$$\begin{aligned} \lambda_i &= \min_{H_{i+1}} \max_{f \in H_{i+1},\, |f|_\mu = 1} \langle -\mathcal{L} f, f \rangle_\mu\,, \\ \widetilde{\lambda}_i &= \min_{\widetilde{H}_{i+1}} \max_{\widetilde{f} \in \widetilde{H}_{i+1},\, |\widetilde{f}|_\nu = 1} \langle -\widetilde{\mathcal{L}} \widetilde{f}, \widetilde{f} \rangle_\nu\,, \end{aligned} \tag{3.41}$$

*where $H_{i+1}$ and $\widetilde{H}_{i+1}$ are $(i+1)$-dimensional subspaces of $H$ and $\widetilde{H}$ respectively.*

Using Theorem 3.1, we obtain the following result on the relation between the $\lambda_i$ and the $\widetilde{\lambda}_i$:

PROPOSITION 5. *For $i \geq 0$, let $\varphi_i$ and $\widetilde{\varphi}_i$ be the normalized eigenfunctions of operator $-\mathcal{L}$ and $-\widetilde{\mathcal{L}}$ corresponding to eigenvalues $\lambda_i$ and $\widetilde{\lambda}_i$, respectively. We have*

$$\lambda_i \leq \widetilde{\lambda}_i \leq \lambda_i + \frac{1}{\beta}\langle a\nabla(\varphi_i - \widetilde{\varphi}_i \circ \xi), \nabla(\varphi_i - \widetilde{\varphi}_i \circ \xi)\rangle_\mu \,. \tag{3.42}$$

*Proof.* For the lower bound, we consider the subspace $\widetilde{S}_{i+1} = \mathrm{span}\{\widetilde{\varphi}_0, \widetilde{\varphi}_1, \cdots, \widetilde{\varphi}_i\} \subset \widetilde{H}$. Then it is direct to verify that the minimum is achieved in the second equation of (3.41) with subspace $\widetilde{S}_{i+1}$. Also define $S_{i+1} = \{\widetilde{f} \circ \xi \mid \widetilde{f} \in \widetilde{S}_{i+1}\}$, which is an $(i+1)$-dimensional subspace of $H$. Applying Theorem 3.1, we have

$$\begin{aligned}
\widetilde{\lambda}_i &= \max_{\widetilde{f}\in\widetilde{S}_{i+1},\, |\widetilde{f}|_\nu=1} \langle -\widetilde{\mathcal{L}}\widetilde{f}, \widetilde{f}\rangle_\nu \\
&= \max_{\widetilde{f}\in\widetilde{S}_{i+1},\, |\widetilde{f}|_\nu=1} \langle -\mathcal{L}(\widetilde{f} \circ \xi), \widetilde{f} \circ \xi\rangle_\mu \\
&= \max_{f\in S_{i+1},\, |f|_\mu=1} \langle -\mathcal{L}f, f\rangle_\mu \;\geq \min_{H_{i+1}} \max_{f\in H_{i+1},\, |f|_\mu=1} \langle -\mathcal{L}f, f\rangle_\mu = \lambda_i \,.
\end{aligned}$$

Notice that we have used the relation (3.23), i.e. $\langle \mathcal{L}(\widetilde{f} \circ \xi), (\widetilde{f} \circ \xi)\rangle_\mu = \langle \widetilde{\mathcal{L}}\widetilde{f}, \widetilde{f}\rangle_\nu$, $\forall \widetilde{f} \in \widetilde{H}$.

For the upper bound, using (2.13) in the reversible case and $|\varphi_i|_\mu = |\widetilde{\varphi}_i|_\nu = |\widetilde{\varphi}_i \circ \xi|_\mu = 1$, we have

$$\begin{aligned}
&\frac{1}{\beta}\langle a\nabla(\varphi_i - \widetilde{\varphi}_i \circ \xi), \nabla(\varphi_i - \widetilde{\varphi}_i \circ \xi)\rangle_\mu \\
=& -\langle \mathcal{L}(\varphi_i - \widetilde{\varphi}_i \circ \xi), \varphi_i - \widetilde{\varphi}_i \circ \xi\rangle_\mu \\
=& \lambda_i\langle \varphi_i, \varphi_i - \widetilde{\varphi}_i \circ \xi\rangle_\mu - \lambda_i\langle \varphi_i, \widetilde{\varphi}_i \circ \xi\rangle_\mu - \langle \mathcal{L}(\widetilde{\varphi}_i \circ \xi), \widetilde{\varphi}_i \circ \xi\rangle_\mu \\
=& \lambda_i - 2\lambda_i\langle \varphi_i, \widetilde{\varphi}_i \circ \xi\rangle_\mu - \langle \widetilde{\mathcal{L}}\widetilde{\varphi}_i, \widetilde{\varphi}_i\rangle_\nu \\
=& \lambda_i - 2\lambda_i\langle \varphi_i, \widetilde{\varphi}_i \circ \xi\rangle_\mu + \widetilde{\lambda}_i \\
=& \widetilde{\lambda}_i - \lambda_i + \lambda_i\langle \varphi_i - \widetilde{\varphi}_i \circ \xi, \varphi_i - \widetilde{\varphi}_i \circ \xi\rangle_\mu \geq \widetilde{\lambda}_i - \lambda_i \,.
\end{aligned}$$

□

REMARK 4. *Proposition 5 shows that all eigenvalues $\lambda_i$ (the time scales $\lambda_i^{-1}$) of the original full dynamics are either preserved or overestimated (underestimated) if we compute them using the effective dynamics. Furthermore, roughly speaking, the approximation error between corresponding eigenvalues is bounded by the error between the corresponding eigenfunctions.*

*Estimating the eigenvalue approximation error by Galerkin methods is an important topic in the field of finite element methods [4, 5]. Our proof above is also based on a basic analysis in [5]. Different from the setting in finite element methods, where a sequence of subspaces corresponding to mesh resolutions are considered, here we have only one subspace which is determined by the reactive coordinate.*

*Eigenvalue (time scales) approximations have also been studied for the transfer operator (which is bounded and whose eigenvalues are between $[0, 1]$) in the context of Markov state models [67], which aim at approximating high-dimensional diffusion processes by jump processes on a finite state space. See [72, 18, 71] for more details.*

Finally, we consider the special case when the reaction coordinate function $\xi$ is defined by the eigenfunctions $\boldsymbol{\varphi}(x) = \big(\varphi_1(x), \varphi_2(x), \cdots, \varphi_m(x)\big) \in \mathbb{R}^m$. Specifically, we assume $\xi = F \circ \boldsymbol{\varphi}(x) \in \mathbb{R}^m$, $\forall x \in \mathbb{R}^n$, where $F : \mathrm{Im}(\boldsymbol{\varphi}) \to \mathbb{R}^m$ is a $\mathcal{C}^2$ diffeomorphism from $\mathrm{Im}(\boldsymbol{\varphi}) \subset \mathbb{R}^m$ to $\mathbb{R}^m$ and we denote $G = F^{-1}$. The reason to introduce the function $F$ is to guarantee that the coefficients (3.11) are well defined on the whole space $\mathbb{R}^m$. Furthermore, it demonstrates that there is a freedom in choosing the reaction coordinate $\xi$.

To compute the coefficients in (3.11), we notice that the projection operator P is the expectation conditioned on

$$\xi(x) = F \circ \boldsymbol{\varphi}(x) = z \in \mathbb{R}^m \quad \Longleftrightarrow \quad \boldsymbol{\varphi}(x) = G(z)\,.$$

For $1 \le l \le m$, denote $F^{(l)}(z), G^{(l)}(z)$ as the $l$th component of function $F, G$ respectively. Then, using the fact that $\varphi_i$ are eigenfunctions, we can write the coefficients of the effective dynamics (3.10) as

$$\begin{aligned}
\widetilde{b}_l(z) &= -\sum_{r=1}^{m} \lambda_r \Big(\frac{\partial F^{(l)}}{\partial z_r} \circ G\Big)(z) G^{(r)}(z) + \sum_{r,r'=1}^{m} \Big(\frac{\partial^2 F^{(l)}}{\partial z_r \partial z_{r'}} \circ G\Big)(z) \mathrm{P}\Big(\sum_{i,j=1}^{n} a_{ij} \frac{\partial \varphi_r}{\partial x_i} \frac{\partial \varphi_{r'}}{\partial x_j}\Big)(z)\,, \\
\widetilde{a}_{kl}(z) &= \sum_{r,r'=1}^{m} \Big(\frac{\partial F^{(k)}}{\partial z_r} \circ G\Big)(z) \Big(\frac{\partial F^{(l)}}{\partial z_{r'}} \circ G\Big)(z)\, \mathrm{P}\Big(\sum_{i,j=1}^{n} a_{ij} \frac{\partial \varphi_r}{\partial x_i} \frac{\partial \varphi_{r'}}{\partial x_j}\Big)(z)\,,
\end{aligned} \tag{3.43}$$

where $z \in \mathbb{R}^m$, $1 \le l, k \le m$. Applying the chain rule and differentiating the identity $G(F(z)) = z$ twice, then using (3.43), we can verify that $\widetilde{\mathcal{L}} G^{(k)} = \lambda_k G^{(k)}$, $1 \le k \le m$, i.e. $\lambda_k$, $G^{(k)}$ are the eigenvalues and eigenfunctions of the infinitesimal operator $\widetilde{\mathcal{L}}$, respectively.

In this case, we see that the corresponding eigenvalues $\lambda_k$ of the original dynamics are preserved in the effective dynamics and the equality is achieved in Proposition 5.

**3.3.3. Reaction rates.** In this part we discuss some issues related to the reaction rates defined in the transition path theory (TPT) [25, 78, 24]. Suppose that two disjoint closed set $A, B \subset \mathbb{R}^n$ are given. In many applications, it is often important to know the frequency for the system to enter set $B$ when starting from set $A$. The reaction rate between set $A$ and $B$ which is defined in the TPT theory is a quantity to characterize this frequency and we will denote it by $k_{AB}$. It turns out that the committor function $q : \mathbb{R}^n \to \mathbb{R}$, which satisfies

$$\begin{aligned}
\mathcal{L} q &= 0\,, \quad x \in (A \cup B)^c\,, \\
q|_A &= 0\,, \quad q|_B = 1\,,
\end{aligned} \tag{3.44}$$

plays an important role for computing the rate $k_{AB}$. In fact, we have

$$\begin{aligned}
k_{AB} =& \frac{1}{\beta} \int_{\mathbb{R}^n} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial q(x)}{\partial x_i} \frac{\partial q(x)}{\partial x_j}\, \delta(q(x) - z)\, \rho(x)\, dx\,, \quad \forall z \in [0,1]\,, \\
=& \frac{1}{\beta} \int_{(A \cup B)^c} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial q(x)}{\partial x_i} \frac{\partial q(x)}{\partial x_j}\, \rho(x)\, dx\,,
\end{aligned} \tag{3.45}$$

see equations (43) and (53) in [78] for more details. In general, however, solving the PDE (3.44) and computing the reaction rate using formula (3.45) is impractical for high dimensions. In order to see whether we can reduce the dimension by using reaction coordinates $\xi$, we first suppose

that $A = \xi^{-1}(\widetilde{A})$ and $B = \xi^{-1}(\widetilde{B})$ for some disjoint closed sets $\widetilde{A}, \widetilde{B} \subset \mathbb{R}^m$, i.e. sets $A, B$ are defined through the reaction coordinate. Now consider that we compute the committor function $q$ from (3.44) by using the Galerkin method using the linear subspace of $\widetilde{H}$ as above and assume that this procedure results in an approximation in the form $\widetilde{q}(\xi(x))$ with $\widetilde{q} : \mathbb{R}^m \to \mathbb{R}$. Then, by proceeding in a very similar way as in Subsection 3.2, it turns out that $\widetilde{q}$ satisfies

$$\begin{aligned} &\widetilde{\mathcal{L}}\,\widetilde{q} = 0\,, \quad z \in (\widetilde{A} \cup \widetilde{B})^c\,, \\ &\widetilde{q}|_{\widetilde{A}} = 0\,, \quad \widetilde{q}|_{\widetilde{B}} = 1\,, \end{aligned} \tag{3.46}$$

i.e. $\widetilde{q}$ is actually the committor function of the effective dynamics (3.10) corresponding to sets $\widetilde{A}$ and $\widetilde{B}$. With the approximate solution $\widetilde{q}(\xi(x))$, using the chain rule and the co-area formula (3.4), formula (3.45) yields

$$\begin{aligned} k_{AB} &\approx \frac{1}{\beta} \int_{(A\cup B)^c} \sum_{i,j=1}^{n} a_{ij} \frac{\partial \widetilde{q}(\xi(x))}{\partial x_i} \frac{\partial \widetilde{q}(\xi(x))}{\partial x_j}\, \rho(x)\, dx \\ &= \frac{1}{\beta} \int_{(A\cup B)^c} \sum_{k,l=1}^{m} \frac{\partial \widetilde{q}(\xi(x))}{\partial z_k} \frac{\partial \widetilde{q}(\xi(x))}{\partial z_l} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial \xi_k(x)}{\partial x_i} \frac{\partial \xi_l(x)}{\partial x_j}\, \rho(x)\, dx \\ &= \frac{1}{\beta} \int_{(\widetilde{A}\cup \widetilde{B})^c} \sum_{k,l=1}^{m} \frac{\partial \widetilde{q}(z)}{\partial z_k} \frac{\partial \widetilde{q}(z)}{\partial z_l} \widetilde{a}_{kl}(z)\, Q(z)\, dz = \widetilde{k}_{\widetilde{A}\widetilde{B}}\,, \end{aligned} \tag{3.47}$$

where $\widetilde{k}_{\widetilde{A}\widetilde{B}}$ is the reaction rate for the effective dynamics between set $\widetilde{A}$ and $\widetilde{B}$.

In summary, if we compute the committor function $q$ by solving the PDE (3.44) using the Galerkin method based on a finite-dimensional linear subspace of $\widetilde{H}$ and then compute the reaction rate based on the approximate solution $\widetilde{q}(\xi(x))$, then what we obtained is actually the reaction rate for the effective dynamics.

In the reversible case, for any smooth function $f : \mathbb{R}^n \to \mathbb{R}$ we define

$$I(f) = \frac{1}{\beta} \int_{(A\cup B)^c} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial f(x)}{\partial x_i} \frac{\partial f(x)}{\partial x_j}\, \rho(x)\, dx\,. \tag{3.48}$$

Then from (3.44) and (3.45) we know

$$k_{AB} = I(q) = \min_f I(f)\,, \tag{3.49}$$

where the minimum is taken wrt all smooth functions $f$ such that $f|_A = 0, f|_B = 1$ [25].

Similarly, we define $\widetilde{I}(\cdot)$ for the effective dynamics. Then, the derivations in (3.47) show that $\widetilde{k}_{\widetilde{A}\widetilde{B}} = \widetilde{I}(\widetilde{q}) = I(\widetilde{q} \circ \xi)$. Applying the calculus of variations to the functional $I(\cdot)$, we obtain

$$\frac{1}{\beta} \int_{(A\cup B)^c} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial q(x)}{\partial x_i} \frac{\partial h(x)}{\partial x_j}\, \rho(x)\, dx = 0\,, \tag{3.50}$$

which holds for all smooth functions $h : \mathbb{R}^n \to \mathbb{R}$, $h|_{A\cup B} = 0$. For all $f : \mathbb{R}^n \to \mathbb{R}$ with $f|_A = 0, f|_B = 1$, since $(f - q)|_{A\cup B} = 0$, (3.50) yields that

$$\begin{aligned} I(f) =& I(q) + \frac{1}{\beta} \int_{(A\cup B)^c} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial (q(x) - f(x))}{\partial x_i} \frac{\partial (q(x) - f(x))}{\partial x_j}\, \rho(x)\, dx \\ =& I(q) + I(q - f)\,. \end{aligned} \tag{3.51}$$

Combining the above results, we conclude

PROPOSITION 6. *Assume that full dynamics (2.1) is reversible. Let $k_{AB}$ and $\widetilde{k}_{\widetilde{A}\widetilde{B}}$ be the reaction rates of the full dynamics between the sets $A$ and $B$ and the effective dynamics between $\widetilde{A}$ and $\widetilde{B}$, respectively, as defined above. Let $q$ and $\widetilde{q}$ denote the corresponding committor functions, respectively. Then, we have*

$$k_{AB} \leq \widetilde{k}_{\widetilde{A}\widetilde{B}} = k_{AB} + \frac{1}{\beta}\int_{(A\cup B)^c} \sum_{i,j=1}^{n} a_{ij} \frac{\partial(q-\widetilde{q}\circ\xi)}{\partial x_i}\frac{\partial(q-\widetilde{q}\circ\xi)}{\partial x_j}\,\rho\,dx\,. \tag{3.52}$$

*Proof.* By inserting $f = \widetilde{q}\circ\xi$ into equation (3.51), we obtain $I(\widetilde{q}\circ\xi) = I(q) + I(q-\widetilde{q}\circ\xi)$. The assertion follows by noticing that $\widetilde{k}_{\widetilde{A}\widetilde{B}} = I(\widetilde{q}\circ\xi)$ and $k_{AB} = I(q)$. □

The above results imply that the reaction rate of the full dynamics is either preserved or overestimated when we compute it using the effective dynamics. Furthermore, the approximation error of the reaction rate depends on the approximation error between the respective committor functions.

Finally we consider the special case when the reaction coordinate is constructed from the committor function (3.44), i.e., the idea is to set $\xi = q$. To this end, we need some preparations. First of all, in order to define a reaction coordinate mapping onto $\mathbb{R}$, we need to modify the value of the committor function $q$ in $\mathring{A}\cup\mathring{B}$, the interior of the set $A\cup B$. We consider a modification $\bar{q}$ of $q$ such that

$$\begin{aligned} &\bar{q}:\mathbb{R}^n\to\mathbb{R} \quad \text{is } \mathcal{C}^2 \text{ smooth}\,,\\ &\bar{q}(\mathring{A}) = (-\infty,0)\,, \quad \bar{q}(\mathring{B}) = (1,+\infty),\\ &\bar{q}(x) = q(x)\,, \quad \forall x\in\overline{(A\cup B)^c}\,. \end{aligned} \tag{3.53}$$

Based on this, we especially know that $\mathcal{L}\bar{q} = 0$ holds on $\overline{(A\cup B)^c}$. Then we consider the reaction coordinate $\xi(x) = \bar{q}(x)$, $x\in\mathbb{R}^n$. From (3.53) and the boundary conditions in (3.44), we know $A = \xi^{-1}((-\infty,0])$, $B = \xi^{-1}([1,+\infty))$, which implies that $\widetilde{A} = (-\infty,0]$, $\widetilde{B} = [1,+\infty)$. Notice that for $z\in(\widetilde{A}\cup\widetilde{B})^c = (0,1)$, (3.53) implies that $\bar{q}^{-1}(z)\subset(A\cup B)^c$. Then, based on (3.11), the coefficients satisfy

$$\widetilde{b}(z) = 0, \quad \widetilde{\sigma}^2(z) = \mathrm{P}\Big(\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial q(x)}{\partial x_i}\frac{\partial q(x)}{\partial x_j}\Big)\,, \quad \forall z\in(0,1)\,, \tag{3.54}$$

where the projection operator P is defined through the expectation conditioning on $\bar{q}(x) = q(x) = z$. We also know from the rate formula (3.45) that $\widetilde{\sigma}^2(z) = \frac{\beta k_{AB}}{Q(z)}$. Therefore, on the interval $(0,1)$, the effective dynamics satisfies the SDE

$$dz(s) = \sqrt{2\frac{k_{AB}}{Q(z(s))}}\,dw(s)\,, \tag{3.55}$$

where $z(s)\in(0,1)$ and $w(s)$ is a one-dimensional Brownian motion. Accordingly, the infinitesimal generator on $(0,1)$ is $\widetilde{\mathcal{L}} = \frac{k_{AB}}{Q(z)}\frac{d^2}{dz^2}$. Let $\widetilde{q}(z)$ be the committor function of the effective dynamics corresponding to set $\widetilde{A}$ and $\widetilde{B}$. Due to the specific form of $\widetilde{\mathcal{L}}$, equation (3.46) can be analytically solved and we obtain that $\widetilde{q}(z) = z$, for $0\leq z\leq 1$. Applying formula (3.47), we can obtain that

$$\widetilde{k}_{\widetilde{A}\widetilde{B}} = \int_0^1 Q(z)|\widetilde{q}\,'(z)|^2\frac{k_{AB}}{Q(z)}dz = k_{AB}\,. \tag{3.56}$$

In conclusion, the reaction rate constant $k_{AB}$ is preserved by the effective dynamics (and equality is achieved in Proposition 6) when we take the committor function $\bar{q}(\cdot)$ as the reaction coordinate. The choice of optimal reaction coordinates is also discussed in [56].

**3.4. Numerical methods.** In this section, we discuss two different options for performing numerical simulations of the effective dynamics. When considering numerical simulations one has to overcome the obstacle that the equations for the coefficients $\widetilde{b}$ and $\widetilde{\sigma}$ determining the effective dynamics are by no means explicit. That is, in most realistic cases the evaluation of the coefficients is a difficult numerical problem in itself. The literature contains several proposals for overcoming this obstacle. The two most prominent ones, the equation-free approach [44, 42, 43, 40, 75, 49] and the heterogeneous multiscale method (HMM) [20, 21, 1, 77], are utilizing short trajectories of the full dynamics in order to compute the coefficients of the effective dynamics and advocate that this is efficient as long as required full trajectories are much shorter than the resulting time-steps of the numerical simulation of the effective dynamics. Next we will demonstrate how these two approaches blend into the general results established above.

Apart from simulations of the effective dynamics there is a huge number of articles related to numerical methods for sampling the free energy space along a reaction coordinate; various algorithms have been proposed [40, 2, 80, 10, 17, 12, 53]. These are only partially related to the present investigation but contain tools that we can utilize, see below.

**3.4.1. Two algorithms for simulating the effective dynamics.** Next we will sketch tow different algorithms for simulations based on the effective dynamics. Subsequently, we will discuss their relation to the equation-free approach and HMM.

For the first algorithm, recall that the coefficients of the effective dynamics (3.10) are defined in (3.11), where the projection operator P given in (3.1) or (3.5) is a conditional expectation on $\Sigma_z$. Therefore for each $z \in \mathbb{R}^m$, the coefficients $\widetilde{b}, \widetilde{a}$ can be calculated by generating a long trajectory of the *constrained* full dynamics which evolves on $\Sigma_z$ with the unique invariant measure $\mu_z$. The realization of this constrained full dynamics has already been discussed , e.g., in [12, 13]; we will denote it by $y(s)$. Based on this, an algorithm for simulating the effective dynamics (3.10) can be sketched as follows:

**Algorithm 1.**

1. Set $t = 0$, $k = 0$ and $z^{(0)} = z \in \mathbb{R}^m$. Choose time step size $\Delta s$, $\Delta t$ and fix the parameters $i_0, M \in \mathbb{N}^+$.
2. At time $t = k\Delta t$, let $z = z^{(k)}$. Generate trajectories $y(s)$ for $(i_0 + M)$ steps with time step size $\Delta s$ and compute coefficients $\widetilde{b}, \widetilde{a}$ by

$$\begin{aligned} \widetilde{b}_l(z) =& \frac{1}{M} \sum_{i=i_0+1}^{i_0+M} \mathcal{L}\xi_l(y^{(i)}) \,, \\ \widetilde{a}_{ll'}(z) =& \frac{1}{M} \sum_{i=i_0+1}^{i_0+M} \Big( \sum_{j,j'=1}^{n} a_{jj'}(y^{(i)}) \frac{\partial \xi_l(y^{(i)})}{\partial x_j} \frac{\partial \xi_{l'}(y^{(i)})}{\partial x_{j'}} \Big) \,, \end{aligned} \tag{3.57}$$

   where $1 \le l, l' \le m$, $y(s)$ is the dynamics on $\Sigma_z$ with invariant measure $\mu_z$ and $y^{(i)} = y(i\Delta s)$.

3. Update $z^{(k+1)}$ due to

$$z_l^{(k+1)} = z_l^{(k)} + \widetilde{b}_l(z)\Delta t + \sum_{i=1}^{m} \sqrt{\frac{2\Delta t}{\beta}} \widetilde{\sigma}_{li}(z)\eta_i \,, \quad 1 \le l \le m \,, \tag{3.58}$$

where $\eta_i$ are independent standard Gaussian variables, $1 \le i \le m$ (higher-order schemes are also possible [37, 46]).

4. Set $k := k + 1$ and go back to step 2, or stop if the terminal time is arrived.

In the above algorithm, conditional expectations in the definition of (3.11) are replaced by the time integration of the constrained full dynamics $y(s)$ on $\Sigma_z$ and only the segment of trajectory $y(s)$ after time $i_0\Delta s$ has been used. Here $\Delta s$ and $\Delta t$ are the time step sizes used for integrating the dynamics $y(s)$ and $z(s)$, respectively. When the probability distribution of $y(s)$ converges quickly to the equilibrium distribution $\mu_z$ on $\Sigma_z$ and the reaction coordinate $\xi$ is chosen appropriately, we can expect that (3.57) will be a good approximation of coefficients $\widetilde{b}, \widetilde{a}$ in (3.11) for moderate $M > 0$, and also a large time step size $\Delta t$ can be used to integrate the effective dynamics $z(s)$. In particular, the above algorithm will accelerate the simulation if we have $\Delta t \gg (M + i_0)\Delta s$. Notice that coefficients $\widetilde{\sigma}$ can be obtained by decomposing the matrix $\widetilde{a}$; the computational cost is negligible compared to the other parts of the algorithm, since typically dimensions $m \ll n$ and $\widetilde{\sigma}$ is only required to integrate dynamics $z(s)$ with a large time step $\Delta t$ (in Step 3).

For the second algorithm, we recall the alternative formulas (3.12) of the coefficients $\widetilde{b}, \widetilde{a}$ of the effective dynamics. Based on these we can calculate the coefficients provided that we approximate the limit $s \to 0+$ by a small but finite time and replace the expectation by an appropriate trajectory ensemble:

**Algorithm 2.**

1. Set $t = 0$, $k = 0$ and $z^{(0)} = z \in \mathbb{R}^m$. Choose time step size $\Delta s$, $\Delta t$. Set $N \in \mathbb{N}^+$ and $\bar{s} = k_0\Delta s$ for some $k_0 \in \mathbb{N}^+$. .
2. At time $t = k\Delta t$, let $z = z^{(k)}$. Generate $N$ trajectories of length $s$ of the unconstrained full dynamics $x(s)$ by discretization of (2.1) and initial distribution $\mu_z$ on $\Sigma_z$ for $k_0$ steps with time step size $\Delta s$ and compute the coefficients $\widetilde{b}, \widetilde{a}$ by

$$\begin{aligned} \widetilde{b}_l(z) =& \frac{1}{N}\sum_{i=1}^{N} \frac{\xi_l(x(\bar{s})) - z_l}{\bar{s}} \,, \\ \widetilde{a}_{ll'}(z) =& \frac{\beta}{2}\Big[\frac{1}{N}\sum_{i=1}^{N} \frac{(\xi_l(x(\bar{s})) - z_l)(\xi_{l'}(x(\bar{s})) - z_{l'})}{\bar{s}} - \widetilde{b}_l\widetilde{b}_{l'}\bar{s}\Big] \,, \end{aligned} \tag{3.59}$$

where $1 \le l, l' \le m$.

3. Update $z^{(k+1)}$ as in Algorithm 1 :

$$z_l^{(k+1)} = z_l^{(k)} + \widetilde{b}_l(z)\Delta t + \sum_{i=1}^{m} \sqrt{\frac{2\Delta t}{\beta}} \widetilde{\sigma}_{li}(z)\eta_i \,, \quad 1 \le l \le m \,, \tag{3.60}$$

where $\eta_i$ are independent standard Gaussian variables, $1 \le i \le m$.

4. Set $k := k + 1$ and go back to step 2, or stop if the terminal time is arrived.

In Algorithm 2, we use trajectories of length $\bar{s} = k_0 \Delta s$ to approximate the limit $s \to 0+$ in (3.12). A correction term was added when computing $\widetilde{a}$ in (3.59) due to this finite time approximation [34, 69, 31, 74]. We expect that both, some small number of steps $k_0$ as well as and a small step size $\Delta s$ will be appropriate and sufficient to provide reasonable approximations. The conditional expectations in (3.12) are replaced by taking the average with $N$ trajectories of the unconstrained full dynamics. The initial state that has to be distributed due $\mu_z$ can be sampled either by using the dynamics $y(s)$ or appropriate MCMC methods. As in the case of Algorithm 1, Algorithm 2 will accelerate the simulation if $\Delta t \gg N k_0 \Delta s$.

**3.4.2. Comparison.** We can readily recognize that Algorithms 1 and 2 fit into the HMM and equation-free approaches. More specifically, if we look at Algorithm 1 in the context of HMM, one supplies the required data/coefficients in the "macro model" (our effective dynamics) by simulating the "micro model" locally (constrained full dynamics). We refer to some generalizations of HMM to the implicit time scale separation case that can be found in [29]. In contrast, Algorithm 2, despite many similarities, utilizes the unconstrained full dynamics (called "legacy code" in the equation-free setting) for computing the coefficient of the effective dynamics "on the fly". We can conclude that, in our setting, both the HMM and the equation-free approach can be viewed as numerical approaches to simulate the same effective dynamics. More discussions on the connections and comparisons of these two approaches can be found in [19, 79]; one obvious advantage of Algorithm 2 is that computing the derivatives of the reaction coordinate $\xi$ can be avoided, while on the other hand bias may be introduced when computing the coefficients using (3.59) due to the approximation of limit $s \to 0+$ by finite time length $\bar{s}$.

**4. Conclusions.** Ergodic diffusion processes are ubiquitous in the real world applications and have received considerable attentions in different disciplines during the past several decades. In this article, we have studied some structural properties of ergodic diffusion processes as well as their model reduction with respect to a low-dimensional reaction coordinate space. We showed that the effective dynamics, which is obtained from model reduction as the low-dimensional approximation of the full dynamics in the reaction coordinate space, inherits some important structual properties like ergodicity and reversibility and that other dynamical properties like inherent time-scales and reaction / transition rates of the full dynamics can be approximated by the effective dynamics. Numerical algorithms for simulating the effective dynamics and their relation to well-known approaches from the literature are discussed as well.

While Langevin dynamics is widely used in modeling chemical and biological systems, the model reduction of this type of dynamics (with degenerate noise) is not covered in the current work and will be studied in the future. However, the short introduction to this case in Appendix D already outlines that many of the results of the present article might be extended to the case of Langevin dynamics.

**Acknowledgement.** This research has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114, and by the Einstein Foundation Berlin through the Einstein Center for Mathematics (ECMath).

**Appendix.** The appendix collects some rather technical aspects that have been taken out of the main text for improving readability.

**A: Proof of Proposition 1.**

*Proof.*

1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be any integrable function. Applying some basic properties of the transportation of measures and the integration by parts formula, we can directly compute

$$\begin{aligned}
&\frac{d}{ds}\int_{\mathbb{R}^n} f\, d\big((\phi_s)_{\#}\mu\big)\\
=&\frac{d}{dt}\int_{\mathbb{R}^n} f\, d\big((\phi_{t+s})_{\#}\mu\big)\Big|_{t=0} = \frac{d}{dt}\int_{\mathbb{R}^n} f(\phi_s(\phi_t))\, d\mu\,\Big|_{t=0}\\
=&\int_{\mathbb{R}^n} \nabla(f\circ\phi_s)(x)\cdot J^{st}(x)\,\rho(x)\,dx = -\int_{\mathbb{R}^n}(f\circ\phi_s)\,\mathrm{div}\big(J^{st}\rho\big)\,dx = 0\,,
\end{aligned}$$

   where equation (2.11) is used in the last step. Therefore we can conclude $(\phi_s)_{\#}\mu \equiv \mu$ and the first assertion follows.

2. For any measurable function $f : \mathbb{R}^n \to \mathbb{R}$, using the invariance $(\phi_s)_{\#}\mu = \mu$ and the change of variables formula, we have

$$\begin{aligned}
&\int_{\mathbb{R}^n} f\rho\,dx = \int_{\mathbb{R}^n} f\,d\mu = \int_{\mathbb{R}^n} f\,d(\phi_s)_{\#}\mu = \int_{\mathbb{R}^n} f(\phi_s)\,d\mu = \int_{\mathbb{R}^n} f(\phi_s)\rho\,dx\,,\\
&\int_{\mathbb{R}^n} f\rho\,dx = \int_{\mathbb{R}^n} f(\phi_s(x))\,\rho(\phi_s(x))\,|\det(\nabla\phi_s)|\,dx\,,
\end{aligned}\tag{4.1}$$

   where $\nabla\phi_s$ denotes the Jacobian matrix of map $\phi_s$. Combining the above two equalities, we can conclude

$$\rho(x) = \rho(\phi_s(x))\,|\det(\nabla\phi_s)|\,,\quad \text{a.s. } x\in\mathbb{R}^n\,.\tag{4.2}$$

   A similar reasoning as in (4.1) gives

$$v(x) = v_s(\phi_s(x))\,|\det(\nabla\phi_s)|\,,\quad \text{a.s. } x\in\mathbb{R}^n\,.\tag{4.3}$$

   Therefore, we can deduce that

$$\begin{aligned}
\mathrm{Ent}(v_s) &= \int_{\mathbb{R}^n} v_s \ln\big(\frac{v_s}{\rho}\big)\,dx = \int_{\mathbb{R}^n} \ln\big(\frac{v_s}{\rho}\big)\,d\big((\phi_s)_{\#}\mu_v\big)\\
&= \int_{\mathbb{R}^n} \ln\Big(\frac{v_s(\phi_s)}{\rho(\phi_s)}\Big)\,d\mu_v = \int_{\mathbb{R}^n} \ln\big(\frac{v}{\rho}\big)\,d\mu_v = \mathrm{Ent}(v)\,,
\end{aligned}\tag{4.4}$$

   which holds for $\forall s \geq 0$.

□

**B: Proof of Proposition 2.**

*Proof.* The conclusion follows by direct calculation. Using (2.9), (2.11), integration by parts formula as well as (2.10), we obtain

$$\begin{aligned}
&\frac{d}{dt}\mathrm{Ent}(u_t)\\
=&\int_{\mathbb{R}^n}\frac{\partial u_t}{\partial t}\Big[\ln\big(\frac{u_t}{\rho}\big)+1\Big]\,dx=\int_{\mathbb{R}^n}J(u_t)\cdot\nabla\big(\frac{u_t}{\rho}\big)\,\rho\,dx\\
=&-\frac{1}{\beta}\sum_{i,j=1}^n\int_{\mathbb{R}^n}a_{ij}\frac{\partial}{\partial x_j}\big(\ln\frac{u_t}{\rho}\big)\frac{\partial}{\partial x_i}\big(\ln\frac{u_t}{\rho}\big)\,u_t\,dx\\
=&-\frac{1}{\beta}\int_{\mathbb{R}^n}\Big(\nabla\ln\big(\frac{u_t}{\rho}\big)\Big)^T a\Big(\nabla\ln\big(\frac{u_t}{\rho}\big)\Big)\,u_t\,dx\\
=&\frac{4}{\beta}\Big\langle\mathcal{L}_s\Big(\frac{u_t}{\rho}\Big)^{\frac12},\Big(\frac{u_t}{\rho}\Big)^{\frac12}\Big\rangle_\mu\le 0\,,
\end{aligned}$$

which implies that the relative entropy is non-increasing. □

**C: Proof of Proposition 3.**

*Proof.* Let us first assume that $f\in D(\mathcal{L})$ and $\widetilde{h}\in\widetilde{H}$ is a $\mathcal{C}^1$ function. The partial derivatives of functions in $\widetilde{H}$ will be denoted as $\frac{\partial}{\partial z_k}$, $1\le k\le m$. Using relation (2.13) for the generator $\mathcal{L}$, the co-area formula (3.4) and the chain rule, we have

$$\begin{aligned}
&\int_{\mathbb{R}^m}(\mathrm{P}\mathcal{L}f)(z)\,\widetilde{h}(z)\,Q(z)\,dz\\
=&\int_{\mathbb{R}^n}\mathcal{L}f(x)\,\widetilde{h}(\xi(x))\,\rho(x)\,dx\\
=&\int_{\mathbb{R}^n}(J^{st}\cdot\nabla f)(x)\,\widetilde{h}\big(\xi(x)\big)\rho(x)\,dx-\frac{1}{\beta}\int_{\mathbb{R}^n}\sum_{k=1}^m\sum_{i,j=1}^n\Big(a_{ij}\frac{\partial f}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)(x)\frac{\partial\widetilde{h}}{\partial z_k}\big(\xi(x)\big)\,\rho(x)\,dx\\
=&\int_{\mathbb{R}^m}\mathrm{P}(J^{st}\cdot\nabla f)(z)\,\widetilde{h}(z)\,Q(z)\,dz+\frac{1}{\beta}\int_{\mathbb{R}^m}\sum_{k=1}^m\frac{\partial}{\partial z_k}\Big[Q(z)\mathrm{P}\Big(\sum_{i,j=1}^n a_{ij}\frac{\partial f}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)(z)\Big]\,\widetilde{h}(z)\,dz\,.
\end{aligned}$$

Therefore we have obtained the identity

$$\mathrm{P}(\mathcal{L}f)=\mathrm{P}(J^{st}\cdot\nabla f)+\frac{1}{\beta Q}\sum_{k=1}^m\frac{\partial}{\partial z_k}\Big[Q\,\mathrm{P}\Big(\sum_{i,j=1}^n a_{ij}\frac{\partial f}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)\Big]\,. \tag{4.5}$$

Especially, taking $f=\xi_l$, $1\le l\le m$, we have

$$\mathrm{P}(\mathcal{L}\xi_l)=\mathrm{P}(J^{st}\cdot\nabla\xi_l)+\frac{1}{\beta Q}\sum_{k=1}^m\frac{\partial}{\partial z_k}\Big[Q\,\mathrm{P}\Big(\sum_{i,j=1}^n a_{ij}\frac{\partial\xi_l}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)\Big]\,. \tag{4.6}$$

Now consider function $f=\widetilde{f}\circ\xi\in H_0$ where $\widetilde{f}\in D(\widetilde{\mathcal{L}})$. Using (4.5), (4.6) and the chain rule, it follows that

$$\begin{aligned}
\mathrm{P}(\mathcal{L}f)=&\sum_{l=1}^m\mathrm{P}(J^{st}\cdot\nabla\xi_l)\frac{\partial\widetilde{f}}{\partial z_l}+\frac{1}{\beta Q}\sum_{k,l=1}^m\frac{\partial}{\partial z_k}\Big[Q\,\mathrm{P}\Big(\sum_{i,j=1}^n a_{ij}\frac{\partial\xi_l}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)\frac{\partial\widetilde{f}}{\partial z_l}\Big]\\
=&\sum_{l=1}^m\mathrm{P}(\mathcal{L}\xi_l)\frac{\partial\widetilde{f}}{\partial z_l}+\frac{1}{\beta}\sum_{k,l=1}^m\mathrm{P}\Big(\sum_{i,j=1}^n a_{ij}\frac{\partial\xi_l}{\partial x_i}\frac{\partial\xi_k}{\partial x_j}\Big)\frac{\partial^2\widetilde{f}}{\partial z_l\partial z_k}\\
=&\widetilde{\mathcal{L}}\widetilde{f}\,.
\end{aligned}\tag{4.7}$$

□

**D: Langevin dynamics.** Consider the Langevin dynamics

$$\begin{aligned} \dot r_i &= v_i \\ m_i \dot v_i &= -\frac{\partial V}{\partial r_i} - \gamma v_i + \sqrt{2\beta^{-1}\gamma}\eta_i(t) \end{aligned} \qquad 1 \le i \le n\,, \tag{4.8}$$

where $r = (r_1, r_2, \cdots, r_n), v = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^n$. The infinitesimal generator is

$$\mathcal{L} = \sum_{i=1}^n \Big( v_i \frac{\partial}{\partial r_i} - \frac{1}{m_i}\frac{\partial V}{\partial r_i}\frac{\partial}{\partial v_i} - \frac{\gamma v_i}{m_i}\frac{\partial}{\partial v_i} + \frac{\gamma}{\beta m_i^2}\frac{\partial^2}{\partial v_i^2}\Big)\,. \tag{4.9}$$

The system has an invariant measure

$$\mu(dr, dv) = \rho(r,v)drdv = \frac{1}{Z}e^{-\beta H(r,v)}drdv, \quad H = V(r) + \frac{1}{2}\sum_{i=1}^n m_i v_i^2\,. \tag{4.10}$$

However it is not reversible, and at equilibrium there is a nonzero flux $J^{st}$ with

$$J_i^{st} = \begin{cases} v_i, & 1 \le i \le n\,, \\ -\frac{1}{m_{i'}}\frac{\partial V}{\partial r_{i'}} & n+1 \le i = i'+n \le 2n\,. \end{cases} \tag{4.11}$$

Now suppose a reaction coordinate $z = \xi(r,v) \in \mathbb{R}^m$ is given. Applying the previous derivations, we have for $f = \widetilde{f} \circ \xi$,

$$\mathrm{P}\mathcal{L}f = \sum_{k=1}^m \mathrm{P}(\mathcal{L}\xi_k)\frac{\partial f}{\partial z_k} + \frac{\gamma}{\beta}\sum_{j,k=1}^m \mathrm{P}\big(\sum_{i=1}^n \frac{1}{m_i^2}\frac{\partial \xi_k}{\partial v_i}\frac{\partial \xi_j}{\partial v_i}\big)\frac{\partial^2 f}{\partial z_j \partial z_k} = \widetilde{\mathcal{L}}f\,. \tag{4.12}$$

The corresponding SDE in $\mathbb{R}^m$ is

$$dz(t) = \widetilde{b}(z(t))dt + \sqrt{2\gamma\beta^{-1}}\,\widetilde{\sigma}(z(t))dw(t) \tag{4.13}$$

where for $1 \le j,k \le m$,

$$\widetilde{b}_j(z) = \mathrm{P}(\mathcal{L}\xi_j), \quad (\widetilde{a})_{jk} = \widetilde{\sigma}\,\widetilde{\sigma}^T(z)_{jk} = \mathrm{P}\Big(\sum_{i=1}^n \frac{1}{m_i^2}\frac{\partial \xi_k}{\partial v_i}\frac{\partial \xi_j}{\partial v_i}\Big)\,. \tag{4.14}$$

When $z = (\xi(r), \xi'(v)) \in \mathbb{R}^m \times \mathbb{R}^m$, we have

$$\begin{aligned} dz_1(t) =& \mathrm{P}\Big(\sum_{i=1}^n v_i \frac{\partial \xi}{\partial r_i}\Big)dt \\ dz_2(t) =& \mathrm{P}\Big[\sum_{i=1}^n \Big( -\frac{1}{m_i}\frac{\partial V}{\partial r_i}\frac{\partial \xi'}{\partial v_i} - \frac{\gamma v_i}{m_i}\frac{\partial \xi'}{\partial v_i} + \frac{\gamma}{\beta m_i^2}\frac{\partial^2 \xi'}{\partial v_i^2}\Big)\Big]dt + \sqrt{2\beta^{-1}\gamma}\widetilde{\sigma}(z)dw(t) \end{aligned}$$

where for $1 \le j,k \le m$,

$$(\widetilde{\sigma}\widetilde{\sigma}^T)_{jk} = \mathrm{P}\Big(\sum_{i=1}^n \frac{1}{m_i^2}\frac{\partial \xi_j'}{\partial v_i}\frac{\partial \xi_k'}{\partial v_i}\Big)\,. \tag{4.15}$$

REFERENCES


[1] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden, *The heterogeneous multiscale method*, Acta Numer., 21 (2012), pp. 1–87.

[2] C. Abrams and G. Bussi, *Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration*, Entropy, 16 (2014), pp. 163–199.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics, Birkhäuser, 2005.

[4] I. Babuška and J. E. Osborn, *Finite element-Galerkin approximation of the eigenvalues and eigenvectors of selfadjoint problems*, Math. Comput., 52 (1989), pp. 275–297.

[5] ———, *Eigenvalue problems*, in Finite Element Methods (Part 1), vol. 2 of Handbook of Numerical Analysis, Elsevier, 1991, pp. 641 – 787.

[6] H. B. G. Casimir, *On Onsager's principle of microscopic reversibility*, Rev. Mod. Phys., 17 (1945), pp. 343–350.

[7] A. J. Chorin, O. H. Hald, and R. Kupferman, *Optimal prediction and the Mori-Zwanzig representation of irreversible processes*, Proc. Natl. Acad. Sci. U.S.A., 97 (2000), pp. 2968–2973.

[8] ———, *Optimal prediction with memory*, Physica D, 166 (2002), pp. 239–257.

[9] A. J. Chorin, R. Kupferman, and D. Levy, *Optimal prediction for Hamiltonian partial differential equations*, J. Comput. Phys., 162 (2000), pp. 267 – 297.

[10] C. D. Christ, A. E. Mark, and W. F. van Gunsteren, *Basic ingredients of free energy calculations: A review*, J. Comput. Chem., 31 (2010), pp. 1569–1582.

[11] P. Ciarlet, *The Finite Element Method for Elliptic Problems*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 2002.

[12] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, *Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics*, ChemPhysChem, 6 (2005), pp. 1809–1814.

[13] G. Ciccotti, T. Lelièvre, and E. Vanden-Eijnden, *Projection of diffusions on submanifolds: Application to mean force computation*, Commun. Pur. Appl. Math., 61 (2008), pp. 371–408.

[14] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, 1996.

[15] E. Darve, J. Solomon, and A. Kia, *Computing generalized Langevin equations and generalized Fokker-Planck equations*, Proc. Natl. Acad. Sci. U.S.A., 106 (2009), pp. 10884–10889.

[16] E. B. Davies, *Spectral Theory and Differential Operators*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1995.

[17] C. Dellago and P. G. Bolhuis, *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, ch. Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events, pp. 167–233.

[18] N. Djurdjevac, M. Sarich, and Ch. Schütte, *Estimating the eigenvalue error of Markov state models*, Multiscale Model. Simul., 10 (2012), pp. 61 – 81.

[19] W. E, *Principles of Multiscale Modeling*, Cambridge University Press, 2011.

[20] W. E and B. Engquist, *The heterognous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.

[21] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden, *Heterogeneous multiscale methods: A review*, Commun. Comput. Phys., 2 (2007), pp. 367–450.

[22] W. E, W. Ren, and E. Vanden-Eijnden, *String method for the study of rare events*, Phys. Rev. B, 66 (2002), p. 052301.

[23] ———, *Finite temperature string method for the study of rare events*, J. Phys. Chem. B, 109 (2005), pp. 6688–6693.

[24] W. E and E. Vanden-Eijnden, *Towards a theory of transition paths*, J. Stat. Phys., 123 (2006), pp. 503–523.

[25] ———, *Transition-path theory and path-finding algorithms for the study of rare events*, Annu. Rev. Phys. Chem., 61 (2010), pp. 391–420.

[26] J. P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, Rev. Mod. Phys., 57 (1985), pp. 617–656.

[27] R. Erban, I. G. Kevrekidis, D. Adalsteinsson, and T. C. Elston, *Gene regulatory networks: A coarse-grained, equation-free approach to multiscale computation*, J. Chem. Phys., 124 (2006).

[28] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, Taylor & Francis, 1991.

[29] I. Fatkullin and E. Vanden-Eijnden, *A computational strategy for multiscale systems with applications to lorenz 96 model*, J. Comput. Phys., 200 (2004), pp. 605 – 638.

[30] H. Feng and J. Wang, *Potential and flux decomposition for dynamical systems and non-equilibrium thermodynamics: Curvature, field, and generalized fluctuation-dissipation theorem*, J. Chem. Phys., 135 (2011).

[31] R. Friedrich, Ch. Renner, M. Siefert, and J. Peinke, *Comment on "indispensable finite time corrections for Fokker-Planck equations from time series data"*, Phys. Rev. Lett., 89 (2002), p. 149401.

[32] D. Givon, R. Kupferman, and A. M. Stuart, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity, 17 (2004), pp. R55–R127.

[33] H. Grabert, *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*, Springer Tracts in Modern Physics, Springer-Verlag, 1982.

[34] J. Gradišek, S. Siegert, R. Friedrich, and I. Grabec, *Analysis of time series from stochastic processes*, Phys. Rev. E, 62 (2000), pp. 3146–3155.

[35] I. Gyöngy, *Mimicking the one-dimensional marginal distributions of processes having an Ito differential*, Probab. Th. Rel. Fields, 71 (1986), pp. 501–516.

[36] M. Haataja, D. J. Srolovitz, and I. G. Kevrekidis, *Apparent hysteresis in a driven system with self-organized drag*, Phys. Rev. Lett., 92 (2004), p. 160603.

[37] D. J. Higham., *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Rev., 43 (2001), pp. 525–546.

[38] C. Hijon, P. Espanol, E. Vanden-Eijnden, and R. Delgado-Buscalioni, *Mori-Zwanzig formalism as a practical computational tool*, Farad. Discuss., 144 (2010), pp. 301–322.

[39] E. P. Hsu, *Stochastic analysis on manifolds*, Graduate Studies in Mathematics, American Mathematical Society, 2002.

[40] G. Hummer and I. G. Kevrekidis, *Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations*, J. Chem. Phys., 118 (2003), pp. 10762–10773.

[41] S. Izvekov, *Microscopic derivation of particle-based coarse-grained dynamics*, J. Chem. Phys., 138 (2013), p. 134106.

[42] I. G. Kevrekidis, C. W. Gear, and G. Hummer, *Equation-free: The computer-aided analysis of complex multiscale systems*, AIChE J., 50 (2004), pp. 1346–1355.

[43] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G Kevrekidid, O. Runborg, and C. Theodoropoulos, *Equation-free, coarse-grained multiscale computation: Enabling mocroscopic simulators to perform system-level analysis*, Commun. Math. Sci., 1 (2003), pp. 715–762.

[44] I. G. Kevrekidis and G. Samaey, *Equation-free multiscale computation: Algorithms and applications*, Annu. Rev. Phys. Chem., 60 (2009), pp. 321–344.

[45] B. Khouider, A. J Majda, and S. N Stechmann, *Climate science in the tropics: waves, vortices and PDEs*, Nonlinearity, 26 (2013), pp. R1–R68.

[46] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics, Springer, Berlin, New York, 1999.

[47] D. I. Kopelevich, A. Z. Panagiotopoulos, and I. G. Kevrekidis, *Coarse-grained kinetic computations for rare events: Application to micelle formation*, J. Chem. Phys., 122 (2005).

[48] S. G. Krantz and H. R. Parks, *Geometric Integration Theory*, Cornerstones, Birkhäuser Boston, 2008.

[49] C. R. Laing, T. Frewen, and I. G. Kevrekidis, *Reduced models for binocular rivalry*, J. Comput. Neurosci., 28 (2010), pp. 459–476.

[50] J. L. Lebowitz and O. Penrose, *Modern ergodic theory*, Phys. Today, 26 (1973), pp. 23–29.

[51] F. Legoll and T. Lelièvre, *Effective dynamics using conditional expectations*, Nonlinearity, 23 (2010), pp. 2131–2163.

[52] F. Legoll, T. Lelièvre, and G. Samaey, *A micro-macro parareal algorithm: application to singularly perturbed ordinary differential equations*, SIAM J. Sci. Comput., 35 (2013), pp. A1951–A1986.

[53] T. Lelièvre, G. Stoltz, and M. Rousset, *Free Energy Computations: A Mathematical Perspective*, Imperial College Press, 2010.

[54] T. Li, X. Li, and X. Zhou, *Finding transition pathways on manifolds*, Multiscale Model. Simul., 14 (2016), pp. 173–206.

[55] Z. Li, X. Bian, B. Caswell, and G. E. Karniadakis, *Construction of dissipative particle dynamics models for complex fluids via the Mori-Zwanzig formulation*, Soft Matter, 10 (2014), pp. 8659–8672.

[56] J. Lu and E. Vanden-Eijnden, *Exact dynamical coarse-graining without time-scale separation*, J. Chem.

Phys., 141 (2014), p. 044109.
[57] A. J. Majda, C. Franzke, and B. Khouider, *An applied mathematics perspective on stochastic modelling for climate*, Philos. Trans. A: Math. Phys. Eng. Sci., 366 (2008), pp. 2427–2453.
[58] C. C. Moore, *Ergodic theorem, ergodic theory, and statistical mechanics*, Proc. Natl. Acad. Sci. U.S.A., 112 (2015), pp. 1907–1911.
[59] H. Mori, *Transport, collective motion, and Brownian motion*, Prog. Theor. Phys., 33 (1965), pp. 423–455.
[60] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 5th ed., 2000.
[61] L. Onsager, *Reciprocal relations in irreversible processes. I.*, Phys. Rev., 37 (1931), pp. 405–426.
[62] A. Papavasiliou, G. A. Pavliotis, and A. M. Stuart, *Maximum likelihood drift estimation for multiscale diffusions*, Stoch. Proc. Appl., 119 (2009), pp. 3173 – 3210.
[63] G. A. Pavliotis and A. M. Stuart, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.
[64] ———, *Multiscale Methods: Averaging and Homogenization*, Texts in Applied Mathematics, Springer New York, 2008.
[65] K. E. Petersen, *Ergodic Theory*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1989.
[66] P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics, Springer New York, 2006.
[67] J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, Ch. Schütte, and F. Noé, *Markov models of molecular kinetics: Generation and validation*, J. Chem. Phys., 134 (2011), p. 174105.
[68] H. Qian, *A decomposition of irreversible diffusion processes without detailed balance*, J. Math. Phys., 54 (2013), p. 053302.
[69] M. Ragwitz and H. Kantz, *Indispensable finite time corrections for Fokker-Planck equations from time series data*, Phys. Rev. Lett., 87 (2001), p. 254501.
[70] H. Risken, *The Fokker-Planck equation : Methods of solution and applications*, vol. 18 of Springer series in synergetics, Springer-Verlag, Berlin, New York, 1996.
[71] M. Sarich, F. Noé, and Ch. Schütte, *On the approximation quality of Markov state models*, Multiscale Model. Simul., 8 (2010), pp. 1154 – 1177.
[72] M. Sarich and Ch. Schütte, *Approximating selected non-dominant timescales by Markov state models*, Comm. Math. Sci., 10 (2012), pp. 1001 – 1013.
[73] Ch. Schütte, W. Huisinga, and P. Deuflhard, *Transfer operator approach to conformational dynamics in biomolecular systems*, in Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems, B. Fiedler, ed., Springer Berlin Heidelberg, 2001, pp. 191–223.
[74] S. Siegert, R. Friedrich, and J. Peinke, *Analysis of data sets of stochastic systems*, Phys. Lett. A, 243 (1998), pp. 275 – 280.
[75] S. Sriraman, I. G. Kevrekidis, and G. Hummer, *Coarse nonlinear dynamics and metastability of filling-emptying transitions: Water in carbon nanotubes*, Phys. Rev. Lett., 95 (2005), p. 130603.
[76] G. Teschl, *Mathematical methods in quantum mechanics : With applications to Schrödinger operators*, vol. 99 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2009.
[77] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Commun. Math. Sci., 1 (2003), pp. 385–391.
[78] ———, *Transition path theory*, in Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1, M. Ferrario, G. Ciccotti, and K. Binder, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 453–493.
[79] ———, *On HMM-like integrators and projective integration methods for systems with multiple time scales*, Commun. Math. Sci., 5 (2007), pp. 495–505.
[80] ———, *Some recent techniques for free energy calculations*, J. Comput. Chem., 30 (2009), pp. 1737–1747.
[81] E. Vanden-Eijnden and M. Venturoli, *Revisiting the finite temperature string method for the calculation of reaction tubes and free energies*, J. Chem. Phys., 130 (2009), p. 194103.
[82] C. Villani, *Topics in optimal transportation*, Graduate studies in mathematics, American Mathematical Society, 2003.
[83] ———, *Optimal transport : old and new*, Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 2009.
[84] Ch. Schütte W. Huisinga, S. Meyn, *Phase transitions and metastability in Markovian and molecular*

*systems*, Ann. Appl. Probab., 14 (2004), pp. 419–458.

[85] F. Wang, *Functional inequalities for empty essential spectrum*, J. Funct. Anal., 170 (2000), pp. 219 – 245.

[86] J. Wang, L. Xu, and E. Wang, *Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations*, Proc. Natl. Acad. Sci. U.S.A., 105 (2008), pp. 12271–12276.

[87] A. V. Weigel, B. Simon, M. M. Tamkun, and D. Krapf, *Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking*, Proc. Natl. Acad. Sci. U.S.A., 108 (2011), pp. 6438–6443.

[88] R. Zwanzig, *Nonlinear generalized Langevin equations*, J. Stat. Phys., 9 (1973), pp. 215–220.