NONINTRUSIVE AND STRUCTURE PRESERVING MULTISCALE INTEGRATION OF STIFF ODEs, SDEs, AND HAMILTONIAN SYSTEMS WITH HIDDEN SLOW DYNAMICS VIA FLOW AVERAGING[*]

MOLEI TAO[†], HOUMAN OWHADI[‡], AND JERROLD E. MARSDEN[†]

Abstract. We introduce a new class of integrators for stiff ODEs as well as SDEs. Examples of subclasses of systems that we treat are ODEs and SDEs that are sums of two terms, one of which has large coefficients. These integrators are as follows: (i) *Multiscale*: They are based on flow averaging and thus do not fully resolve the fast variables and have a computational cost determined by slow variables. (ii) *Versatile*: The method is based on averaging the flows of the given dynamical system (which may have hidden slow and fast processes) instead of averaging the instantaneous drift of assumed separated slow and fast processes. This bypasses the need for identifying explicitly (or numerically) the slow or fast variables. (iii) *Nonintrusive*: A pre-existing numerical scheme resolving the microscopic time scale can be used as a black box and easily turned into one of the integrators in this paper by turning the large coefficients on over a microscopic time scale and off during a mesoscopic time scale. (iv) *Convergent over two scales*: They converge strongly over slow processes and in the sense of measures over fast ones. We introduce the related notion of two-scale flow convergence and analyze the convergence of these integrators under the induced topology. (v) *Structure preserving*: They inherit the structure preserving properties of the legacy integrators from which they are derived. Therefore, for stiff Hamiltonian systems (possibly on manifolds), they can be made to be symplectic, time-reversible, and symmetry preserving (symmetries are group actions that leave the system invariant) in all variables. They are explicit and applicable to arbitrary stiff potentials (that need not be quadratic). Their application to the Fermi–Pasta–Ulam problems shows accuracy and stability over four orders of magnitude of time scales. For stiff Langevin equations, they are symmetry preserving, time-reversible, Boltzmann–Gibbs-reversible, quasi-symplectic on all variables, and conformally symplectic with isotropic friction.

1. Overview of the integrator on ODEs.
Consider the following ODE on $\mathbb{R}^d$:

$$(1.1) \qquad \dot{u}^\epsilon = G(u^\epsilon) + \frac{1}{\epsilon} F(u^\epsilon).$$

In subsections 1.9, 2.1, 3.1, 3.5, and 4.1, we will consider more general ODEs, stiff deterministic Hamiltonian systems (2.1), SDEs ((3.1) and (3.15)), and Langevin equations ((4.1) and (4.2)); however, for the sake of clarity, we will start the description of our method with (1.1).

CONDITION 1.1. *Assume that there exists a diffeomorphism* $\eta := (\eta^x, \eta^y)$, *from* $\mathbb{R}^d$ *onto* $\mathbb{R}^{d-p} \times \mathbb{R}^p$ *(with uniformly bounded* C^1, C^2 *derivatives), separating slow and*

[†]Applied & Computational Mathematics, and Control & Dynamical Systems, California Institute of Technology, MC 217-50, Pasadena, CA 91125 (mtao@caltech.edu, marsden@cds.caltech.edu).

[‡]Corresponding author. Applied & Computational Mathematics, and Control & Dynamical Systems, California Institute of Technology, MC 217-50, Pasadena, CA 91125 (owhadi@caltech.edu).

fast variables, i.e., such that (for all $\epsilon > 0$) the process $(x_t^\epsilon, y_t^\epsilon) = (\eta^x(u_t^\epsilon), \eta^y(u_t^\epsilon))$ satisfies an ODE system of the form

$$(1.2) \qquad \begin{cases} \dot{x}^\epsilon = g(x^\epsilon, y^\epsilon), & x_0^\epsilon = x_0, \\ \dot{y}^\epsilon = \frac{1}{\epsilon} f(x^\epsilon, y^\epsilon), & y_0^\epsilon = y_0. \end{cases}$$

CONDITION 1.2. *Assume that the fast variables in* (1.2) *are locally ergodic with respect to a family of measures μ drifted by slow variables. More precisely, we assume that there exists a family of probability measures $\mu(x, dy)$ on $\mathbb{R}^p$ indexed by $x \in \mathbb{R}^{d-p}$ and a positive function $T \mapsto E(T)$ such that $\lim_{T \to \infty} E(T) = 0$ and such that for all x_0, y_0, T, and ϕ uniformly bounded and Lipschitz the solution to*

$$(1.3) \qquad \dot{Y}_t = f(x_0, Y_t), \quad Y_0 = y_0$$

satisfies

$$(1.4) \qquad \left| \frac{1}{T} \int_0^T \phi(Y_s) ds - \int_{\mathbb{R}^p} \phi(y) \mu(x_0, dy) \right| \le \chi\big(\|(x_0, y_0)\|\big) E(T)(\|\phi\|_{L^\infty} + \|\nabla \phi\|_{L^\infty}),$$

where $r \mapsto \chi(r)$ is bounded on compact sets.

Under Conditions 1.1 and 1.2, it is known (we refer the reader, for instance, to [99] or to Theorem 14 of section 3 of Chapter II of [108] or to [92]) that x^ϵ converges towards x_t defined as the solution to the ODE

$$(1.5) \qquad \dot{x} = \int g(x, y) \mu(x, dy), \qquad x|_{t=0} = x_0,$$

where $\mu(x, dy)$ is the ergodic measure associated with the solution to the ODE

$$(1.6) \qquad \dot{y} = f(x, y).$$

It follows that the slow behavior of solutions of (1.1) can be simulated over coarse time steps by first identifying the slow process x^ϵ and then using numerical approximations of solutions of (1.2) to approximate x^ϵ. Two classes of integrators have been founded on this observation: the equation-free method [67, 68] and the heterogeneous multi-scale method (HMM) [37, 41, 36, 5]. One shared characteristic of the original form of those integrators is, after identification of the slow variables, to use a microsolver to approximate the effective drift in (1.5) by averaging the instantaneous drift g with respect to numerical solutions of (1.6) over a time span larger than the mixing time of the solution to (1.6).

1.1. FLAVORS. In this paper, we propose a new method based on the averaging of the instantaneous flow of the ODE (1.1) with hidden slow and fast variables instead of the instantaneous drift of x^ϵ in ODE (1.2) with separated slow and fast variables. We call the resulting class of numerical integrators *FLow AVeraging integratORS (FLAVORS)*. Since FLAVORS are directly applied to (1.1), hidden slow variables do not need to be identified, either explicitly or numerically. Furthermore, FLAVORS can be implemented using an arbitrary legacy integrator $\Phi_h^{\frac{1}{\epsilon}}$ for (1.1) in which the parameter $\frac{1}{\epsilon}$ can be controlled (Figure 1.1). More precisely, assume that there exists a constant $h_0 > 0$ such that Φ_h^α satisfies for all $h \le h_0 \min(\frac{1}{\alpha}, 1)$ and $u \in \mathbb{R}^d$

$$(1.7) \qquad \left| \Phi_h^\alpha(u) - u - hG(u) - \alpha h F(u) \right| \le Ch^2(1 + \alpha)^2;$$
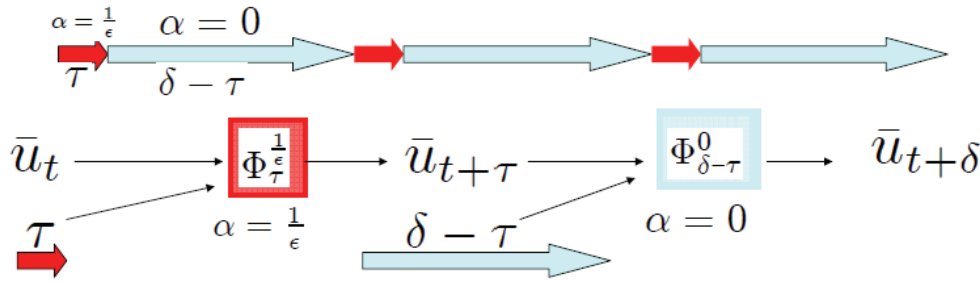
FIG. 1.1. *A pre-existing numerical scheme resolving the microscopic time scale can be used as a black box and turned into a FLAVOR by simply turning stiff parameters on and off over a microscopic time scale τ (on) and a mesoscopic time scale δ (off). The bottom line of the approach is to (repeatedly) compose an accurate, short time integration of the complete set of equations with an accurate, intermediate-time integration of the nonstiff part of the system. While the integration over short time intervals is accurate (in a strong sense), this is extended to intermediate time integration (in the sense of measures) using the interplay between the short time integration and the mesoscopic integration. The computational cost remains bounded independently from the stiff parameter $1/\epsilon$ because of the following: (i) The whole system is integrated only over extremely short ($\tau \ll \epsilon$) time intervals during every intermediate (δ) time intervals. (ii) The intermediate time step δ (that of the nonstiff part of the system) is limited not by the fast time scales (ϵ) but by the slow ones ($\mathcal{O}(1)$).*

then FLAVOR can be defined as the algorithm simulating the process

$$(1.8) \qquad \bar{u}_t = \left(\Phi^0_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_\tau \right)^k (u_0) \quad \text{for} \quad k\delta \le t < (k+1)\delta,$$

where τ is a fine time step resolving the fast time scale ($\tau \ll \epsilon$) and δ is a mesoscopic time step independent of the fast time scale satisfying $\tau \ll \epsilon \ll \delta \ll 1$ and

$$(1.9) \qquad \left(\frac{\tau}{\epsilon} \right)^2 \ll \delta \ll \frac{\tau}{\epsilon}.$$

In our numerical experiments, we have used the "rule of thumb" $\delta \sim \gamma \frac{\tau}{\epsilon}$, where γ is a small parameter (0.1, for instance).

By switching stiff parameters, FLAVOR approximates the flow of (1.1) over a coarse time step h (resolving the slow time scale) by the flow

$$(1.10) \qquad \Phi_h := \left(\Phi^0_{\frac{h}{M}-\tau} \circ \Phi^{\frac{1}{\epsilon}}_\tau \right)^M,$$

where M is a positive integer corresponding to the number of "samples" used to average the flow (δ has to be identified with $\frac{h}{M}$). We refer the reader to subsection 1.4 for the distinction between macro- and mesosteps and for the rationale and mechanism behind FLAVORS and the limits (1.9).

Since FLAVORS are obtained by flow composition, we will show in sections 2 and 4 that they inherit the structure preserving properties (for instance, symplecticity and symmetries under a group action) of the legacy integrator for Hamiltonian systems and Langevin equations.

Under conditions (1.9) on τ and δ, we show that (1.8) is strongly accurate with respect to (hidden) slow variables and weakly (in the sense of measures) accurate with respect to (hidden) fast variables. Motivated by this observation, we introduce the related notion of *two-scale flow convergence* in analogy with homogenization theory for elliptic PDEs [90, 3] and call it *F*-convergence for short. *F*-convergence is close in spirit to the Young measure approach to computing slowly advancing fast oscillations introduced in [10, 9].

1.2. Two-scale flow convergence. Let $(\xi_t^\epsilon)_{t\in\mathbb{R}^+}$ be a sequence of processes on $\mathbb{R}^d$ (functions from $\mathbb{R}^+$ to $\mathbb{R}^d$) indexed by $\epsilon > 0$. Let $(X_t)_{t\in\mathbb{R}^+}$ be a process on $\mathbb{R}^{d-p}$ ($p \geq 0$). Let $x \mapsto \nu(x, dz)$ be a function from $\mathbb{R}^{d-p}$ into the space of probability measures on $\mathbb{R}^d$.

DEFINITION 1.1. *We say that the process ξ_t^ϵ F-converges to $\nu(X_t, dz)$ as $\epsilon \downarrow 0$ and write $\xi_t^\epsilon \xrightarrow[\epsilon\to 0]{F} \nu(X_t, dz)$ if and only if for all functions φ bounded and uniformly Lipschitz continuous on $\mathbb{R}^d$, and for all $t > 0$,*

$$(1.11) \qquad \lim_{h\to 0}\lim_{\epsilon\to 0}\frac{1}{h}\int_t^{t+h}\varphi(\xi_s^\epsilon)\,ds = \int_{\mathbb{R}^d}\varphi(z)\nu(X_t, dz).$$

1.3. Asymptotic convergence result. Our convergence theorem requires that u_t^ϵ and $\bar{u}_t$ do not blow up as $\epsilon \downarrow 0$; more precisely, we will assume that the following condition is satisfied.

CONDITION 1.3.
1. *F and G are Lipschitz continuous.*
2. *For all u_0, $T > 0$, the trajectories $(u_t^\epsilon)_{0\leq t\leq T}$ are uniformly bounded in ϵ.*
3. *For all u_0, $T > 0$, the trajectories $(\bar{u}_t^\epsilon)_{0\leq t\leq T}$ are uniformly bounded in ϵ, $0 < \delta \leq h_0$, and $\tau \leq \min(\tau_0\epsilon, \delta)$.*

For π, an arbitrary measure on $\mathbb{R}^d$, we define $\eta^{-1} * \pi$ to be the push forward of the measure π by η^{-1}.

THEOREM 1.2. *Let u_t^ϵ be the solution to (1.1) and $\bar{u}_t$ be defined by (1.8). Assume that (1.7) and Conditions 1.1, 1.2, and 1.3 are satisfied; then the following hold:*

- *u_t^ϵ F-converges to $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ as $\epsilon \downarrow 0$, where X_t is the solution to*

$$(1.12) \qquad \dot{X}_t = \int g(X_t, y)\,\mu(X_t, dy), \qquad X_0 = x_0.$$

- *$\bar{u}_t$ F-converges to $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ for $\epsilon \leq \delta/(-C\ln\delta)$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\epsilon}{\tau}\delta \downarrow 0$, and $(\frac{\tau}{\epsilon})^2\frac{1}{\delta} \downarrow 0$.*

REMARK 1.1. *The F-convergence of u_t^ϵ to $\eta^{-1} * \big(\delta_{X_t} \otimes \mu(X_t, dy)\big)$ can be restated as*

$$(1.13) \qquad \lim_{h\to 0}\lim_{\epsilon\to 0}\frac{1}{h}\int_t^{t+h}\varphi(u_s^\epsilon)\,ds = \int_{\mathbb{R}^p}\varphi(\eta^{-1}(X_t, y))\mu(X_t, dy)$$

for all functions φ bounded and uniformly Lipschitz continuous on $\mathbb{R}^d$ and for all $t > 0$.

REMARK 1.2. *Observe that g comes from (1.5). It is not explicitly known and does not need to be explicitly known for the implementation of the proposed method.*

REMARK 1.3. *The limits on ϵ, τ, and δ are in essence stating that FLAVOR is accurate provided that $\tau \ll \epsilon$ (τ resolves the stiffness of (1.1)) and (1.9) is satisfied.*

REMARK 1.4. *Throughout this paper, C will refer to an appropriately large enough constant independent from ϵ, δ, τ. To simplify the presentation of our results, we use the same letter C for expressions such as $2Ce^C$ instead of writing it as a new constant C_1 independent from ϵ, δ, τ.*

1.4. Rationale and mechanism behind FLAVORS. We will now explain the rationale and mechanism behind FLAVORS. We refer the reader to subsection A.1 of the appendix for the detailed proof of Theorem 1.2. Let us start by considering

the case where η is the identity diffeomorphism. Let $\varphi^{\frac{1}{\epsilon}}$ be the flow of (1.2). Observe that φ^0 (obtained from $\varphi^{\frac{1}{\epsilon}}$ by setting the parameter $\frac{1}{\epsilon}$ to zero) is the flow of (1.2) with y^ϵ frozen, i.e.,

$$(1.14) \qquad \varphi^0(x, y) = (\hat{x}_t, y), \quad \text{where } \hat{x}_t \text{ solves} \quad \frac{d\hat{x}}{dt} = g(\hat{x}, y), \quad \hat{x}_0 = x.$$
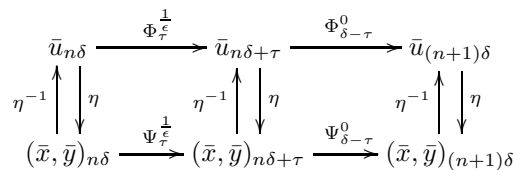
The main effect of FLAVORS is to average the flow of (1.2) with respect to fast degrees of freedom via splitting and resynchronization. By splitting we refer to the substitution of the flow $\varphi^{\frac{1}{\epsilon}}_\delta$ by composition of $\varphi^0_{\delta-\tau}$ and $\varphi^{\frac{1}{\epsilon}}_\tau$, and by resynchronization we refer to the distinct time steps δ and τ whose effects are to advance the internal clock of fast variables by τ every step of length δ. By averaging we refer to the fact that FLAVORS approximate the flow $\varphi^{\frac{1}{\epsilon}}_h$ by the flow

$$(1.15) \qquad \varphi_h := \left(\varphi^0_{\frac{h}{M}-\tau} \circ \varphi^{\frac{1}{\epsilon}}_\tau\right)^M,$$

where h is a coarse time step resolving the slow time scale associated with x^ϵ, M is a positive integer corresponding to the number of samples used to average the flow (δ is identified with $\frac{h}{M}$), and τ is a fine time step resolving the fast time scale, of the order of ϵ, and associated with y^ϵ. In general, analytical formulae are not available for φ^0 and $\varphi^{\frac{1}{\epsilon}}$, and numerical approximations are used instead.

Observe that when FLAVORS are applied to systems with explicitly separated slow and fast processes, they lead to integrators that are locally in the neighborhood of those obtained with HMM (or the equation-free method) with a reinitialization of the fast variables at macrotime n by their final value at macrotime step $n-1$ and with only one microstep per macrostep [38, 40].

We will now consider the situation where η is not the identity diffeomorphism and give the rationale behind the limits (1.9):

$$
\begin{array}{ccccc}
\bar{u}_{n\delta} & \xrightarrow{\Phi^{\frac{1}{\epsilon}}_\tau} & \bar{u}_{n\delta+\tau} & \xrightarrow{\Phi^0_{\delta-\tau}} & \bar{u}_{(n+1)\delta} \\
\eta^{-1} \Big\uparrow \Big\downarrow \eta & & \eta^{-1} \Big\uparrow \Big\downarrow \eta & & \eta^{-1} \Big\uparrow \Big\downarrow \eta \\
(\bar{x}, \bar{y})_{n\delta} & \xrightarrow{\Psi^{\frac{1}{\epsilon}}_\tau} & (\bar{x}, \bar{y})_{n\delta+\tau} & \xrightarrow{\Psi^0_{\delta-\tau}} & (\bar{x}, \bar{y})_{(n+1)\delta}
\end{array}
$$

As illustrated in the above diagram, since $(\bar{x}_t, \bar{y}_t) = \eta(\bar{u}_t)$, simulating $\bar{u}_{n\delta}$ defined in (1.8) is equivalent to simulating the discrete process

$$(1.16) \qquad (\bar{x}_{n\delta}, \bar{y}_{n\delta}) := \left(\Psi^{\frac{1}{\epsilon}}_{\delta-\tau} \circ \Psi^0_\tau\right)^n (x_0, y_0),$$

where

$$(1.17) \qquad \Psi^\alpha_h := \eta \circ \Phi^\alpha_h \circ \eta^{-1}.$$

Observe that the accuracy (in the topology induced by F-convergence) of $\bar{u}_t$ with respect to u^ϵ_t, the solution of (1.1), is equivalent to that of $(\bar{x}_t, \bar{y}_t)$ with respect to $(x^\epsilon_t, y^\epsilon_t)$ defined by (1.2). Now, for the clarity of the presentation, assume that

$$(1.18) \qquad \Phi^\alpha_h(u) = u + hG(u) + \alpha h F(u).$$

Using Taylor's theorem and (1.18), we obtain that
(1.19)
$$\Psi_h^\alpha(x,y) = (x,y) + h\big(g(x,y),0\big) + \alpha h\big(0,f(x,y)\big) + \int_0^1 v^T \operatorname{Hess}\eta(u+tv)v(1-t)^2\,dt$$

with

$$(1.20) \qquad u := \eta^{-1}(x,y) \quad \text{and} \quad v := h(G+\alpha F)\circ\eta^{-1}(x,y).$$

It follows from (1.19) and (1.20) that $\Psi_h^{\frac{1}{\epsilon}}$ is a first order accurate integrator approximating the flow of (1.2) and Ψ_h^0 is a first order accurate integrator approximating the flow of (1.14). Let h be a coarse time step and δ be a mesostep. Since $\bar{x}$ remains nearly constant over the coarse time step, the switching (on and off) of the stiffness parameter $\frac{1}{\epsilon}$ averages the drift g of $\bar{x}$ with respect to the trajectory of $\bar{y}$ over h. Since the coarse step h is composed of $\frac{h}{\delta}$ mesosteps, the internal clock of the fast process is advanced by $\frac{h}{\delta}\times\frac{\tau}{\epsilon}$. Since h is of the order of one, the trajectory of $\bar{y}$ is mixing with respect to the local ergodic measure μ provided that $\frac{\tau}{\delta\epsilon}\gg 1$, i.e.,

$$(1.21) \qquad \delta \ll \frac{\tau}{\epsilon}.$$

Equation (1.21) corresponds to the right-hand side of (1.9). If η is a nonlinear diffeomorphism (with a nonzero Hessian), it also follows from (1.19) and (1.20) that each invocation of the integrator $\Psi_\tau^{\frac{1}{\epsilon}}$ causes an error (on the accuracy of the slow process) proportional to $(\frac{\tau}{\epsilon})^2$. Since, during the coarse time step h, $\Psi_\tau^{\frac{1}{\epsilon}}$ is called $\frac{h}{\delta}$-times, it follows that the error accumulation during h is $\frac{h}{\delta}\times(\frac{\tau}{\epsilon})^2$. Hence, the accuracy of the integrator requires that $\frac{1}{\delta}\times(\frac{\tau}{\epsilon})^2 \ll 1$, i.e.,

$$(1.22) \qquad \left(\frac{\tau}{\epsilon}\right)^2 \ll \delta.$$

Equation (1.22) corresponds to the left-hand side of (1.9).

Observe that if η is linear, its Hessian is null and the remainder in the right-hand side of (1.19) is zero. It follows that if η is linear, the error accumulation due to fine time steps on slow variables is zero and condition (1.21) is sufficient for the accuracy of the integrator.

It has been observed in [39] and in section 5 of [114] that slow variables do not need to be identified with HMM/averaging-type integrators if the relation between original and slow variables is linear or a permutation and if

$$(1.23) \qquad \frac{\Delta t}{M} \ll \frac{\tau}{\epsilon},$$

where M is the number of fine-step iterations used by HMM to compute the average of the drift of slow variables and Δt is the coarse time step (in HMM) along the direction of the averaged drift. The analysis of FLAVORS associated with (1.19) reaches a similar conclusion if η is linear in the sense that the error caused by the Hessian of η in (1.19) is zero and the (sufficient) condition (1.21) is analogous to (1.23) for $M=1$. It is also stated on page 2 of [39] that *"there are counterexamples showing that algorithms of the same spirit do not work for deterministic ODEs with separated time scales if the slow variables are not explicitly identified and made use*

of. But in the present context, the slow variables are linear functions of the original variables, and this is the reason why the seamless algorithm works." Here, the analysis of FLAVORS associated with (1.19) shows that an algorithm based on an averaging principle would indeed, in general, not work if η is nonlinear (and (1.22) not satisfied) due to the error accumulation (on slow variables) associated with the Hessian of η. However, the above analysis also shows that if condition (1.22) is satisfied, then, although η may be *nonlinear*, flow averaging integrators will *always* work without *identifying slow variables*.

1.5. Nonasymptotic convergence result.

THEOREM 1.3. *Under assumptions and notations of Theorem 1.2, there exists $C > 0$ such that for $\delta < h_0$, $\tau < \tau_0 \epsilon$, and $t > 0$*

$$|x_t^\epsilon - \eta^x(\bar{u}_t)| \leq C e^{Ct} \chi_1(u_0, \epsilon, \delta, \tau) \tag{1.24}$$

and

$$\left| \frac{1}{T} \int_t^{t+T} \varphi(\bar{u}_s) \, ds - \int_{\mathbb{R}^p} \varphi(\eta^{-1}(X_t, y)) \mu(X_t, dy) \right|$$

$$\leq \chi_2(u_0, \epsilon, \delta, \tau, T, t)(\|\varphi\|_{L^\infty} + \|\nabla\varphi\|_{L^\infty}), \tag{1.25}$$

where χ_1 and χ_2 are functions converging towards zero as $\epsilon \leq \delta/(C \ln \frac{1}{\delta})$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\epsilon}{\tau}\delta \downarrow 0$, and $(\frac{\tau}{\epsilon})^2 \frac{1}{\delta} \downarrow 0$ (and $T \downarrow 0$ for χ_2).

REMARK 1.5. *For $\epsilon \leq \delta/(-C \ln \delta)$ and $\delta\frac{\epsilon}{\tau} + \frac{\tau}{\epsilon} \leq 1$, the following holds:*

$$\chi_1(u_0, \epsilon, \delta, \tau) \leq \sqrt{\delta} + \left(\frac{\tau}{\epsilon}\right)^2 \frac{1}{\delta} + E\left(\frac{1}{C} \ln \frac{1}{\delta}\right) + \left(\frac{\delta\epsilon}{\tau}\right)^{\frac{1}{2}} + \left(\frac{\tau}{\epsilon}\right)^{\frac{1}{2}}$$

$$+ E\left(\frac{1}{C} \ln\left(\left(\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon}\right)^{-1}\right)\right), \tag{1.26}$$

and χ_2 satisfies a similar inequality.

REMARK 1.6. *Choosing $\tau \sim \gamma\epsilon$ and $\delta \sim \gamma\frac{\tau}{\epsilon}$, where γ is a small constant independent from ϵ, Theorem 1.3 shows that the approximation error of FLAVOR is bounded by a function of γ converging towards zero as $\gamma \downarrow 0$. It follows that the speedup is of the order of $\frac{\delta}{\tau} \sim \frac{\gamma}{\epsilon}$, i.e., scales like $\frac{1}{\epsilon}$ at fixed accuracy. In order to be able to compare FLAVOR with integrators resolving all the fine time steps, we have limited the speedup in the numerical experiments to 200x (but the latter can be arbitrarily large as $\epsilon \downarrow 0$). For sufficiently small ϵ, we observe that FLAVORS with microstep τ and mesostep δ overperform their associated legacy integrator with the same microstep τ over large simulation times (we refer the reader to section 6.3 on the Fermi–Pasta–Ulam problem). This phenomenon is caused by an error accumulation at each tick (microstep) of the clock of fast variables. Since FLAVORS (indirectly, i.e., without identifying fast variables) slow down the speed of this clock from $\frac{1}{\epsilon}$ to a value $\frac{\tau}{\delta\epsilon} \sim \frac{1}{\gamma}$ independent from ϵ, this error does not blow up as $\epsilon \downarrow 0$ (as opposed to an integrator resolving all fine time steps). For this reason, if this error accumulation on fast variables is exponential, then the speedup at fixed accuracy does not scale like $\frac{1}{\epsilon}$, but like $e^{\frac{T}{\epsilon}}$, where T is the total simulation time. A consequence of this phenomenon can be seen in Figure 6.5 (associated with the Fermi–Pasta–Ulam problem), where Velocity Verlet fails to capture the $\mathcal{O}(\epsilon^{-1})$ dynamics with a time step $h = 10^{-5}$, whereas FLAVORS remain accurate with $\tau = 10^{-4}$ and $\delta = 2 \cdot 10^{-3}$.*

REMARK 1.7. *The reader should not be surprised by the presence of the exponential factor e^{Ct} in* (1.24). *It is known that global errors for numerical approximations of ODEs grow, in general, exponentially with time (see, for instance,* [58]). *These bounds are, however, already tight; consider, for instance, how error propagates in a generic numerical scheme applied to the special system of $\dot{x} = x$. It is possible to show that the increase of global errors is linear in time only for a restricted class of ODEs (using techniques from Lyapunov's theory of stability* [118]). *Notice that the constant C in the exponential of our bound does not scale with ϵ^{-1}, and therefore the bound is uniform and rather tight.*

REMARK 1.8. *We refer the reader to* [41] *for higher order averaging based methods. In particular,* [41] *shows how, after identification of slow variables, balancing the different error contributions yields an explicit stable integration method having the order of the macroscheme.*

1.6. Natural FLAVORS. Although convenient, it is not necessary to use legacy integrators to obtain FLAVORS. More precisely, Theorems 1.2 and 1.3 remain valid if FLAVORS are defined to be algorithms simulating the discrete process

$$(1.27) \qquad \bar{u}_t := \left(\theta_{\delta-\tau}^G \circ \theta_\tau^\epsilon\right)^k(u_0) \quad \text{for} \quad k\delta \leq t < (k+1)\delta,$$

where θ_τ^ϵ and $\theta_{\delta-\tau}^G$ are two mappings from $\mathbb{R}^d$ onto $\mathbb{R}^d$ (the former approximating the flow of the whole system (1.1) for time τ and the latter approximating the flow of $\dot{v} = G(v)$ for time $\delta - \tau$), satisfying the following condition.

CONDITION 1.4.
1. *There exists h_0, $C > 0$, such that for $h \leq h_0$ and any $u \in \mathbb{R}^d$*

$$(1.28) \qquad \left|\theta_h^G(u) - u - hG(u)\right| \leq Ch^2.$$

2. *There exists τ_0, $C > 0$, such that for $\frac{\tau}{\epsilon} \leq \tau_0$ and any $u \in \mathbb{R}^d$*

$$(1.29) \qquad \left|\theta_\tau^\epsilon(u) - u - \tau G(u) - \frac{\tau}{\epsilon}F(u)\right| \leq C\left(\frac{\tau}{\epsilon}\right)^2.$$

3. *For all u_0, $T > 0$, the discrete trajectories $\left(\left(\theta_{\delta-\tau}^G \circ \theta_\tau^\epsilon\right)^k(u_0)\right)_{0 \leq k \leq T/\delta}$ are uniformly bounded in ϵ, $0 < \delta \leq h_0$, and $\tau \leq \min(\tau_0\epsilon, \delta)$.*

Observe that (1.8) is a particular case of (1.27) in which $\theta^\epsilon = \Phi^{\frac{1}{\epsilon}}$ and the mapping θ^G is obtained from the legacy integrator Φ^α by setting α to zero.

1.7. Related work. *"Dynamical systems with multiple time scales pose a major problem in simulations because the small time steps required for stable integration of the fast motions lead to large numbers of time steps required for the observation of slow degrees of freedom"* [113, 57]. Traditionally, stiff dynamical systems have been separated into two classes with distinct integrators: stiff systems with fast transients and stiff systems with rapid oscillations [6, 36, 100]. The former have been solved using implicit schemes [49, 35, 57, 59], Chebyshev methods [73, 1], or the projective integrator approach [51]. The latter have been solved using filtering techniques [48, 69, 104] or Poincaré map techniques [50, 94]. We also refer to methods based on highly oscillatory quadrature [32, 62, 63], an area that has undergone significant developments in the last few years [64]. It has been observed that *"at the present time, there exists no unified strategy for dealing with both classes of problems"* [36]. When slow variables can be identified, effective equations can be obtained by averaging the

instantaneous drift driving those slow variables [108]. Two classes of numerical methods have been built on this observation: the equation-free method [67, 68] and HMM [37, 41, 36, 5]. Observe that FLAVORS apply in a unified way to both stiff systems with fast transients and stiff systems with rapid oscillations, with or without noise, with a mesoscopic integration time step chosen independently from the stiffness.

1.8. Limitations of the method. The proof of the accuracy of the method (Theorems 1.2 and 1.3) is based on an averaging principle; hence, if ϵ is not small (the stiffness of the ODE is weak), although the method may be stable, there is no guarantee of accuracy. More precisely, the global error of the method is an increasing function of ϵ, δ, $\frac{\tau}{\epsilon}$, $\frac{\delta\epsilon}{\tau}$, $(\frac{\tau}{\epsilon})^2\delta$. Writing $\gamma := \frac{\tau}{\epsilon}$, the accuracy requires $\gamma^2 \ll \delta \ll \gamma$. Choosing $\delta = \gamma^{\frac{3}{2}}$, the condition $\epsilon \ll \delta \ll 1$ (related to computational gain) requires $\epsilon^{\frac{2}{3}} \ll \gamma \ll 1$, which can be satisfied only if ϵ is small.

The other limitation of the method lies in the fact that a stiff parameter $\frac{1}{\epsilon}$ needs to be clearly identified. In many examples of interest (Navier–Stokes equations, Maxwell's equations, etc.), stiffness is a result of nonlinearity, initial conditions, or boundary conditions and not of the existence of a large parameter $\frac{1}{\epsilon}$. Molecular dynamics can also create widely separated time scales from nonlinear effects; we refer the reader, for instance, to [119] and the references therein.

1.9. Generic stiff ODEs. FLAVORS have a natural generalization to systems of the form

$$(1.30) \qquad \dot{u}^{\alpha,\epsilon} = F(u^{\alpha,\epsilon}, \alpha, \epsilon),$$

where $u \mapsto F(u, \alpha, \epsilon)$ is Lipschitz continuous.

CONDITION 1.5. *Assume that the following hold:*
1. *$\epsilon \mapsto F(u, \alpha, \epsilon)$ is uniformly continuous in the neighborhood of 0.*
2. *There exists a diffeomorphism $\eta := (\eta^x, \eta^y)$, from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^p$, independent from ϵ, α, with uniformly bounded C^1, C^2 derivatives, such that the process $(x_t^\alpha, y_t^\alpha) = \left(\eta^x(u_t^{\alpha,0}), \eta^y(u_t^{\alpha,0})\right)$ satisfies, for all $\alpha \geq 1$, the ODE*

$$(1.31) \qquad \dot{x}^\alpha = g(x^\alpha, y^\alpha), \qquad x_0^\alpha = x_0,$$

   where $g(x, y)$ is Lipschitz continuous in x and y on bounded sets.
3. *There exists a family of probability measures $\mu(x, dy)$ on $\mathbb{R}^p$ such that for all $x_0, y_0, T \left((x_0, y_0) := \eta(u_0)\right)$, and φ uniformly bounded and Lipschitz*

$$\left| \frac{1}{T} \int_0^T \varphi(y_s^\alpha) \, ds - \int_{\mathbb{R}^p} \varphi(y) \mu(x_0, dy) \right|$$
$$(1.32) \qquad \leq \chi\big(\|(x_0, y_0)\|\big)\big(E_1(T) + E_2(T\alpha^\nu)\big)\|\nabla\varphi\|_{L^\infty},$$

   where $r \mapsto \chi(r)$ is bounded on compact sets and $E_2(r) \to 0$ as $r \to \infty$ and $E_1(r) \to 0$ as $r \to 0$.
4. *For all u_0, $T > 0$, the trajectories $(u_t^{\alpha,0})_{0 \leq t \leq T}$ are uniformly bounded in $\alpha \geq 1$.*

REMARK 1.9. *Observe that slow variables are not kept frozen in* (1.32). *The error on local invariant measures induced by the (slow) drift of x^α is controlled by E_2. More precisely, the convergence of the right-hand side of* (1.32) *towards zero requires the convergence of T towards zero and (at the same time) the divergence of $T\alpha^\nu$ towards infinity.*

Assume that we are given a mapping $\Phi_h^{\alpha,\epsilon}$ from $\mathbb{R}^d$ onto $\mathbb{R}^d$ approximating the flow of (1.30). If the parameter α can be controlled, then $\Phi_h^{\alpha,\epsilon}$ can be used as a black box for accelerating the computation of solutions of (1.30).

CONDITION 1.6. *Assume that the following hold:*

1. *There exists a constant $h_0 > 0$ such that $\Phi^{\alpha,\epsilon}$ satisfies for all $h \leq h_0 \min(\frac{1}{\alpha^\nu}, 1)$,*
   $0 < \epsilon \leq 1 \leq \alpha$

$$(1.33) \qquad \left| \Phi_h^{\alpha,\epsilon}(u) - u - hF(u,\alpha,\epsilon) \right| \leq C(u)h^2(1 + \alpha^{2\nu}),$$

   where $C(u)$ is bounded on compact sets.

2. *For all u_0, $T > 0$, the discrete trajectories $\left(\left(\Phi_{\delta-\tau}^{0,\epsilon} \circ \Phi_\tau^{\frac{1}{\epsilon},\epsilon} \right)^k (u_0) \right)_{0 \leq k \leq T/\delta}$ are uniformly bounded in $0 < \epsilon \leq 1$, $0 < \delta \leq h_0$, and $\tau \leq \min(h_0 \epsilon^\nu, \delta)$.*

FLAVOR can be defined as the algorithm given by the process

$$(1.34) \qquad \bar{u}_t = \left(\Phi_{\delta-\tau}^{0,\epsilon} \circ \Phi_\tau^{\frac{1}{\epsilon},\epsilon} \right)^k (u_0) \quad \text{for} \quad k\delta \leq t < (k+1)\delta.$$

The theorem below shows the accuracy of FLAVORS for $\delta \ll h_0$, $\tau \ll \epsilon^\nu$, and $\left(\frac{\tau}{\epsilon^\nu} \right)^2 \ll \delta \ll \frac{\tau}{\epsilon^\nu}$.

THEOREM 1.4. *Let $u_t^{\frac{1}{\epsilon},\epsilon}$ be the solution to (1.30) with $\alpha = 1/\epsilon$, and let $\bar{u}_t$ be defined by (1.34). Assume that Conditions 1.5 and 1.6 are satisfied. Then the following hold:*

- *$u_t^{\frac{1}{\epsilon},\epsilon}$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy) \right)$ as $\epsilon \downarrow 0$, where X_t is the solution to*

$$(1.35) \qquad \dot{X}_t = \int_{\mathbb{R}^p} g(X_t, y)\, \mu(X_t, dy), \qquad X_0 = x_0.$$

- *As $\epsilon \downarrow 0$, $\tau \epsilon^{-\nu} \downarrow 0$, $\delta \frac{\epsilon^\nu}{\tau} \downarrow 0$, and $\frac{\tau^2}{\epsilon^{2\nu}\delta} \downarrow 0$, $\bar{u}_t$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy) \right)$ as $\epsilon \downarrow 0$, where X_t is the solution of (1.35).*

Proof. The proof of Theorem 1.4 is similar to that of Theorems 1.2 and 3.2. Only the idea of the proof will be given here. The condition $\epsilon \ll 1$ is needed for the approximation of $u^{\alpha,\epsilon}$ by $u^{\alpha,0}$ and for the F-convergence of $u^{\frac{1}{\epsilon},0}$. Since $y_t^\alpha = \eta^y(u_t^{\alpha,0})$, the condition $\tau \ll \epsilon^\nu$ is used along with (1.33) for the accuracy of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ in (locally) approximating y_t^α. The condition $\delta \ll \frac{\tau}{\epsilon^\nu}$ allows for the averaging of g to take place prior to a significant change of x_t^α; more precisely, it allows for $m \gg 1$ iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ prior to a significant change of x_t^α. The condition $\left(\frac{\tau}{\epsilon^\nu} \right)^2 \ll \delta$ is required in order to control the error accumulated by m iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$. $\square$

REMARK 1.10. *It is easy to see that Theorem 1.4 remains valid if item 4 of Condition 1.5 and item 2 of Condition 1.6 do not hold for all u_0 but only for a subset of initial conditions $u_0 \in \mathcal{I}$ and if the trajectories of u and $\bar{u}$ remain in that subset for all ϵ.*

We also observe that Theorem 1.4 can easily be generalized to situations where η is noninjective, for instance to a situation where η is a differentiable mapping from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^q$, where $q < p$. In that situation, item 4 of Condition 1.5 and item 2 of Condition 1.6 should be replaced by the condition that $\eta(u)$ and $\eta(\bar{u})$ do not blow up as $\epsilon \downarrow 0$. Furthermore, the convergence of u and $\bar{u}$ is only partial in the sense that $\eta(u)$ F-converges towards $\delta_{X_t} \otimes \mu(X_t, dy)$ but the projection of u on the kernel of η (i.e., $\eta^{-1}(0,0)$) may not F-converge.

2. Deterministic mechanical systems: Hamiltonian equations. Since averaging with FLAVORS is obtained by flow composition, FLAVORS have an inherent extension to multiscale structure preserving integrators for stiff Hamiltonian systems, i.e., ODEs of the form

$$(2.1) \qquad \dot{p} = -\partial_q H(p,q), \qquad \dot{q} = \partial_p H(p,q),$$

where the Hamiltonian

$$(2.2) \qquad H(q,p) := \frac{1}{2} p^T M^{-1} p + V(q) + \frac{1}{\epsilon} U(q)$$

represents the total energy of a mechanical system with Euclidean phase space $\mathbb{R}^d \times \mathbb{R}^d$ or a cotangent bundle $T^*\mathcal{M}$ of a configuration manifold $\mathcal{M}$.

Structure preserving numerical methods for Hamiltonian systems have been developed in the framework of geometric numerical integration [57, 75] and variational integrators [82, 78]. *"The subject of geometric numerical integration deals with numerical integrators that preserve geometric properties of the flow of a differential equation, and it explains how structure preservation leads to an improved long-time behavior"* [56]. Variational integration theory derives integrators for mechanical systems from discrete variational principles that are characterized by a discrete Noether theorem. These methods have excellent energy behavior over long integration runs because they are symplectic; i.e., by backward error analysis, they simulate a nearby mechanical system instead of nearby differential equations. Furthermore, statistical properties of the dynamics such as Poincaré sections are well preserved even with large time steps [16]. Preservation of structures is especially important for long time simulations. Consider integrations of a harmonic oscillator, for example: no matter how small a time step is used, the amplitude given by forward Euler/backward Euler will increase/decrease unboundedly, whereas the amplitude given by variational Euler (also known as symplectic Euler) will be oscillatory with a variance controlled by the step length.

These long term behaviors of structure preserving numerical integrators motivated their extension to multiscale or stiff Hamiltonian systems. We refer the reader to [31] for a recent review on numerical integrators for highly oscillatory Hamiltonian systems. *"Symplectic integrators are natural for the integration of Hamiltonian systems since they reproduce at the discrete level an important geometric property of the exact flow"* [20]. For symplectic integrators primarily for (but not limited to) stiff quadratic potentials, we refer the reader to the impulse method, the mollified impulse method, and their variations [54, 113, 47, 101], which require an explicit form of the flow map of the stiff process. In the context of variational integrators, by defining a discrete Lagrangian with an explicit trapezoidal approximation of the soft potential and a midpoint approximation for the fast potential, a symplectic (IMEX—IMplicit– EXplicit) scheme for stiff Hamiltonian systems has been proposed in [109]. The resulting scheme is explicit for quadratic potentials and implicit for nonquadratic stiff potentials. We also refer the reader to Le Bris and Legoll's (Hamilton–Jacobi derived) homogenization method [20]. Asynchronous variational integrators [77] provide a way to derive conservative symplectic integrators for PDEs where the solution advances nonuniformly in time; however, stiff potentials require a fine time step discretization over the whole time evolution. In addition, multiple time step methods [110] evaluate forces to different extents of accuracies by approximating less important forces via Taylor expansions, but they have issues on long time behavior, stability, and accuracy, as described in section 5 of [76]. Fixman froze the fastest bond oscillations in

polymers to remove stiffness by adding a log term resemblant of entropy-based free energy to compensate [44]. This approach is successful in studying statistics of the system but does not always reconstruct the correct dynamics [95, 93, 14].

Several approaches to the homogenization of Hamiltonian systems (in analogy with classical homogenization [11, 65]) have been proposed. We refer the reader to $\mathcal{M}$-convergence introduced in [105, 15], to the two-scale expansion of solutions of the Hamilton–Jacobi form of Newton's equations with stiff quadratic potentials [20], and to PDE methods in weak KAM theory [42]. We also refer the reader to [26, 61, 99].

Obtaining explicit symplectic integrators for Hamiltonian systems with nonquadratic stiff potentials is known to be an important and nontrivial problem. By using Verlet/leap-frog macrosolvers, methods that are symplectic on slow variables (when those variables can be identified) have been proposed in the framework of HMM in [106, 24]. A "reversible averaging" method has been proposed in [74] for mechanical systems with separated fast and slow variables. More recently, a reversible multiscale integration method for mechanical systems was proposed in [6] in the context of HMM. By tracking slow variables, [6] enforces reversibility in all variables as an optimization constraint at each coarse step when minimizing the distance between the effective drift obtained from the microsolver (in the context of HMM) and the drift of the macrosolver. We also refer the reader to [103] for a symmetric HMM for mechanical systems with stiff potentials of the form $\frac{1}{\epsilon} \sum_{j=1}^{\nu} g_j(q)^2$.

2.1. FLAVORS for mechanical systems on manifolds. Assume that we are given a first order accurate legacy integrator for (2.1) in which the parameter $1/\epsilon$ can be controlled, i.e., a mapping Φ_h^α acting on the phase space such that for $h \leq h_0 \min(1, \alpha^{-\frac{1}{2}})$

$$(2.3) \qquad \left| \Phi_h^\alpha(q, p) - (q, p) - h\left(M^{-1}p, -V(q) - \alpha U(q)\right) \right| \leq Ch^2(1 + \alpha).$$

Write Θ_δ as the FLAVOR discrete mapping approximating solutions of (2.1) over time steps $\delta \gg \epsilon$, i.e.,

$$(2.4) \qquad (q_{(n+1)\delta}, p_{(n+1)\delta}) := \Theta_\delta(q_{n\delta}, p_{n\delta}).$$

FLAVOR can then be defined by

$$(2.5) \qquad \Theta_\delta := \Phi_{\delta-\tau}^0 \circ \Phi_\tau^{\frac{1}{\epsilon}}.$$

Theorem 1.4 establishes the accuracy of this integrator under Conditions 1.5 and 1.6 provided that $\tau \ll \sqrt{\epsilon} \ll \delta$ and $\frac{\tau^2}{\epsilon} \ll \delta \ll \frac{\tau}{\sqrt{\epsilon}}$.

REMARK 2.1. *We also refer the reader to Remark* 1.10 *for the application of Theorem* 1.4 *to Hamiltonian systems. Consider, for instance, the linear Hamiltonian system* $H(q_1, q_2, p_1, p_2) := \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 + \frac{1}{2}q_1^2 + \frac{1}{\epsilon}(q_2 - q_1)^2$. *If the system is started from* $q_2(0) - q_1(0) = \mathcal{O}(\sqrt{\epsilon})$, *then the energy remains bounded as* $\epsilon \downarrow 0$ *and* (q_1, q_2, p_1, p_2) *F-converges due to the first part of Remark* 1.10.

For that same example, if the system is started from a point such that $q_2(0) - q_1(0) \neq \mathcal{O}(\sqrt{\epsilon})$, *then the energy in the system blows up as* $\epsilon \downarrow 0$ *and the range of* $p_2 - p_1$ *blows up and therefore cannot converge, even in the sense of measures. However, the (slow) process* $(q_1 + q_2, p_1 + p_2)$ *satisfies an equation of the type* (1.31), *where the dependence on fast variables is only through* $q_2 - q_1$ *(* $y^\alpha = q_2 - q_1$ *in* (1.31)) *and* $q_2 - q_1$ *is locally ergodic (as defined in item 3 of Condition* 1.5) *and does converge in the sense of distributions. Henceforth, if* $q_2(0) - q_1(0) \neq 0$, *then the generalization of Theorem* 1.4 *(see the second part of Remark* 1.10) *applies with* η *being noninjective.*

2.1.1. Structure preserving properties of FLAVORS. We will now show that FLAVORS inherit the structure preserving properties of their legacy integrators.

THEOREM 2.1. *If, for all $h, \epsilon > 0$, Φ_h^ϵ is symmetric under a group action, then Θ_δ is symmetric under the same group action.*

THEOREM 2.2. *If Φ_h^α is symplectic on the cotangent bundle $T^*\mathcal{M}$ of a configuration manifold $\mathcal{M}$, then Θ_δ defined by (2.5) is symplectic on the cotangent bundle $T^*\mathcal{M}$.*

Theorems 2.1 and 2.2 can be resolved by noting that "the overall method is symplectic—as a composition of symplectic transformations, and it is symmetric—as a symmetric composition of symmetric steps" (see Chapter XIII.1.3 of [57]).

Write

$$\Phi_h^* := \left(\Phi_{-h}\right)^{-1}. \tag{2.6}$$

Let us recall the following definition corresponding to Definition 1.4 of Chapter V of [57].

DEFINITION 2.3. *A numerical one-step method Φ_h is called time-reversible if it satisfies $\Phi_h^* = \Phi_h$.*

The following theorem, whose proof is straightforward, shows how to derive a "symplectic and symmetric and time-reversible" FLAVOR from a symplectic legacy integrator and its adjoint. Since this derivation applies to manifolds, it also leads to structure preserving FLAVORS for constrained mechanical systems.

THEOREM 2.4. *If Φ_h^α is symplectic on the cotangent bundle $T^*\mathcal{M}$ of a configuration manifold $\mathcal{M}$, then*

$$\Theta_\delta := \Phi_{\frac{1}{\frac{\epsilon}{\tau}}}^{\frac{1}{\frac{\epsilon}{\tau}},*} \circ \Phi_{\frac{\delta-\tau}{2}}^{0,*} \circ \Phi_{\frac{\delta-\tau}{2}}^{0} \circ \Phi_{\frac{\frac{1}{\epsilon}}{2}}^{\frac{1}{\epsilon}} \tag{2.7}$$

is symplectic and time-reversible on the cotangent bundle $T^*\mathcal{M}$.

REMARK 2.2. *Observe that (except for the first and last steps) iterating Θ_δ defined by (2.7) is equivalent to iterating*

$$\Theta_\delta := \Phi_{\frac{\delta-\tau}{2}}^{0,*} \circ \Phi_{\frac{\delta-\tau}{2}}^{0} \circ \Phi_{\frac{\frac{1}{\epsilon}}{2}}^{\frac{1}{\epsilon}} \circ \Phi_{\frac{\frac{1}{\epsilon}}{2}}^{\frac{1}{\epsilon},*}. \tag{2.8}$$

It follows that a symplectic, symmetric, and reversible FLAVOR can be obtained in a nonintrusive way from a Störmer/Verlet integrator for (2.1) [56, 58, 117].

2.1.2. An example of a symplectic FLAVOR. If the phase space is $\mathbb{R}^d \times \mathbb{R}^d$, then an example of a symplectic FLAVOR is obtained from Theorem 2.2 by choosing Φ_h^α to be the symplectic Euler (also known as variational Euler or VE for short) integrator defined by

$$\Phi_h^\alpha(q, p) = \begin{pmatrix} q \\ p \end{pmatrix} + h \begin{pmatrix} M^{-1}\big(p - h\big(V(q) + \alpha U(q)\big)\big) \\ -V(q) - \alpha U(q) \end{pmatrix} \tag{2.9}$$

and letting Θ_δ be defined by (2.5).

2.1.3. An example of a symplectic and time-reversible FLAVOR. If the phase space is the Euclidean space $\mathbb{R}^d \times \mathbb{R}^d$, then an example of a symplectic and time-reversible FLAVOR is obtained by letting Θ_δ be defined by (2.7) of Theorem 2.4 by choosing Φ_h^α to be the symplectic Euler integrator defined by (2.9) and

$$\Phi_h^{\alpha,*}(q, p) = \begin{pmatrix} q \\ p \end{pmatrix} + h \begin{pmatrix} M^{-1}p \\ -V(q + hM^{-1}p) - \alpha U(q + hM^{-1}p) \end{pmatrix}. \tag{2.10}$$

2.1.4. An artificial FLAVOR. There is not a unique way of averaging the flows of (2.2). We present below an alternative method based on the freezing and unfreezing of degrees of freedom associated with fast potentials. We have called this method "artificial" because it is intrusive. With this method, the discrete flow approximating solutions of (2.1) is given by (2.4) with

$$(2.11) \qquad \Theta_\delta := \theta_{\delta-\tau}^{tr} \circ \theta_\tau^\epsilon \circ \theta_\delta^V,$$

where θ_δ^V is a symplectic map corresponding to the flow of $H^{slow}(q,p) := V(q)$, approximating the effects of the soft potential on momentum over the mesoscopic time step δ and defined by

$$(2.12) \qquad \theta_\delta^V(q,p) = (q, p - \delta\nabla V(q)).$$

θ_τ^ϵ is a symplectic map approximating the flow of $H^{fast}(q,p) := \frac{1}{2}p^T M^{-1}p + \frac{1}{\epsilon}U(q)$ over a microscopic time step τ:

$$(2.13) \qquad \theta_\tau^\epsilon(q,p) = \left(q + \tau M^{-1}p, p - \frac{\tau}{\epsilon}\nabla U(q + tM^{-1}p)\right).$$

$\theta_{\delta-\tau}^{tr}$ is a map approximating the flow of the Hamiltonian $H^{free}(q,p) := \frac{1}{2}p^T M^{-1}p$ under holonomic constraints imposing the freezing of stiff variables. Velocities along the direction of constraints have to be stored and set to be 0 before the constrained dynamics, i.e., frozen, and the stored velocities should be restored after the constrained dynamics, i.e., unfrozen; geometrically speaking, one projects to the constrained sub-symplectic manifold, runs the constrained dynamics, and lifts back to the original full space. Oftentimes, the exact solution to the constrained dynamics can be found (examples are given in subsections 5.3, 5.2, 6.2, 6.3, and 6.4).

When the exact solution to the constrained dynamics cannot be found easily, one may want to employ integrators for constrained dynamics such as SHAKE [98] or RATTLE [4] instead. This has to be done with caution because symplecticity of the translational flow may be lost. The composition of projection onto the constrained manifold (freezing), evolution on the constrained manifold, and lifting from it to the unconstrained space (unfreezing) preserves symplecticity in the unconstrained space only if the evolution on the constrained manifold preserves the inherited symplectic form. A numerical integration preserves the discrete symplectic form on the constrained manifold but not necessarily the projected continuous symplectic form.

REMARK 2.3. *This artificial FLAVOR is locally a perturbation of nonintrusive FLAVORS. By splitting theory* [85, 57],

$$(2.14) \qquad \theta_{\delta-\tau}^{tr} \circ \theta_\tau^\epsilon \circ \theta_\delta^V \approx \theta_{\delta-\tau}^{tr} \circ \theta_{\delta-\tau}^V \circ \theta_\tau^\epsilon \circ \theta_\tau^V \approx \theta_{\delta-\tau}^{tr} \circ \theta_{\delta-\tau}^V \circ \Phi_\tau^{\frac{1}{\epsilon}},$$

whereas $\Phi_{\delta-\tau}^0 \circ \Phi_\tau^{\frac{1}{\epsilon}} \approx \theta_{\delta-\tau}^{free} \circ \theta_{\delta-\tau}^V \circ \Phi_\tau^{\frac{1}{\epsilon}}$, *where* θ^{free} *is the flow of* $H^{free}(q,p)$ *under no constraint. The only difference is that constraints are treated in* θ^{tr} *but not in* θ^{free}.

REMARK 2.4. *This artificial FLAVOR can be formally regarded as* $\Phi_{\delta-\tau}^\infty \circ \Phi_\tau^{\frac{1}{\epsilon}}$. *In contrast, the natural FLAVOR is* $\Phi_{\delta-\tau}^0 \circ \Phi_\tau^{\frac{1}{\epsilon}}$.

The advantage of this artificial FLAVOR lies in the fact that only $\tau \ll \sqrt{\epsilon} \ll \delta$ and $\delta \ll \frac{\tau}{\sqrt{\epsilon}}$ are required for its accuracy (and not $\frac{\tau^2}{\epsilon} \ll \delta$). We also observe that, in general, the artificial FLAVOR overperforms the nonintrusive FLAVOR in Fermi–Pasta–Ulam long time ($\mathcal{O}(\omega^2)$) simulations (we refer the reader to subsection 6.3).

2.2. Variational derivation of FLAVORS. FLAVORS based on variational legacy integrators [82] are variational too. Recall that discrete Lagrangian L_d is an approximation of the integral of the continuous Lagrangian over one time step, and the discrete Euler–Lagrangian (DEL) equation is obtained by applying the variational (least action) principle to the discrete action, which is a sum of discrete Lagrangians. The following diagram commutes:

$$
\begin{array}{ccc}
Singlescale \; L_d & \xrightarrow{\;FLAVORization\;} & Multiscale \; L_d \\
\Big\downarrow {\scriptstyle variational\; principle} & & \Big\downarrow {\scriptstyle variational\; principle} \\
Singlescale \; DEL & \xrightarrow{\;FLAVORization\;} & Multiscale \; DEL
\end{array}
$$

For example, recall that VE (i.e., symplectic Euler) for system (2.2) with time step h

$$
(2.15) \qquad \begin{cases} p_{k+1} = p_k - h[\nabla V(q_k) + \frac{1}{\epsilon}\nabla U(q_k)], \\ q_{k+1} = q_k + hp_{k+1} \end{cases}
$$

can be obtained by applying the variational principle to the following discrete Lagrangian:

$$
(2.16) \qquad L_{d_h}^{1/\epsilon}(q_k, q_{k+1}) = h\left[\frac{1}{2}\left(\frac{q_{k+1}-q_k}{h} \right)^2 - \left(V(q_k) + \frac{1}{\epsilon}U(q_k) \right) \right].
$$

Meanwhile, FLAVORized VE with small step τ and mesostep δ

$$
(2.17) \qquad \begin{cases} p'_k = p_k - \tau[\nabla V(q_k) + \frac{1}{\epsilon}\nabla U(q_k)], \\ q'_k = q_k + \tau p'_k, \\ p_{k+1} = p'_k - (\delta - \tau)\nabla V(q'_k), \\ q_{k+1} = q'_k + (\delta - \tau)p_{k+1} \end{cases}
$$

can be obtained by applying the variational principle to the FLAVORized discrete Lagrangian

$$
(2.18)
$$
$$
L_{d\delta}(q_k, q'_k, q_{k+1}) = L_{d\tau}^{1/\epsilon}(q_k, q'_k) + L_{d\delta-\tau}^{0}(q'_k, q_{k+1})
$$
$$
= \tau\left[\frac{1}{2}\left(\frac{q'_k - q_k}{\tau} \right)^2 - \left(V(q_k) + \frac{1}{\epsilon}U(q_k) \right) \right] + (\delta - \tau)\left[\frac{1}{2}\left(\frac{q_{k+1}-q'_k}{\delta - \tau} \right)^2 - V(q'_k) \right].
$$

FLAVORizations of other variational integrators such as Velocity Verlet follow similarly.

3. SDEs. *"Asymptotic problems for stochastic differential equations arose and were solved simultaneously with the very beginnings of the theory of such equations"* [108]. Here, we refer the reader to the early work of Gihman [52], Kryloff and Bogoliouboff [70, 71], Bogolyubov [13], and Papanicolaou and Kohler [91]. We refer the reader in particular to Skorokhod's detailed monograph [108]. As for ODEs, effective equations for stiff SDEs can be obtained by averaging the instantaneous coefficients (drift and the diffusivity matrix squared) with respect to the fast components; we

refer the reader to section 3 of Chapter II of [108] for a detailed analysis including error bounds. Numerical methods such as HMM [38] and the equation-free method [7] have been extended to SDEs based on this averaging principle. *"Implicit methods in general fail to capture the effective dynamics of the slow time scale because they cannot correctly capture non-Dirac invariant distributions"* [79] (we refer to a non-Dirac invariant distribution as a measure of probability on the configuration space whose support is not limited to a single point). Another idea is to treat fast variables by conditioning; here, we refer the reader to optimal prediction [28, 27, 29] that has also been used for model reduction. We also refer the reader to [8, 55, 112, 23, 22, 79, 2].

Since FLAVORS are obtained via flow averaging, they have a natural extension to the SDEs developed in this section. As for ODEs, FLAVORS are directly applied to SDEs with mixed (hidden) slow and fast variables without prior (analytical or numerical) identification of slow variables. Furthermore, they can be implemented using a pre-existing scheme by turning the stiff parameters on and off.

For the sake of clarity, we will start the description with the following SDE on $\mathbb{R}^d$:

$$(3.1) \qquad du_t^\epsilon = \left(G(u_t^\epsilon) + \frac{1}{\epsilon} F(u_t^\epsilon) \right) dt + \left(H(u_t^\epsilon) + \frac{1}{\sqrt{\epsilon}} K(u_t^\epsilon) \right) dW_t, \qquad u_0^\epsilon = u_0,$$

where $(W_t)_{t \geq 0}$ is a d-dimensional Brownian motion; F and G are vector fields on $\mathbb{R}^d$; and H and K are $d \times d$ matrix fields on $\mathbb{R}^d$. In subsection 3.5, we will consider the more general form (3.15).

CONDITION 3.1. *Assume that the following hold:*
1. *F, G, H, and K are uniformly bounded and Lipschitz continuous.*
2. *There exists a diffeomorphism $\eta := (\eta^x, \eta^y)$, from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^p$, independent of ϵ, with uniformly bounded C^1, C^2, and C^3 derivatives, such that the process $(x_t^\epsilon, y_t^\epsilon) = (\eta^x(u_t^\epsilon), \eta^y(u_t^\epsilon))$ satisfies the SDE*

$$(3.2) \qquad \begin{cases} dx^\epsilon = g(x^\epsilon, y^\epsilon) \, dt + \sigma(x^\epsilon, y^\epsilon) dW_t, & x_0^\epsilon = x_0, \\ dy^\epsilon = \frac{1}{\epsilon} f(x^\epsilon, y^\epsilon) \, dt + \frac{1}{\sqrt{\epsilon}} Q(x^\epsilon, y^\epsilon) dW_t, & y_0^\epsilon = y_0, \end{cases}$$

   where g is a $(d-p)$-dimensional vector field; f is a p-dimensional vector field; σ is a $((d-p) \times d)$-dimensional matrix field; Q is a $(p \times d)$-dimensional matrix field; and W_t is a d-dimensional Brownian motion.
3. *Let Y_t be the solution to*

$$(3.3) \qquad dY_t = f(x_0, Y_t) \, dt + Q(x_0, Y_t) \, dW_t, \qquad Y_0 = y_0;$$

   there exists a family of probability measures $\mu(x, dy)$ on $\mathbb{R}^p$ indexed by $x \in \mathbb{R}^{d-p}$ and a positive function $T \mapsto E(T)$ such that $\lim_{T \to \infty} E(T) = 0$ and such that for all x_0, y_0, T, and ϕ with uniformly bounded C^r derivatives for $r \leq 3$

$$(3.4) \qquad \left| \frac{1}{T} \int_0^T \mathbb{E}\big[\phi(Y_s)\big] - \int \phi(y)\mu(x_0, dy) \right| \leq \chi\big(\|(x_0, y_0)\|\big) E(T) \max_{r \leq 3} \|\phi\|_{C^r},$$

   where $r \mapsto \chi(r)$ is bounded on compact sets.
4. *For all u_0, $T > 0$, $\sup_{0 \leq t \leq T} \mathbb{E}\big[\chi\big(\|u_t^\epsilon\|\big)\big]$ is uniformly bounded in ϵ.*

REMARK 3.1. *As in the proof of Theorem 1.2, the uniform regularity of F, G, H, and K can be relaxed to local regularity by adding a control on the rate of escape*

of the process towards infinity. To simplify the presentation, we will use the global uniform regularity.

We will now extend the definition of two-scale flow convergence introduced in subsection 1.2 to stochastic processes.

3.1. Two-scale flow convergence for SDEs. Let $\left(\xi^\epsilon(t, \omega)\right)_{t \in \mathbb{R}^+, \omega \in \Omega}$ be a sequence of stochastic processes on $\mathbb{R}^d$ (progressively measurable mappings from $\mathbb{R}^+ \times \Omega$ to $\mathbb{R}^d$) indexed by $\epsilon > 0$. Let $(X_t)_{t \in \mathbb{R}^+}$ be a (progressively measurable) stochastic process on $\mathbb{R}^{d-p}$ $(p \geq 0)$. Let $x \mapsto \nu(x, dz)$ be a function from $\mathbb{R}^{d-p}$ into the space of probability measures on $\mathbb{R}^d$.

DEFINITION 3.1. *We say that the process ξ_t^ϵ F-converges to $\nu(X_t, dz)$ as $\epsilon \downarrow 0$ and write $\xi_t^\epsilon \xrightarrow[\epsilon \to 0]{F} \nu(X_t, dz)$ if and only if for all functions φ bounded and uniformly Lipschitz continuous on $\mathbb{R}^d$, and for all $t > 0$,*

$$(3.5) \qquad \lim_{h \to 0} \lim_{\epsilon \to 0} \frac{1}{h} \int_t^{t+h} \mathbb{E}\left[\varphi(\xi_s^\epsilon)\right] ds = \mathbb{E}\left[\int_{\mathbb{R}^d} \varphi(z) \nu(X_t, dz)\right].$$

3.2. Nonintrusive FLAVORS for SDEs. Let ω be a random sample from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Phi_h^\alpha(\cdot, \omega)$ be a random mapping from $\mathbb{R}^d$ onto $\mathbb{R}^d$ approximating the flow of (3.1) for $\alpha = 1/\epsilon$. If the parameter α can be controlled, then Φ_h^α can be used as a black box for accelerating the computation of solutions of (3.1) without prior identification of slow variables. Indeed, assume that there exist a constant $h_0 > 0$ and a normal random vector $\xi(\omega)$ such that for $h \leq h_0 \min(\frac{1}{\alpha}, 1)$
(3.6)

$$\left(\mathbb{E}\left[\left|\Phi_h^\alpha(u, \omega) - u - hG(u) - \alpha h F(u) - \sqrt{h} H(u)\xi(\omega) - \sqrt{\alpha h} K(u)\xi(\omega)\right|^2\right]\right)^{\frac{1}{2}} \leq C h^{\frac{3}{2}}(1+\alpha)^{\frac{3}{2}};$$

then FLAVOR can be defined as the algorithm simulating the stochastic process

$$(3.7) \qquad \begin{cases} \bar{u}_0 = u_0, \\ \bar{u}_{(k+1)\delta} = \Phi_{\delta-\tau}^0(., \omega_k') \circ \Phi_\tau^{\frac{1}{\epsilon}}(\bar{u}_{k\delta}, \omega_k), \\ \bar{u}_t = \bar{u}_{k\delta} \quad \text{for} \quad k\delta \leq t < (k+1)\delta, \end{cases}$$

where ω_k, ω_k' are independent and identically distributed (i.i.d.) samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\delta \leq h_0$, and $\tau \in (0, \delta)$ such that $\tau \leq \tau_0 \epsilon$. Theorem 3.2 establishes the asymptotic accuracy of FLAVOR for $\tau \ll \epsilon \ll \delta$ and

$$(3.8) \qquad \left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}} \ll \delta \ll \frac{\tau}{\epsilon}.$$

3.3. Convergence theorem.
THEOREM 3.2. *Let u^ϵ be the solution to (3.1), and let $\bar{u}_t$ be defined by (3.7). Assume that (3.6) and Condition 3.1 are satisfied; then the following hold:*
- *u_t^ϵ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy)\right)$ as $\epsilon \downarrow 0$, where X_t is the solution to*

$$(3.9) \qquad dX_t = \int g(X_t, y) \mu(X_t, dy) dt + \bar{\sigma}(X_t) dB_t, \qquad X_0 = x_0,$$

where $\bar{\sigma}$ is a $(d-p) \times (d-p)$ matrix field defined by

$$(3.10) \qquad \bar{\sigma}\bar{\sigma}^T = \int \sigma\sigma^T(x,y)\,\mu(x,dy)$$

and B_t is a $(d-p)$-dimensional Brownian motion.

- $\bar{u}_t$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t,dy)\right)$ as $\epsilon \downarrow 0$, $\tau \leq \delta$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\delta\epsilon}{\tau} \downarrow 0$, and $\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\frac{1}{\delta} \downarrow 0$.

The proof of convergence of SDEs of type (3.2) is classical, and a comprehensive monograph can be found in Chapter II of [108]. A proof of (mean squared) convergence of HMM applied to (3.2) (separated slow and fast variables) with $\sigma = 0$ has been obtained in [38]. A proof of (mean squared) convergence of the equation-free method applied to (3.2) with $\sigma \neq 0$ but independent of fast variables has been obtained in [53]. Theorem 3.2 proves the convergence in distribution of FLAVOR applied to SDE (3.1) with hidden slow and fast processes. One of the main difficulties of the proof of Theorem 3.2 lies in the fact that we are not assuming that the noise on (hidden) slow variables is null or independent from fast variables. Without this assumption, x_t^ϵ converges only weakly towards X_t, the convergence of u^ϵ can be only weak, and techniques for strong convergence cannot be used. The proof of Theorem 3.2 relies on a powerful result by Skorokhod (Theorem 1 of Chapter II of [108]) stating that the convergence in distribution of a sequence of stochastic processes is implied by the convergence of their generators. We refer the reader to subsection A.2 of the appendix for the detailed proof of Theorem 3.2.

3.4. Natural FLAVORS. As for ODEs, it is not necessary to use legacy integrators to obtain FLAVORS for SDEs. More precisely, Theorem 3.2 remains valid if FLAVORS are defined to be algorithms simulating the discrete process

$$(3.11) \qquad \begin{cases} \bar{u}_0 = u_0, \\ \bar{u}_{(k+1)\delta} = \theta_{\delta-\tau}^G(.,\omega_k') \circ \theta_\tau^\epsilon(\bar{u}_{k\delta},\omega_k), \\ \bar{u}_t = \bar{u}_{k\delta} \quad \text{for} \quad k\delta \leq t < (k+1)\delta, \end{cases}$$

where ω_k, ω_k' are i.i.d. samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and θ_τ^ϵ and $\theta_{\delta-\tau}^G$ are two random mappings from $\mathbb{R}^d$ onto $\mathbb{R}^d$ satisfying Condition 3.2. More precisely, $\theta_\tau^\epsilon(.,\omega)$ approximates in distribution the flow of (3.1) over time steps $\tau \ll \epsilon$. $\theta_h^G(.,\omega)$ approximates in distribution the flow of

$$(3.12) \qquad dv_t^\epsilon = G(v_t^\epsilon)\,dt + H(v_t^\epsilon)\,dW_t$$

over time steps $h \ll 1$.

CONDITION 3.2. *Assume that the following hold:*

1. *There exist h_0, $C > 0$, and a d-dimensional centered Gaussian vector $\xi(\omega)$ with identity covariance matrix such that for $h \leq h_0$*

$$(3.13) \qquad \left(\mathbb{E}\left[\left|\theta_h^G(u,\omega) - u - hG(u) - \sqrt{h}H(u)\xi(\omega)\right|^2\right]\right)^{\frac{1}{2}} \leq Ch^{\frac{3}{2}}.$$

2. *There exist τ_0, $C > 0$, and a d-dimensional centered Gaussian vector $\xi(\omega)$ with identity covariance matrix such that for $\frac{\tau}{\epsilon} \leq \tau_0$*

(3.14)

$$\left(\mathbb{E}\left[\left|\theta_\tau^\epsilon(u,\omega) - u - \tau G(u) - \frac{\tau}{\epsilon}F(u) - \sqrt{\tau}H(u)\xi(\omega) - \sqrt{\frac{\tau}{\epsilon}}K(u)\xi(\omega)\right|^2\right]\right)^{\frac{1}{2}}$$

$$\leq C\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}.$$

3. *For all u_0, $T > 0$, $\sup_{0 \leq n \leq T/\delta} \mathbb{E}\big[\chi\big(\|\bar{u}_{n\delta}\|\big)\big]$ is uniformly bounded in ϵ, $0 < \delta \leq h_0$, and $\tau \leq \min(\tau_0\epsilon, \delta)$, where $\bar{u}$ is defined by (3.11).*

3.5. FLAVORS for generic stiff SDEs. FLAVORS for stochastic systems have a natural generalization to SDEs on $\mathbb{R}^d$ of the form

(3.15) $$du^{\alpha,\epsilon} = F(u^{\alpha,\epsilon}, \alpha, \epsilon)\, dt + K(u^{\alpha,\epsilon}, \alpha, \epsilon)\, dW_t,$$

where $(W_t)_{t \geq 0}$ is a d-dimensional Brownian motion, and F and K are Lipschitz continuous in u.

CONDITION 3.3. *Assume that the following hold:*

1. *$\gamma \mapsto F(u, \alpha, \gamma)$ and $\gamma \mapsto K(u, \alpha, \gamma)$ are uniformly continuous in the neighborhood of 0.*

2. *There exists a diffeomorphism $\eta := (\eta^x, \eta^y)$, from $\mathbb{R}^d$ onto $\mathbb{R}^{d-p} \times \mathbb{R}^p$, independent from ϵ, α, with uniformly bounded C^1, C^2, and C^3 derivatives, such that the stochastic process $(x_t^\alpha, y_t^\alpha) = (\eta^x(u_t^{\alpha,0}), \eta^y(u_t^{\alpha,0}))$ satisfies for all $\alpha \geq 1$ the SDE*

   (3.16) $$dx^\alpha = g(x^\alpha, y^\alpha)\, dt + \sigma(x^\alpha, y^\alpha)\, dW_t, \qquad x_0^\alpha = x_0,$$

   where g is a $(d-p)$-dimensional vector field, σ is a $((d-p) \times d)$-dimensional matrix field, and g and σ are uniformly bounded and Lipschitz continuous in x and y.

3. *There exists a family of probability measures $\mu(x, dy)$ on $\mathbb{R}^p$ such that for all x_0, y_0, T $\big((x_0, y_0) := \eta(u_0)\big)$, and φ with uniformly bounded C^r derivatives for $r \leq 3$*

   $$\left| \frac{1}{T}\int_0^T \mathbb{E}\big[\varphi(y_s^\alpha)\big]\, ds - \int \varphi(y)\mu(x_0, dy) \right|$$
   (3.17) $$\leq \chi\big(\|(x_0, y_0)\|\big)\big(E_1(T) + E_2(T\alpha^\nu)\big) \max_{r \leq 3} \|\varphi\|_{C^r},$$

   where $r \mapsto \chi(r)$ is bounded on compact sets and $E_2(r) \to 0$ as $r \to \infty$ and $E_1(r) \to 0$ as $r \to 0$.

4. *For all u_0, $T > 0$, $\sup_{0 \leq t \leq T} \mathbb{E}\big[\chi\big(\|u_t^{\alpha,0}\|\big)\big]$ is uniformly bounded in $\alpha \geq 1$.*

REMARK 3.2. *As in the proof of Theorem 1.2, the uniform regularity of g and σ can be relaxed to local regularity by adding a control on the rate of escape of the process towards infinity. To simplify the presentation, we have used the global uniform regularity.*

Let ω be a random sample from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Phi_h^{\alpha,\epsilon}(., \omega)$ be a random mapping from $\mathbb{R}^d$ onto $\mathbb{R}^d$ approximating in distribution the flow of (3.15) over time steps $\tau \ll \epsilon$. If the parameter α can be controlled, then $\Phi_h^{\alpha,\epsilon}$ can be used as a black box for accelerating the computation of solutions of (3.15). The acceleration is obtained without prior identification of the slow variables.

CONDITION 3.4. *Assume that the following hold:*

1. *There exist h_0, $C > 0$, and a d-dimensional centered Gaussian vector $\xi(\omega)$ with identity covariance matrix such that for $h \leq h_0$, $0 < \epsilon \leq 1 \leq \alpha$, and $h \leq h_0 \min(\frac{1}{\alpha^\nu}, 1)$*

(3.18)

$$\left(\mathbb{E}\left[\left| \Phi_h^{\alpha,\epsilon}(u) - u - hF(u, \alpha, \epsilon) - \sqrt{h}\xi(\omega)K(u, \alpha, \epsilon) \right|^2 \right] \right)^{\frac{1}{2}} \leq Ch^{\frac{3}{2}}(1 + \alpha^{\frac{3\nu}{2}}).$$

2. *For all u_0, $T > 0$, $\sup_{0 \leq n \leq T/\delta} \mathbb{E}\left[\chi\left(\|\bar{u}_{n\delta}\|\right)\right]$ is uniformly bounded in ϵ, $0 < \delta \leq h_0$, and $\tau \leq \min(h_0\epsilon^\nu, \bar{\delta})$, where $\bar{u}$ is defined by (3.19).*

FLAVORS. Let $\delta \leq h_0$ and $\tau \in (0, \delta)$ such that $\tau \leq \tau_0\epsilon^\nu$. We define FLAVORS as the class of algorithms simulating the stochastic process $t \mapsto \bar{u}_t$ defined by

(3.19)
$$\begin{cases} \bar{u}_0 = u_0, \\ \bar{u}_{(k+1)\delta} = \Phi_{\delta-\tau}^{0,\epsilon}(., \omega_k') \circ \Phi_\tau^{\frac{1}{\epsilon},\epsilon}(\bar{u}_{k\delta}, \omega_k), \\ \bar{u}_t = \bar{u}_{k\delta} \quad \text{for} \quad k\delta \leq t < (k+1)\delta, \end{cases}$$

where ω_k, ω_k' are i.i.d. samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

REMARK 3.3. *ω_k simulates the randomness of the increment of the Brownian motion between times δk and $\delta k + \tau$. ω_k' simulates the randomness of the increment of the Brownian motion between times $\delta k + \tau$ and $\delta(k+1)$. The independence of ω_k and ω_k' is a reflection of the independence of the increments of a Brownian motion.*

The following theorem shows that the flow averaging integrator is accurate with respect to F-convergence for $\tau \ll \epsilon^\nu \ll \delta$ and

(3.20)
$$\left(\frac{\tau}{\epsilon^\nu}\right)^{\frac{3}{2}} \ll \delta \ll \frac{\tau}{\epsilon^\nu}.$$

THEOREM 3.3. *Let $u_t^{\frac{1}{\epsilon},\epsilon}$ be the solution to (3.15) with $\alpha = 1/\epsilon$, and let $\bar{u}_t$ be defined by (3.19). Assume that Conditions 3.3 and 3.4 are satisfied; then the following hold:*

- *$u_t^{\frac{1}{\epsilon},\epsilon}$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy)\right)$ as $\epsilon \downarrow 0$, where X_t is the solution to*

(3.21)
$$dX_t = \int g(X_t, y)\, \mu(X_t, dy) + \bar{\sigma}(X_t)\, dB_t, \qquad X_0 = x_0,$$

where $\bar{\sigma}$ is a $(d-p) \times (d-p)$ matrix field defined by

(3.22)
$$\bar{\sigma}\bar{\sigma}^T = \int \sigma\sigma^T(x, y)\, \mu(x, dy)$$

and B_t is a $(d-p)$-dimensional Brownian motion.

- *As $\epsilon \downarrow 0$, $\tau\epsilon^{-\nu} \downarrow 0$, $\delta\frac{\epsilon^\nu}{\tau} \downarrow 0$, and $\left(\frac{\tau}{\epsilon^\nu}\right)^{\frac{3}{2}}\frac{1}{\delta} \downarrow 0$, the numerical solution $\bar{u}_t$ F-converges towards $\eta^{-1} * \left(\delta_{X_t} \otimes \mu(X_t, dy)\right)$ as $\epsilon \downarrow 0$, where X_t is the solution to (3.21).*

Proof. The proof of Theorem 3.3 is similar to the proof of Theorem 3.2. The condition $\epsilon \ll 1$ is needed for the approximation of $u^{\alpha,\epsilon}$ by $u^{\alpha,0}$ and for the F-convergence of $u^{\frac{1}{\epsilon},0}$. Since $y_t^\alpha = \eta^y(u_t^{\alpha,0})$, the condition $\tau \ll \epsilon^\nu$ is used along with (3.18) for the accuracy of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ in (locally) approximating y_t^α. The condition $\delta \ll \frac{\tau}{\epsilon^\nu}$

allows for the averaging of g and σ to take place prior to a significant change of $x\alpha_t$; more precisely, it allows for $m \gg 1$ iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$ prior to a significant change of $x\alpha_t$. The condition $\left(\frac{\tau}{\epsilon^\nu}\right)^{\frac{3}{2}} \ll \delta$ is required in order to control the error accumulated by m iterations of $\Phi_\tau^{\frac{1}{\epsilon},\epsilon}$. $\square$

4. Stochastic mechanical systems: Langevin equations. Since the foundational work of Bismut [12], the field of stochastic geometric mechanics has grown in response to the demand for tools to analyze the structure of continuous and discrete mechanical systems with uncertainty [107, 60, 115, 30, 87, 86, 89, 72, 81, 18, 17, 19]. Like their deterministic counterparts, these integrators are structure preserving in terms of statistical invariants.

In this section, FLAVORS are developed to be structure preserving integrators for stiff stochastic mechanical systems, i.e., stiff Langevin equations of the form

$$(4.1) \qquad \begin{cases} dq = M^{-1}p, \\ dp = -\nabla V(q)\,dt - \frac{1}{\epsilon}\nabla U(q)\,dt - cp\,dt + \sqrt{2\beta^{-1}}c^{\frac{1}{2}}\,dW_t \end{cases}$$

and of the form

$$(4.2) \qquad \begin{cases} dq = M^{-1}p, \\ dp = -\nabla V(q)\,dt - \frac{1}{\epsilon}\nabla U(q)\,dt - \frac{c}{\epsilon}p\,dt + \sqrt{2\beta^{-1}}\frac{c^{\frac{1}{2}}}{\sqrt{\epsilon}}\,dW_t, \end{cases}$$

where c is a positive symmetric $d \times d$ matrix.

REMARK 4.1. *Provided that hidden fast variables remain locally ergodic, one can also consider Hamiltonians with a mixture of both slow and fast noise and friction. For the sake of clarity, we have restricted our presentation to (4.1) and (4.2).*

Equations (4.1) and (4.2) model a mechanical system with Hamiltonian

$$(4.3) \qquad H(q,p) := \frac{1}{2}p^T M^{-1} p + V(q) + \frac{1}{\epsilon}U(q).$$

The phase space is the Euclidean space $\mathbb{R}^d \times \mathbb{R}^d$ or a cotangent bundle $T^*\mathcal{M}$ of a configuration manifold $\mathcal{M}$.

REMARK 4.2. *If c is not constant and $\mathcal{M}$ is not the usual $\mathbb{R}^d \times \mathbb{R}^d$ Euclidean space, one should use the Stratonovich integral instead of the Itô integral.*

4.1. FLAVORS for stochastic mechanical systems on manifolds. As in section 2, we assume that we are given a mapping Φ_h^α acting on the phase space such that for $h \le h_0 \min(1, \alpha^{-\frac{1}{2}})$

$$(4.4) \qquad \left| \Phi_h^\alpha(q,p) - (q,p) - h\left(M^{-1}p, -V(q) - \alpha U(q)\right) \right| \le Ch^2(1+\alpha).$$

Next, consider the following Ornstein–Uhlenbeck equation:

$$(4.5) \qquad dp = -\alpha c p\,dt + \sqrt{\alpha}\sqrt{2\beta^{-1}}c^{\frac{1}{2}}\,dW_t.$$

The stochastic flow of (4.5) is defined by the following stochastic evolution map:

$$(4.6) \qquad \Psi_{t_1,t_2}^\alpha(q,p) = \left(q, e^{-c\alpha(t_2-t_1)}p + \sqrt{2\beta^{-1}\alpha}c^{\frac{1}{2}}\int_{t_1}^{t_2} e^{-c\alpha(t_2-s)}\,dW_s\right).$$

Let $\delta \leq h_0$ and $\tau \in (0, \delta)$ such that $\tau \leq \tau_0/\sqrt{\alpha}$. FLAVOR for (4.1) can then be defined by

$$
(4.7) \qquad \begin{cases} (\bar{q}_0, \bar{p}_0) = (q_0, p_0), \\ (\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Phi^0_{\delta-\tau} \circ \Psi^1_{k\delta+\tau,(k+1)\delta} \circ \Phi^{\frac{1}{\epsilon}}_{\tau} \circ \Psi^1_{k\delta,k\delta+\tau}(\bar{q}_{k\delta}, \bar{p}_{k\delta}) \end{cases}
$$

and FLAVOR for (4.2) can be defined by

$$
(4.8) \qquad \begin{cases} (\bar{q}_0, \bar{p}_0) = (q_0, p_0), \\ (\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Phi^0_{\delta-\tau} \circ \Phi^{\frac{1}{\epsilon}}_{\tau} \circ \Psi^{\frac{1}{\epsilon}}_{k\delta,k\delta+\tau}(\bar{q}_{k\delta}, \bar{p}_{k\delta}). \end{cases}
$$

Theorem 3.3 establishes the accuracy of these integrators under Conditions 3.3 and 3.4 provided that $\tau \ll \sqrt{\epsilon} \ll \delta$ and $\left(\frac{\tau}{\sqrt{\epsilon}}\right)^{\frac{3}{2}} \ll \delta \ll \frac{\tau}{\sqrt{\epsilon}}$.

4.2. Structure preserving properties of FLAVORS for stochastic mechanical systems on manifolds. First, observe that if Φ^α_h and $\Psi^{\frac{1}{\epsilon}}_h$ are symmetric under a group action for all $\epsilon > 0$, then the resulting FLAVOR, as a symmetric composition of symmetric steps, is symmetric under the same group action (see the comment below Theorem 2.4).

Similarly, the following theorem shows that FLAVORS inherit structure preserving properties from those associated with Φ^α_h (the component approximating the Hamiltonian part of the flow).

THEOREM 4.1.
- *If Φ^α_h is symplectic, then the FLAVORS defined by (4.7) and (4.8) are quasi-symplectic as defined in Conditions RL1 and RL2 of [88] (it degenerates to a symplectic method if friction is set equal to zero and the Jacobian of the flow map is independent of (q, p)).*
- *If, in addition, c is isotropic, then FLAVOR defined by (4.7) is conformally symplectic; i.e., it preserves the precise symplectic area change associated with the flow of inertial Langevin processes [84].*

Proof. Those properties are a consequence of the fact that FLAVORS are splitting schemes. The quasi symplecticity and symplectic conformality of the geometric Langevin algorithm (GLA) has been obtained in a similar way in [17]. □

4.2.1. Example of quasi-symplectic FLAVORS. An example of a quasi-symplectic FLAVOR can be obtained by choosing Φ^α_h to be the symplectic Euler integrator defined by (2.9). This integrator is also conformally symplectic if c is isotropic and friction is slow.

4.2.2. Example of quasi-symplectic and time-reversible FLAVORS. Defining Φ^α_h by (2.9) and $\Phi^{\alpha,*}_h$ by (2.10), an example of a quasi-symplectic and time-reversible FLAVOR can be obtained by using the symmetric Strang splitting:

$$
(4.9) \quad (\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Psi^1_{k\delta+\frac{\delta}{2},(k+1)\delta} \circ \Phi^{\frac{1}{\epsilon},*}_{\frac{\tau}{2}} \circ \Phi^{0,*}_{\frac{\delta-\tau}{2}} \circ \Phi^0_{\frac{\delta-\tau}{2}} \circ \Phi^{\frac{1}{\epsilon}}_{\frac{\tau}{2}} \circ \Psi^1_{k\delta,k\delta+\frac{\delta}{2}}(q, p)
$$

for (4.1) and

$$
(4.10) \quad (\bar{q}_{(k+1)\delta}, \bar{p}_{(k+1)\delta}) = \Psi^{\frac{1}{\epsilon}}_{(k+1)\delta-\frac{\tau}{2},(k+1)\delta} \circ \Phi^{\frac{1}{\epsilon},*}_{\frac{\tau}{2}} \circ \Phi^{0,*}_{\frac{\delta-\tau}{2}} \circ \Phi^0_{\frac{\delta-\tau}{2}} \circ \Phi^{\frac{1}{\epsilon}}_{\frac{\tau}{2}} \circ \Psi^{\frac{1}{\epsilon}}_{k\delta,k\delta+\frac{\tau}{2}}(q, p)
$$

for (4.2). This integrator is also conformally symplectic if c is isotropic and friction is slow.

4.2.3. Example of Boltzmann–Gibbs reversible Metropolis-adjusted FLAVORS. Since the probability density of Ψ_{t_1,t_2} can be computed explicitly, it follows that the probability densities of (4.9) and (4.10) can be computed explicitly, and these algorithms can be metropolized and made reversible with respect to the Gibbs distribution, as has been shown in [19] for GLA introduced in [17]. This metropolization leads to stochastically stable (and ergodic if the noise applied on momentum is not degenerate) algorithms. We refer the reader to [19] for details. Observe that if the proposed move is rejected, the momentum has to be flipped and the acceptance probability involves a momentum flip. It is proved in [19] that GLA [17] remains strongly accurate after a metropolization involving local momentum flips. Whether this preservation of accuracy over trajectories transfers in a weak sense (in distributions) to FLAVORS remains to be investigated.

5. Numerical analysis of FLAVOR based on VE.

5.1. Stability. Consider the following linear Hamiltonian system:

$$(5.1) \qquad H(x, y, p_x, p_y) = \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}x^2 + \frac{\omega^2}{2}(y - x)^2$$

with $\omega \gg 1$. Here, $\frac{x+y}{2}$ is the slow variable and $y - x$ is the fast variable.

It can be shown that, when applied to (5.1), symplectic Euler (2.9) is stable if and only if $h \leq \sqrt{2}/\omega$. Write $\Theta_{\delta,\tau}$ as the nonintrusive FLAVOR (2.5) obtained by using symplectic Euler (2.9) as a legacy integrator. Write $\Theta_{\delta,\tau}^a$ as the artificial FLAVOR described in subsection 2.1.4.

THEOREM 5.1. *The nonintrusive FLAVOR $\Theta_{\delta,\tau}$ with $1/\sqrt{\tau} \gg \omega \gg 1$ is stable if and only if $\delta \in (0, 2)$. The artificial FLAVOR $\Theta_{\delta,\tau}^a$ with $1/\tau \gg \omega \gg 1$ is stable if and only if $\delta \in (0, 2\sqrt{2})$.*

Proof. The numerical scheme associated with $\Theta_{\delta,\tau}$ can be written as

$$(5.2) \qquad \begin{bmatrix} y_{n+1} \\ x_{n+1} \\ (p_y)_{n+1} \\ (p_x)_{n+1} \end{bmatrix} = T \begin{bmatrix} y_n \\ x_n \\ (p_y)_n \\ (p_x)_n \end{bmatrix}$$

with

$$T = \begin{bmatrix} 1 & 0 & \delta - \tau & 0 \\ 0 & 1 & 0 & \delta - \tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \tau - \delta & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \tau & 0 \\ 0 & 1 & 0 & \tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\tau(\omega^2 + 1) & \tau\omega^2 & 1 & 0 \\ \tau\omega^2 & -\tau\omega^2 & 0 & 1 \end{bmatrix}.$$

The characteristic polynomial of T is

$$(5.3)$$
$$\lambda^4 + (-4 + \delta^2 - \delta^2\tau^2 + 2\delta\tau^3 - \tau^4 + 2\delta\tau\omega^2 - \delta^2\tau^2\omega^2 + 2\delta\tau^3\omega^2 - \tau^4\omega^2)\lambda^3 + (6 - 2\delta^2$$
$$+ 2\delta^2\tau^2 - 4\delta\tau^3 + 2\tau^4 - 4\delta\tau\omega^2 + \delta^3\tau\omega^2 + 2\delta^2\tau^2\omega^2 - 4\delta\tau^3\omega^2 - \delta^3\tau^3\omega^2 + 2\tau^4\omega^2$$
$$+ 2\delta^2\tau^4\omega^2 - \delta\tau^5\omega^2)\lambda^2$$
$$+ (-4 + \delta^2 - \delta^2\tau^2 + 2\delta\tau^3 - \tau^4 + 2\delta\tau\omega^2 - \delta^2\tau^2\omega^2 + 2\delta\tau^3\omega^2 - \tau^4\omega^2)\lambda + 1.$$

Since $\omega \gg 1$, $\tau \ll 1/\omega^2$, as long as $\delta \lesssim 1$, roots to the above polynomial are (by continuity; we refer the reader, for instance, to Theorem 1 of [34]) close to roots to
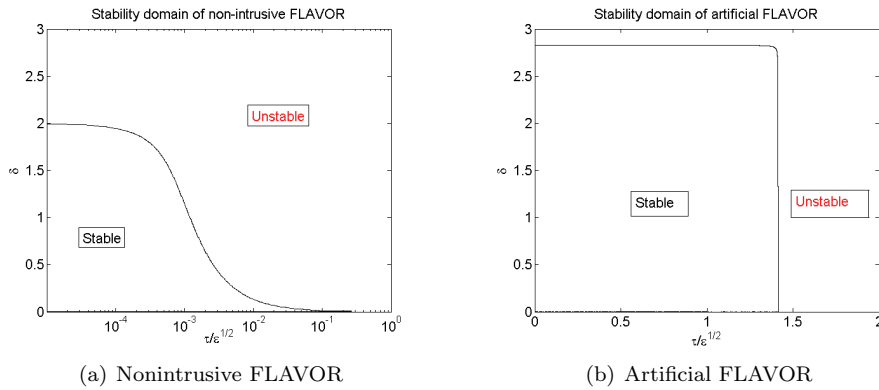
(a) Nonintrusive FLAVOR

(b) Artificial FLAVOR

FIG. 5.1. *Stability domain of the nonintrusive and artificial FLAVORS applied to* (5.1) *as a function of δ and τ/ϵ. $\omega = 1/\sqrt{\epsilon} = 1000$.*

the asymptotic polynomial

$$\lambda^4 + (\delta^2 - 4)\lambda^3 + (6 - 2\delta^2)\lambda^2 + (\delta^2 - 4)\lambda + 1, \tag{5.4}$$

which can be shown to be 1 with multiplicity 2 and $\frac{1}{2}(2 - \delta^2 \pm \delta\sqrt{\delta^2 - 4})$. It is easy to see that all roots are complex numbers with moduli less than or equal to one if and only if $|\delta| \leq 2$.

The numerical scheme associated with $\Theta^a_{\delta,\tau}$ can be written as in (5.2) with

$$T = \begin{bmatrix} 1 & 0 & \frac{\delta-\tau}{2} & \frac{\delta-\tau}{2} \\ 0 & 1 & \frac{\delta-\tau}{2} & \frac{\delta-\tau}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\tau\omega^2 & \tau\omega^2 & 1 & 0 \\ \tau\omega^2 & -\tau\omega^2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \tau & 0 \\ 0 & 1 & 0 & \tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\delta & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.5}$$

The characteristic polynomial of T is

$$2\lambda^4 + (4\omega^2\tau^2 + \tau\delta + \delta^2 - 8)\lambda^3 + (12 - 2\delta^2 - 2\delta\tau - 8\tau^2\omega^2 + 2\delta^2\tau^2\omega^2)\lambda^2$$
$$+ (4\omega^2\tau^2 + \tau\delta + \delta^2 - 8)\lambda + 2. \tag{5.6}$$

Similarly, since $\omega \gg 1$, $\tau \ll 1/\omega$, as long as $\delta \lesssim 1$, roots to the above polynomial are close to roots to the asymptotic polynomial

$$2\lambda^4 + (\delta^2 - 8)\lambda^3 + (12 - 2\delta^2)\lambda^2 + (\delta^2 - 8)\lambda + 1, \tag{5.7}$$

which can be shown to be 1 with multiplicity 2 and $\frac{1}{4}(4 - \delta^2 \pm \delta\sqrt{\delta^2 - 8})$. All roots are complex numbers with moduli less than or equal to one if and only if $|\delta| \leq 2\sqrt{2}$. ☐

Figures 5.1(a) and 5.1(b) illustrate the domain of stability of the nonintrusive FLAVOR (based on symplectic Euler (2.5) and (2.9)) and artificial FLAVOR (2.11) applied to the flow of (5.1), i.e., values of δ and τ/ϵ ensuring stable numerical integrations. We observe that the artificial FLAVOR has a much larger stability domain than the nonintrusive FLAVOR. Specifically, for the nonintrusive FLAVOR and large values of δ, $\tau = o(\sqrt{\epsilon})$ is not enough and one needs $\tau = o(\epsilon)$ for a stable integration, whereas the artificial FLAVOR requires only $\tau = \sqrt{2\epsilon}$, a minimum requirement for a stable symplectic Euler integration of the fast dynamics.

Notice that there is no resonance behavior in terms of stability; everything below the two curves is stable and everything outside is not stable (plots not shown).

5.2. Error analysis. The flow of (5.1) has been explicitly computed and compared with solutions obtained from the nonintrusive FLAVOR based on symplectic Euler ((2.5) and (2.9)) and with artificial FLAVOR (2.11).

The total simulation time is $T = 10$, and absolute errors on the slow variable have been computed with respect to the Euclidean norm of the difference in positions between analytical and numerical solutions. Stability is investigated using the same technique used in subsection 5.1. Figures 5.2(a) and 5.2(b) illustrate errors as functions of mesostep δ and renormalized small step τ/ϵ. Observe that the given δ errors are minimized at specific values of τ/ϵ for both integrators, but the accuracy of the nonintrusive FLAVOR is less sensitive to τ/ϵ. Figures 5.2(c) and 5.2(d) plot the optimal value of τ/ϵ as a function of δ and the associated error. Observe also that for the nonintrusive FLAVOR the dependence of the optimal value of τ/ϵ on δ is weak, whereas for the artificial FLAVOR the optimal value of τ/ϵ roughly scales linearly with δ. Figures 5.2(e) and 5.2(f) describe how error changes with small step τ for mesostep δ fixed. Figure 5.2(e) can be viewed in correspondence with the condition $\delta \ll \tau/\epsilon$ required for accuracy. This requirement, however, is just a sufficient condition to obtain an error bound, as we can see in Figure 5.2(f). There the weak dependence of the error on τ/ϵ for a fixed δ shows that one does not have to choose the microstep with too much care or optimize the integrator with respect to its value if the artificial FLAVOR is used. As a matter of fact, all the numerical experiments illustrated in this paper (except for Figures 5.2(c) and 5.2(d)) have been performed without any tuning of the value τ/ϵ. We have simply used the rule of thumb $\delta \sim \gamma \frac{\tau}{\epsilon}$, where γ is a small parameter (0.1, for instance).

Therefore, it appears that the benefits of artificial FLAVORS lie in their superior accuracy and stability.

Notice that there is no resonant value of δ or τ.

5.3. Numerical error analysis for nonlinear systems. In this subsection, we will consider the nonlinear Hamiltonian system

$$(5.8) \quad H(x, y, z, p_x, p_y, p_z) = \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}p_z^2 + x^4 + \epsilon^{-1}\frac{\omega_1}{2}(y-x)^2 + \epsilon^{-1}\frac{\omega_2}{2}(z-y)^2.$$

Thus, the potential is $U = \frac{\omega_1}{2}(y-x)^2 + \frac{\omega_2}{2}(z-y)^2$ and $V = x^4$. Here, $\frac{x+y+z}{3}$ acts as a slow degree of freedom and $y - x$ and $z - y$ act as fast degrees of freedom.

Figure 5.3 illustrates $t \mapsto \frac{x(t)+y(t)+z(t)}{3}$ (slow variable, convergent point-wisely) and $t \mapsto (y(t) - x(t), z(t) - y(t))$ (fast variables, convergent in measure) computed with symplectic Euler and with the induced symplectic FLAVOR (2.5). Define $q := (x, y, z)$. To illustrate the F-convergence property of FLAVOR, we fix $H = 1$, vary the mesostep $\delta = H/M$ by changing M, and show the Euclidean norm error of the difference between $\frac{1}{M}\sum_{i=0}^{M-1} q(T - ih/M)$ computed with FLAVOR and computed with symplectic Euler in Figure 5.4(a). Notice that, without an averaging over time length h, the error will be no longer monotonically but oscillatorily decreasing as δ changes (plots not shown) because fast variables are captured only in the sense of measure. As shown in Figure 5.4(a), the error scales linearly with $\frac{1}{M}$ for M not too small, and therefore the global error is a linear function of the mesostep δ and the method is first order convergent. Figure 5.4(b) shows that the error in general grows linearly with the total simulation time, and this linear growth of the error has
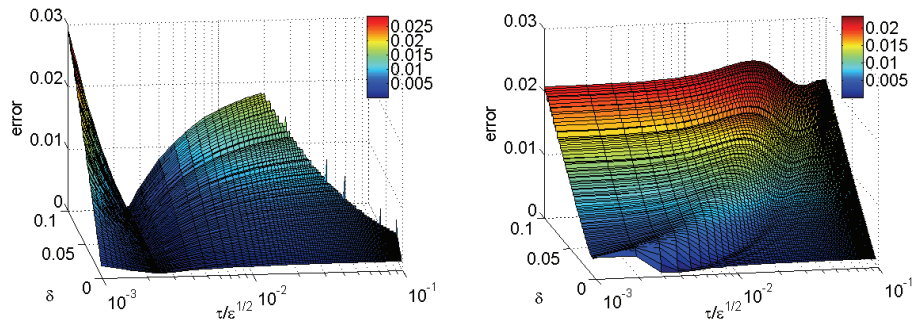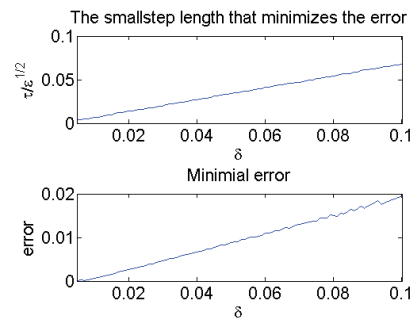
(a) Error of nonintrusive FLAVOR as a function of δ and $\tau/\sqrt{\epsilon}$. Notice that not all pairs of step lengths lead to stable integrations.
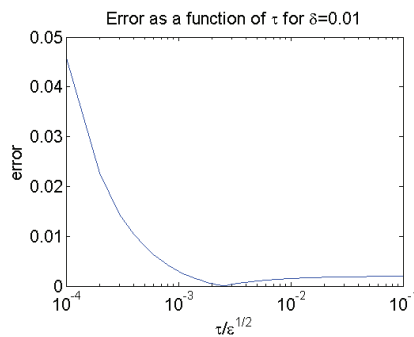

(b) Error of artificial FLAVOR as a function of δ and $\tau/\sqrt{\epsilon}$
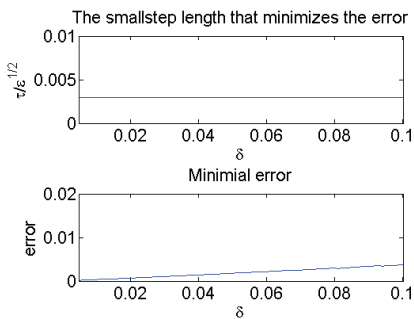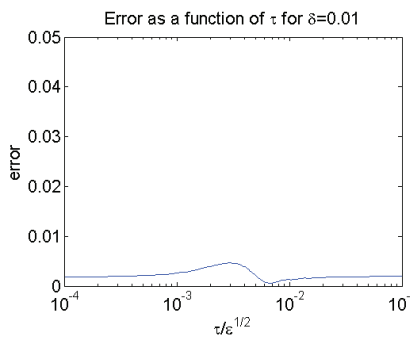

(c) Optimal $\tau/\sqrt{\epsilon}$ and error of nonintrusive FLAVOR as functions of δ

(d) Optimal $\tau/\sqrt{\epsilon}$ and error of artificial FLAVOR as functions of δ

(e) Error dependence on $\tau/\sqrt{\epsilon}$ for a given δ: nonintrusive FLAVOR

(f) Error dependence on $\tau/\sqrt{\epsilon}$ for a given δ: artificial FLAVOR

FIG. 5.2. *Error analysis of* (5.1). *Parameters are* $\omega = \sqrt{\epsilon} = 10^3$, *$x(0) = 0.8$, and $y(0) = x(0) + 1.1/\omega$.*

been observed for a simulation time larger than ω ($\epsilon^{-1/2}$). Figure 5.4(c) shows that the error does not depend on ω ($\epsilon^{-1/2}$) for a fixed δ, as long as ϵ is not too large (i.e., ω is not too small); the error is instead controlled by M. This is not caused by reaching the limit of machine accuracy; it is a characteristic of the method: the plateau for large ω corresponds to the complete scale separation regime of FLAVOR
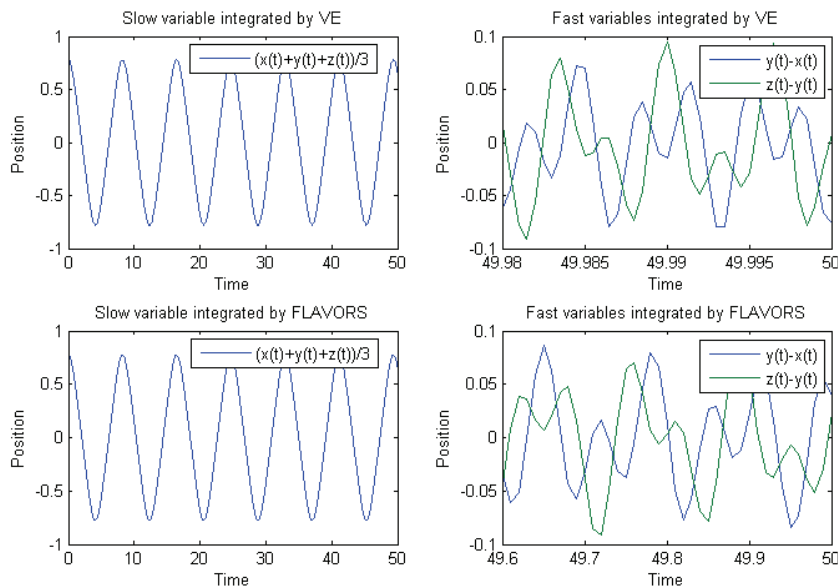
FIG. 5.3. *Comparison between trajectories integrated by VE and FLAVOR (defined by* (2.5) *and* (2.9)*). FLAVOR uses mesostep* $\delta = 0.01$ *and microstep* $\tau = 0.0005$, *and symplectic Euler uses time step* $\tau = 0.0005$. *Time axes in the right column are zoomed in (by different ratios) to illustrate the fact that fast variables are captured in the sense of measure. FLAVOR accelerated the computation by roughly* $20x$ *(*$\delta = 20\tau$*). In this experiment,* $\epsilon = 10^{-6}$, $\omega_1 = 1.1$, $\omega_2 = 0.97$, $x(0) = 0.8$, $y(0) = 0.811$, $z(0) = 0.721$, $p_x(0) = 0$, $p_y(0) = 0$, *and* $p_z(0) = 0$. *Simulation time is* $T = 50$.

(a) Asymptotically linear error dependence on $\delta = 1/M$

(b) Asymptotically linear error dependence on total simulation time T

(c) Asymptotically independent of the scaling factor ω
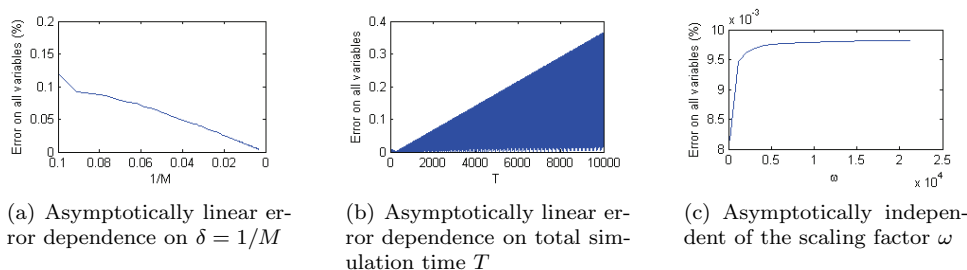
FIG. 5.4. *Error dependence on parameters in a FLAVOR simulation of* (5.8).

as a multiscale method.

Notice that there is no resonant value of δ in the sense of convergence.

The fact that the error scales linearly with total simulation time is a much stronger (numerical) result than our (theoretical) error analysis for FLAVORS (in which the error is bounded by a term growing exponentially with the total simulation time). We conjecture that the linear growth of the error is a consequence of the fact that FLAVOR is symplectic and is true only for a subclass of systems, possibly integrable systems. A rigorous analysis of the effects of the structure preservation of FLAVORS on long term behavior remains to be done.

6. Numerical experiments.

6.1. Hidden Van der Pol oscillator (ODE). Consider the following system of ODEs:

$$
(6.1) \quad \begin{cases} \dot{r} = \frac{1}{\epsilon}(r\cos\theta + r\sin\theta - \frac{1}{3}r^3\cos^3\theta)\cos\theta - \epsilon\, r\cos\theta\sin\theta, \\ \dot{\theta} = -\epsilon\cos^2\theta - \frac{1}{\epsilon}(\cos\theta + \sin\theta - \frac{1}{3}r^2\cos^3\theta)\sin\theta, \end{cases}
$$

where $\epsilon \ll 1$. Taking the transformation from polar coordinates to Cartesian coordinates by $[x, y] = [r\sin\theta, r\cos\theta]$ as the local diffeomorphism, we obtained the hidden system

$$
(6.2) \quad \begin{cases} \dot{x} = -\epsilon y, \\ \dot{y} = \frac{1}{\epsilon}(x + y - \frac{1}{3}y^3). \end{cases}
$$

Taking the second time derivative of y, the system can also be written as the second order ODE

$$
(6.3) \quad \ddot{y} + y = \frac{1}{\epsilon}(1 - y^2)\dot{y}.
$$

The latter is a classical Van der Pol oscillator [116]. The nonintrusive FLAVOR as defined by (1.34) can be directly applied to (6.1) (with hidden slow and fast processes) by turning the stiff parameter $\frac{1}{\epsilon}$ on and off. More precisely, defining $\Phi^{\epsilon,\alpha}(r, \theta)$ by

$$
(6.4) \quad \Phi_h^{\alpha,\epsilon}(r,\theta) := \begin{pmatrix} r \\ \theta \end{pmatrix} + \alpha h \begin{pmatrix} (r\cos\theta + r\sin\theta - \frac{1}{3}r^3\cos^3\theta)\cos\theta \\ -(\cos\theta + \sin\theta - \frac{1}{3}r^2\cos^3\theta)\sin\theta \end{pmatrix} - \epsilon h \begin{pmatrix} r\cos\theta\sin\theta \\ \cos^2\theta \end{pmatrix},
$$

FLAVOR is defined by (1.34) with $\bar{u} := (\bar{r}, \bar{\theta})$, i.e.,

$$
(6.5) \quad (\bar{r}_t, \bar{\theta}_t) = \left(\Phi_{\delta-\tau}^{0,\epsilon} \circ \Phi_\tau^{\frac{1}{\epsilon},\epsilon}\right)^k (r_0, \theta_0) \quad \text{for} \quad k\delta \le t < (k+1)\delta.
$$

We refer the reader to Figure 6.1 for a comparison of integrations by forward Euler, used as a benchmark, and FLAVORS. FLAVORS give trajectories close to forward Euler and correctly capture the $\mathcal{O}(\frac{1}{\epsilon})$ period [116] of the relaxation oscillation. Moreover, a 200x acceleration is achieved using FLAVOR.

6.2. Hamiltonian system with nonlinear stiff and soft potentials. In this subsection, we will apply the symplectic Euler FLAVOR defined by (2.5) and (2.9) to the mechanical system whose Hamiltonian is

$$
(6.6) \quad H(y, x, p_y, p_x) := \frac{1}{2}p_y^2 + \frac{1}{2}p_x^2 + \epsilon^{-1}y^6 + (x - y)^4.
$$

Here, stiff potential $\epsilon^{-1}U = \epsilon^{-1}y^6$ and soft potential $V = (x-y)^4$ are both nonlinear.

Figure 6.2 illustrates $t \mapsto y(t)$ (dominated by a fast process), $t \mapsto x(t) - y(t)$ (a slow process modulated by a fast process), and $t \mapsto H(t)$ computed with symplectic Euler, the induced symplectic FLAVOR ((2.5) and (2.9)), and IMEX [109]. Notice that $x - y$ is not a purely slow variable but contains some fast component, and therefore the FLAVOR integration of it contains a modulation of local oscillations, which could be interpreted as that fast component slowed down by FLAVOR. It is not easy to find a purely slow variable or a purely fast variable in the form of (1.2) for this example, but the integrated trajectory for such a slow variable will not contain these slowed-down local oscillations.
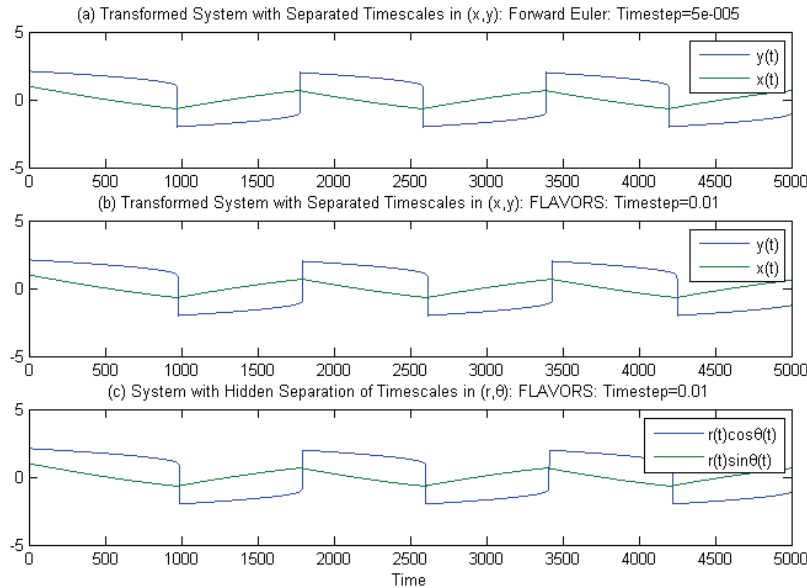
FIG. 6.1. *Over a timespan of $5/\epsilon$ are shown* (a) *the direct forward Euler simulation of* (6.2) *with time steps resolving the fast time scale,* (b) *the (nonintrusive* (1.34)*) FLAVOR simulation of* (6.2)*, and* (c) *the polar to Cartesian image of the (nonintrusive* (1.34)*) FLAVOR simulation of* (6.1) *with hidden slow and fast variables. Forward Euler uses time step* $h = 0.05\epsilon = 0.00005$. *The two FLAVOR simulations use* $\delta = 0.01$ *and* $\tau = 0.00005$. *Parameters are* $\frac{1}{\epsilon} = 1000$, $x(0) = 1$, *and* $y(0) = 1$.

6.3. Fermi–Pasta–Ulam problem. In this subsection, we will consider the Fermi–Pasta–Ulam (FPU) problem [43] illustrated by Figure 6.3 and associated with the Hamiltonian

$$(6.7) \qquad H(q,p) := \frac{1}{2} \sum_{i=1}^{m} (p_{2i-1}^2 + p_{2i}^2) + \frac{\omega^2}{4} \sum_{i=1}^{m} (q_{2i} - q_{2i-1})^2 + \sum_{i=0}^{m} (q_{2i+1} - q_{2i})^4.$$

The FPU problem is a well-known benchmark problem [83, 57] for multiscale integrators because it exhibits different behaviors over widely separated time scales. The stiff springs behave nearly like harmonic oscillators with period $\sim \mathcal{O}(\omega^{-1})$. Then the centers of masses linked by stiff springs (i.e., the midpoints of stiff springs) change over a time scale $\mathcal{O}(1)$. The third time scale, $\mathcal{O}(\omega)$, is associated with the rate of energy exchange between stiff springs. Energy exchange among stiff springs extends to even slower time scales, in either a periodic or a chaotic fashion [46, 45]. On the other hand, the total energy of the stiff springs behaves almost like a constant. This wide separation of time scales can be seen in Figures 6.4, 6.5, and 6.6, where four subplots address different scales: Subplot1 shows the fast variables $(q_{2i} - q_{2i-1})/\sqrt{2}$; Subplot2 shows one of the slow variables $(q_2 + q_1)/\sqrt{2}$; Subplot3 shows the energy transfer pattern among stiff springs, which is even slower; and Subplot4 shows the near-constant total energy of three stiff springs. All four subplots are time series. A comprehensive survey on geometric integration of the FPU problem can be found in [57].

Figures 6.4(a) and 6.4(b) compare symplectic Euler (with time steps fine enough to resolve FPU over the involved long time scale) and with the artificial FLAVOR
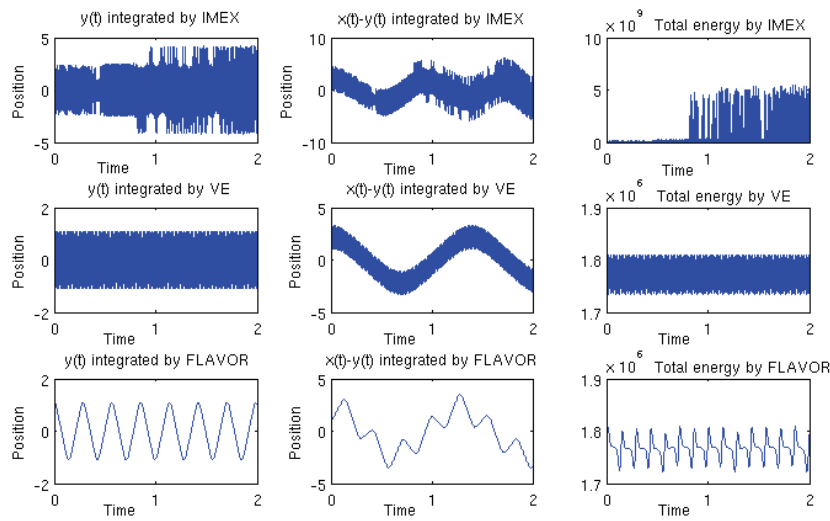
FIG. 6.2. *In this experiment, $\epsilon = 10^{-6}$, $y(0) = 1.1$, $x(0) = 2.2$, $p_y(0) = 0$, and $p_x(0) = 0$. Simulation time is $T = 2$. FLAVOR (defined by (2.5) and (2.9)) uses mesostep $\delta = 10^{-3}$ and microstep $\tau = 10^{-5}$, VE uses small time step $\tau = 10^{-5}$, and IMEX uses mesostep $\delta = 10^{-3}$. Since the fast potential is nonlinear, IMEX is an implicit method and nonlinear equations have to be solved at every step, and IMEX turns out to be slower than VE. FLAVOR is strongly accurate with respect to slow variables and accurate in the sense of measures with respect to fast variables. Compared to symplectic Euler, FLAVOR accelerated the computation by roughly 100x.*
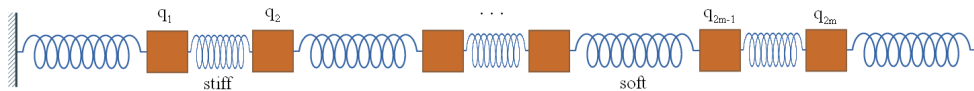

FIG. 6.3. *Fermi–Pasta–Ulam problem [43]—one-dimensional chain of alternatively connected harmonic stiff and nonharmonic soft springs.*

(2.11). On a time scale $\mathcal{O}(\omega)$ ($\omega \gg 1$), FLAVOR captured the slow variable's periodic behavior with the correct period and phase, as well as the slower process of energy transfer. At the same time, FLAVOR accelerated the computation by roughly 40x (since $\delta = 40\tau'$).

It is not worrisome that the artificial FLAVOR produces stiff spring energy trajectories with rapid local oscillations, which exhibit both thicker individual energy curves and total energy with larger variance. In fact, these local oscillations do not seem to affect the global transfer pattern or its period and are caused by the numerical error associated with microstep τ. This can be inferred by using the artificial FLAVOR introduced in subsection 2.1.4 with θ_τ^ϵ corresponding to the exact flow of H^{fast} (rather than its VE approximation: this specific artificial Euler resembles the impulse method, but the impulse method will yield unbounded trajectories if one runs even longer time simulations, whereas FLAVORS do not seem to have an error growing exponentially with the total simulation time). As illustrated in Figure 6.7, exact flow helps to obtain thin energy curves of stiff springs with no rapid local oscillations as well as a total energy with a variance smaller than that given by the fine VE (Figure 6.4(a)), with the transfer pattern similar to Figure 6.4(b).
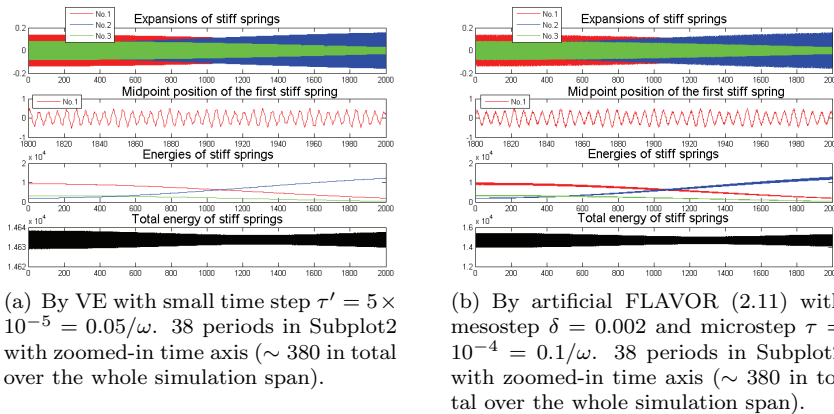
(a) By VE with small time step $\tau' = 5 \times 10^{-5} = 0.05/\omega$. 38 periods in Subplot2 with zoomed-in time axis (~ 380 in total over the whole simulation span).

(b) By artificial FLAVOR (2.11) with mesostep $\delta = 0.002$ and microstep $\tau = 10^{-4} = 0.1/\omega$. 38 periods in Subplot2 with zoomed-in time axis (~ 380 in total over the whole simulation span).

FIG. 6.4. *Simulations of the FPU problem over $T = 2\omega$. Subplot2 of both figures has zoomed-in time axes so that if phase lag or any other distortion of trajectory exists, it could be investigated closely. In this experiment, $m = 3$, $\omega = 10^3$, $x(0) = [0.4642, -0.4202, 0.0344, 0.1371, 0.0626, 0.0810]$ is randomly chosen, and $y(0) = [0, 0, 0, 0, 0, 0]$.*

(a) By Velocity Verlet with small time step $h = 10^{-5}$.

(b) By artificial FLAVOR (2.11) with mesostep $\delta = 0.002$ and microstep $\tau = 0.0005 = 0.1/\omega$.

(c) By IMEX with mesostep $\delta = 0.002$.

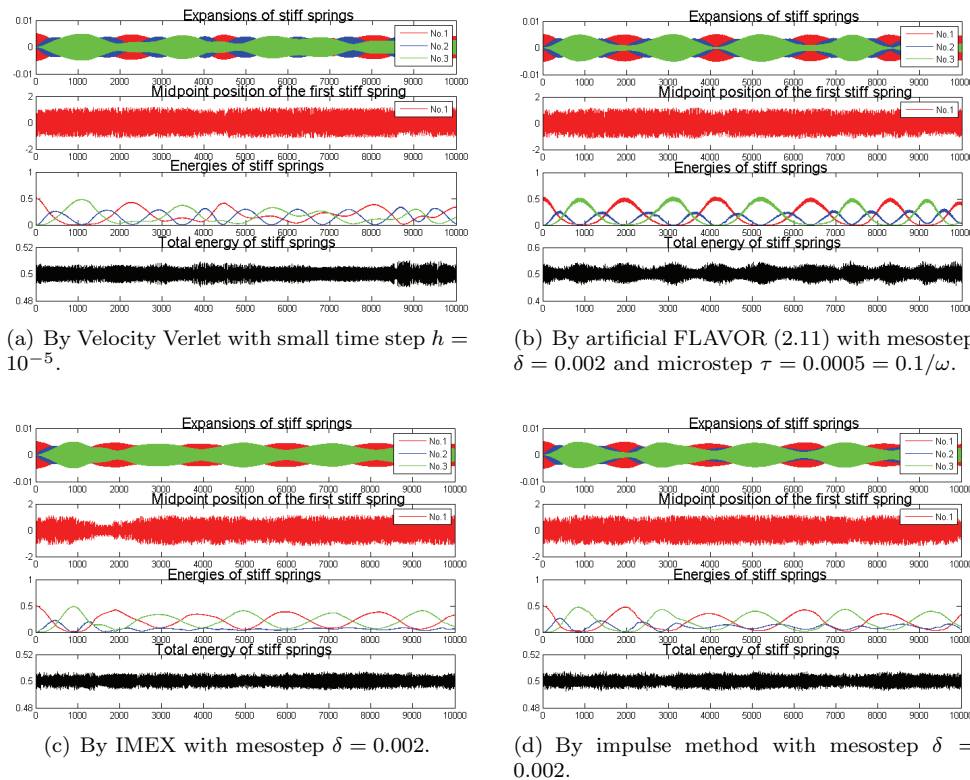(d) By impulse method with mesostep $\delta = 0.002$.

FIG. 6.5. *Simulations of the FPU problem over $T = \frac{1}{4}\omega^2$. Initial conditions are $x(0) = [1, 0, 0, 1/\omega, 0, 0]$ and $y(0) = [0, 0, 0, 0, 0, 0]$ so that energy starts concentrated on the leftmost soft and stiff springs, propagates to the right, bounces back, and oscillates among springs. We chose a smaller $\omega = 200$ because with a larger ω it would take weeks to run Velocity Verlet on a laptop.*
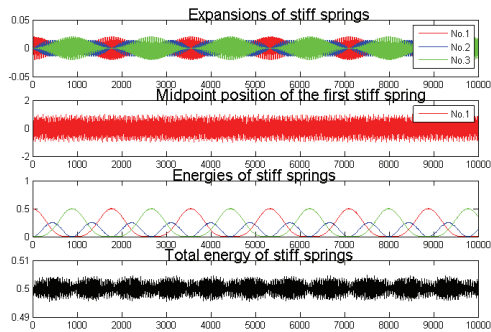
FIG. 6.6. *Harmonic FPU, $T = 50\omega$, exact solution.*
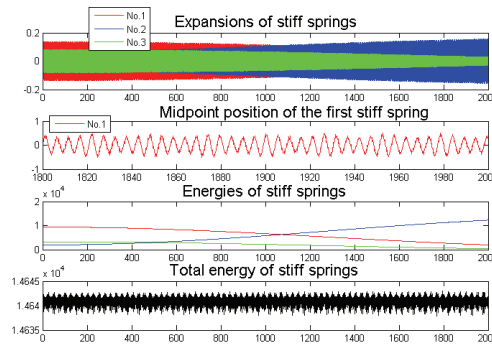

FIG. 6.7. *By artificial FLAVORS (subsection 2.1.4) based on exact fast flow with mesostep $\delta = 0.002$ and microstep $\tau = 10^{-4}$. Less oscillatory stiff spring energies. 38 periods in Subplot2 with zoomed-in time axis (~ 380 in total over the whole simulation span).*

Now we reach further to $\mathcal{O}(\omega^2)$ total integration time to investigate different integrators' performances in capturing long time energy exchange patterns (Figure 6.5).

There is a significant difference among stiff spring energy transfer patterns produced by Velocity Verlet, FLAVOR, IMEX, and the impulse method. Here, there is no analytic solution or provably accurate method for comparison. FLAVOR is the only method that shows periodic behavior on the long time scale, and convergence tests show that FLAVOR's trajectories remain stable under small variations of step sizes.

Notice that the system would be integrable and periodic if nonlinearity did not exist (see Figure 6.6 for integration of a system in which the slow potential is quadratic; that system, by the way, will be perfectly integrated by all FLAVORS, Velocity Verlet, IMEX, and the impulse method). When the number of springs is fixed and very small, in general only strong nonlinearity will destroy the periodicity of the system (see, for instance, [80] for an example of a chaotic threshold on nonlinearity), which should not be the case here since ω is very large.

It is worth discussing why Velocity Verlet with a time step much smaller than the characteristic length of the fast scale ($\mathcal{O}(1/\omega)$) is still not satisfactory. Being a second order method, it has an error bound of $\mathcal{O}(e^T h^2)$. On the other hand, backward error analysis guarantees that the energy of the integrated trajectory oscillates around the true conserved energy, hence eliminating the possibility of exponential growth of the numerical solution. Nevertheless, at this moment there is no result known to the authors to link these two analytical results to guarantee long term accuracy on the stiff springs' energies. The energy exchange among stiff springs is in fact a delicate phenomenon, and a slight distortion in stiff spring lengths could easily disrupt its period or even its periodicity.

These numerical observations seem to indicate that symplectic FLAVORS may have special long time properties. Specifically, although we could not quantify the error here because there is no benchmark to compare to when the total simulation time is $\mathcal{O}(\omega^2)$, the long term behavior seems to indicate an error growing much slower than exponentially (please refer to Remark 1.7 for a discussion on exponential error bounds and Figure 5.4(b) for another example of conjectured linear error growth). A rigorous investigation on FLAVORS' long time behavior remains to be done.
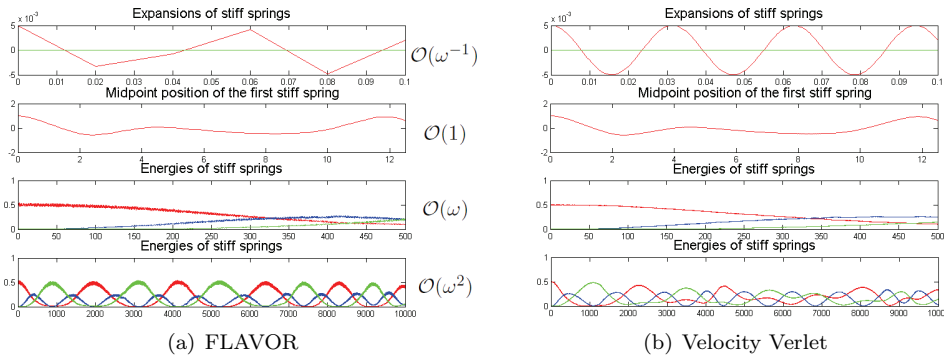
(a) FLAVOR   (b) Velocity Verlet

FIG. 6.8. *Quantities of interest in integrations of FPU over different time scales. FLAVOR (2.11) captures the fastest time scale in the sense of measure, while Velocity Verlet cannot accurately capture the slowest ($\mathcal{O}(\omega^2)$) time scale despite the small time step it uses. Here, FLAVOR is 200 times faster than Velocity Verlet. All parameters are the same as in Figures 6.5(a) and 6.5(b), e.g., $\omega = 200$, $\delta = 0.002$, $\tau = 0.0005$, and $h = 10^{-5}$.*

Figure 6.8 summarizes FLAVOR's performance on various time scales in a comparison to Velocity Verlet.

Notice that there are many sophisticated methods designed for integrating the FPU problem (see [57] for a review), as well as general multiscale methods that can be applied to the FPU problem. HMM, as one state-of-the-art method in the latter category, together with identification of slow variables [5], captured the energy transfer between stiff springs over a time span of the order of ω. Simulations shown here are over a time span of the order of $\mathcal{O}(\omega^2)$.

6.3.1. On resonances. Multiscale in time integrators are usually plagued by two kinds of resonances.

The first kind, called Takens resonance [111], is related to the case in which there are no closed equations for slow variables [15]. FLAVORS avoid Takens resonance because, thanks to F-convergence, the information on the local invariant measure of fast variables is not lost. Observe that the FPU problem has Takens resonance (the eigenfrequencies of the strong potential are identical). Nevertheless, FLAVORS still capture the solution trajectories given any large value of ω with mesostep $\delta \gg 1/\omega$ independent of ω.

The second kind [25] is related to instabilities created by interactions between parameters ϵ, τ, and δ. For instance, if $\epsilon^{-1} = \omega^2$, resonance might happen at $\omega\delta$ or $\omega\tau$ equal to multiples of $\pi/2$. The analysis provided in section 5 shows that such unstable interaction does not occur, either in the sense of stability or in terms of peaks of error function. This can be intuitively understood by observing that FLAVORS never approximate $\cos(\delta\omega)$, while, on the other hand, it does approximate $\cos(\tau\omega)$, whose resonance frequency $\tau = 2\pi/\omega$ is ruled out by the requirement that $\tau \ll \epsilon$ for the nonintrusive FLAVOR and $\tau \ll \sqrt{\epsilon}$ for the artificial FLAVOR.

6.4. Nonlinear two-dimensional primitive molecular dynamics. Now consider a two-dimensional, two degrees of freedom example in which a point mass is linked through a spring to a massless fixed hinge at the origin. While the spring as well as the point mass are allowed to rotate around the hinge (the spring remains straight), the more the spring mass tilts away from its equilibrium angle the more restorative force it will experience. This example is a simplified version of prevailing
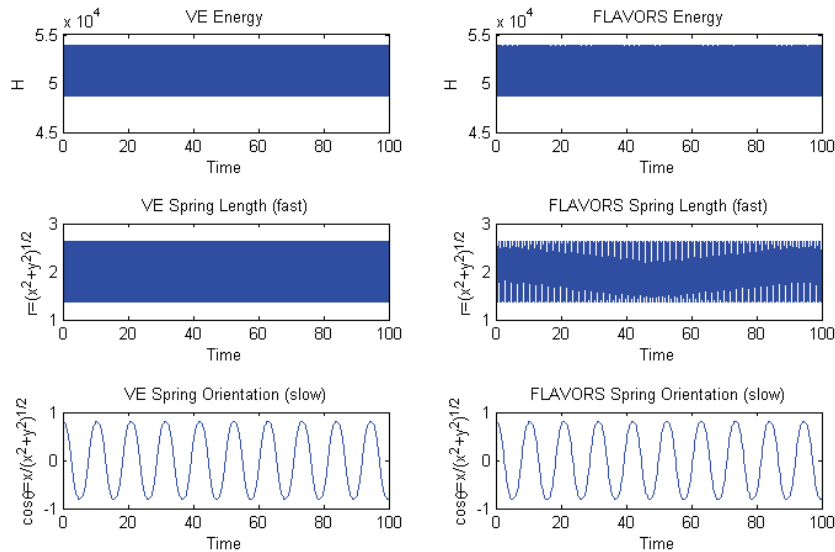
FIG. 6.9. *Simulation of* (6.8). *Symplectic Euler uses small time step* $\tau = 0.0002$, *and the induced symplectic FLAVOR (*(2.5) *and* (2.9)*) uses mesostep* $\delta = 0.01$ *and microstep* $\tau = 0.0002$. *In this simulation,* $\omega = 500$, $x(0) = 1.1$, $y(0) = 0.8$, $p_x(0) = 0$, $p_y(0) = 0$, *and simulation time* $T = 100$.

molecular dynamics models, in which bond lengths and angles between neighboring bonds are both spring-like; other potential energy terms are ignored.

Denote by x and y the Euclidean coordinates of the mass and by p_x and p_y the corresponding momentums. Also introduce polar coordinates (r, θ), with $x = r\cos\theta$ and $y = r\sin\theta$. Then the Hamiltonian reads as

$$
\begin{aligned}
H &= \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}\omega^2(r - r_0)^2 + (\cos\theta)^2 \\
(6.8) \qquad &= \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}\omega^2(\sqrt{x^2 + y^2} - r_0)^2 + \frac{x^2}{x^2 + y^2},
\end{aligned}
$$

where r_0 is an equilibrium bond length parameter and ω is a large number denoting bond oscillation frequency.

REMARK 6.1. *This seemingly trivial example is not easy to integrate.*

1. *If the system is viewed in Euclidean coordinates* (x, y, p_x, p_y), *it is completely nonlinear with a nonpolynomial potential, and hence the impulse method or its variations* [54, 113, 47, 101], *or IMEX* [109], *or the homogenization method introduced in* [20] *cannot be applied using a mesostep.*

2. *If the Hamiltonian is rewritten in generalized coordinates* $(r, \theta, p_r, p_\theta)$, $H = \frac{1}{2}p_r^2 + \frac{1}{2}\frac{p_\theta^2}{r^2} + \frac{1}{2}\omega^2(r - r_0)^2 + \frac{1}{2}\cos(\theta)^2$, *a fast quadratic potential can be identified. However, the mass matrix* $\left[\begin{smallmatrix} 1 & 0 \\ 0 & r^2 \end{smallmatrix}\right]$ *is not constant, but rapidly oscillating, and hence methods that work for quasi-quadratic fast potentials (i.e., a "harmonic oscillator" with a slowly changing frequency) (*[20], *for example) cannot be applied.*

Figure 6.9 compares symplectic Euler with the induced symplectic FLAVOR ((2.5) and (2.9)) applied to (6.8) in Euclidean coordinates.
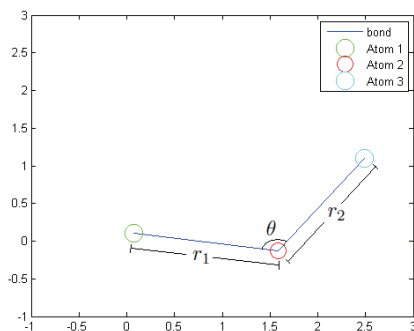
FIG. 6.10. *One example configuration of a propane molecule.*

FLAVOR reproduced the slow θ trajectory while accelerating the simulation time by roughly 50x (since $\delta = 50\tau$). It can also be seen from both energy fluctuations and the trajectory of the fast variable that the fast process's amplitude is well captured, although its period has been lengthened.

6.5. Nonlinear two-dimensional molecular clipper. We now consider a united-atom representation of a three atom polymer with two bonds (e.g., propane or water molecule). This is a simplified version of several prevailing molecular dynamics force fields (for example, CHARMM [21], AMBER [33], or a simpler example of butane [96, 97]). Using conservation of momentum, we fix the coordinate system in the two-dimensional plane defined by the three atoms. We introduce both Cartesian coordinates $(x_1, y_1, x_2, y_2, x_3, y_3)$, as well as generalized coordinates $r_1 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ and $r_2 = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$ for bond lengths and θ for the angle between the two bonds (Figure 6.10). The kinetic energy is

$$(6.9) \qquad K.E. = \frac{1}{2}m_1(\dot{x}_1^2 + \dot{y}_1^2) + \frac{1}{2}m_2(\dot{x}_2^2 + \dot{y}_2^2) + \frac{1}{2}m_3(\dot{x}_3^2 + \dot{y}_3^2),$$

where m_1, m_2, and m_3 denote the masses of the atoms.

The potential energy consists of a bond term and a bond angle term, both of which are of harmonic oscillator type:

$$(6.10) \qquad P.E. = V_{bond} + V_{angle},$$

$$(6.11) \qquad V_{bond} = \frac{1}{2}K_r[(r_1 - r_0)^2 + (r_2 - r_0)^2],$$

$$(6.12) \qquad V_{angle} = \frac{1}{2}K_\theta(\cos(\theta) - \cos(\theta_0))^2.$$

Notice that the system is in fact fully nonlinear: if written in generalized coordinates, the kinetic energy will correspond to a nonlinear and position-dependent mass matrix, whereas in Cartesian coordinates, both terms in the potential energy are nonpolynomial functions of the configuration.

In the case of propane, $m_1 = 15\mu$, $m_2 = 14\mu$, and $m_3 = 15\mu$, where $\mu = 1.67 \cdot 10^{-27}$kg, $r_0 = 1.53$Å, $K_r = 83.7$kcal/(molÅ^2), $\theta_0 = 109.5°$, and $K_\theta = 43.1$kcal/mol [96].

The propane system is characterized by a separation of time scales to some extent: bond stretching and bond-angle bending are characterized by 10^{14} and 10^{13} Hz
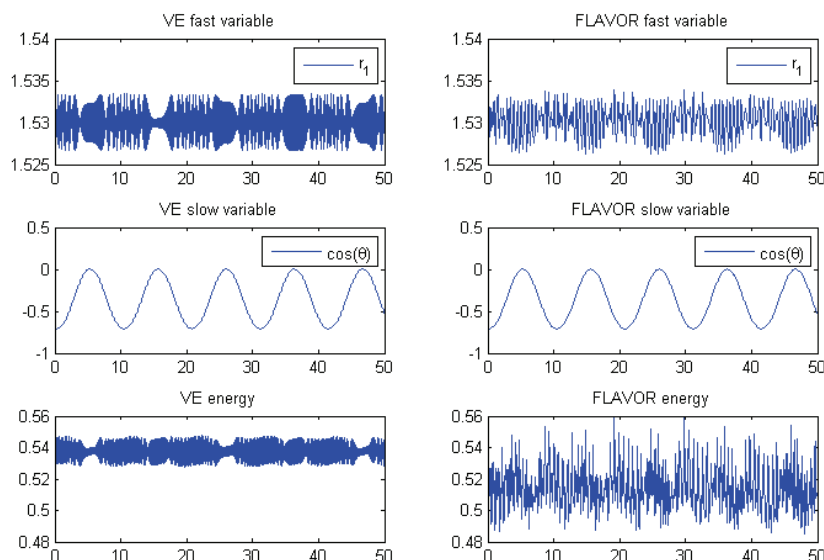
FIG. 6.11. *Simulations of an exaggerated propane molecule (subsection* 6.5*). Symplectic Euler uses* $h = 0.01$*, and the induced symplectic FLAVOR (*(2.5) *and* (2.9)*) parameters are* $\delta = 0.1$ *and* $\tau = 0.01$*. Initial conditions are* $[x_1, y_1, x_2, y_2, x_3, y_3] = [0, 0, 1.533, 0, 2.6136, 1.0826]$ *and* $[m_1\dot{x}_1, m_1\dot{y}_1, m_2\dot{x}_2, m_2\dot{y}_2, m_3\dot{x}_3, m_3\dot{y}_3] = [-0.4326, -1.6656, 0.1253, 0.2877, -1.1465, 1.1909]$.

vibrational frequencies, respectively [120]. To examine FLAVORS, we use unitless parameters and exaggerate the time scale separation by setting K_r to be 8370 and K_θ to be 4.31. We also let $\mu = 1$, without loss of generality, for arithmetic considerations.

In this system, the bond potential is the fast potential and the bond-angle potential is the slow one. It is well known that using a large time step at the time scale corresponding to the bond-angle potential by freezing bond lengths produces biased results, and many physics-based methods have been proposed to remedy this difficulty (for example, by the approach of Fixman [44]; see also a review in [120]). On the other hand, few multiscale methods work for this fully nonlinear system.

Figure 6.11 compares symplectic Euler with the induced symplectic FLAVOR ((2.5) and (2.9)) applied in Euclidean coordinates. 10x acceleration is achieved. A simulation movie is also available at http://www.cds.caltech.edu/∼mtao/Propane.avi and http://www.acm.caltech.edu/∼owhadi/.

6.6. Forced nonautonomous mechanical system: Kapitza's inverted pendulum. As the famous Kapitza's inverted pendulum shows [66] (for recent references, see [6] for numerical integration and [102] for generalization to the stochastic setting), the up position of a single pendulum can be stabilized if the pivot of the pendulum experiences external forcing in the form of vertical oscillation. Specifically, if the position of the pivot is given by $y = \sin(\omega t)$, the system is governed by

$$(6.13) \qquad\qquad l\ddot{\theta} = [g + \omega^2 \sin(2\pi\omega t)] \sin\theta,$$

where θ denotes the clockwise angle of the pendulum from the positive y direction, l is the length of the pendulum, and g is the gravitational constant. In this case, the
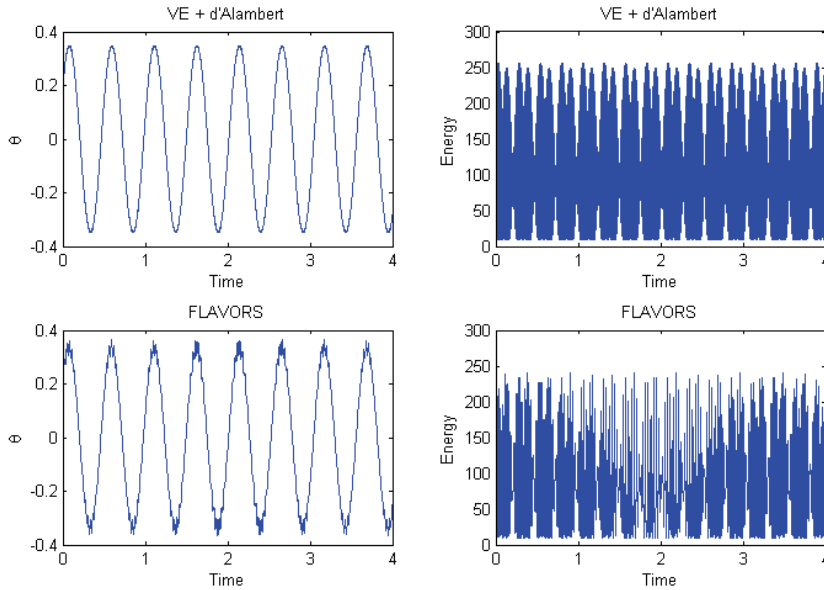
FIG. 6.12. *Simulations of the inverted pendulum. The integration by VE and the d'Alembert principle uses time step $h = 0.2/\omega/\sqrt{l} \approx 0.000067$, while FLAVOR (defined by (6.15)) uses $\delta = 0.002$ and $\tau = 0.2/\omega/\sqrt{l}$. Also, $g = 9.8$, $l = 9$, $\theta(0) = 0.2$, $\dot{\theta}(0) = 0$, and $\omega = 1000$.*

rapid vibration causes the pendulum to oscillate slowly around the positive y direction with a $\mathcal{O}(1)$ frequency.

A single scale integration of this system could be done by VE with the discrete d'Alembert principle for external forces [82]

$$(6.14) \quad \begin{cases} f_i = \omega^2 \sin(2\pi\omega i h), \\ p_{i+1} = p_i + h[g + f_i]\sin\theta_i, \\ \theta_{i+1} = \theta_i + hp_{i+1}/l, \end{cases}$$

where the time step length h has to be smaller than $\mathcal{O}(1/\omega)$.

FLAVOR is given by

$$(6.15) \quad \begin{cases} q_{n\delta+\tau} = q_{n\delta} + \tau p_{n\delta}/l, \\ p_{n\delta+\tau} = p_{n\delta} + \tau g \sin(q_{n\delta+\tau}) + \omega^2 \sin(2\pi\omega n\tau), \\ q_{(n+1)\delta} = q_{n\delta+\tau} + (\delta - \tau)p_{n\delta+\tau}/l, \\ p_{(n+1)\delta} = p_{n\delta+\tau} + (\delta - \tau)g \sin(q_{(n+1)\delta}). \end{cases}$$

Observe that the time-dependent force is synchronized on the τ time scale instead of the δ time scale, specifically $\omega^2 \sin(2\pi\omega n\tau)$ instead of $\omega^2 \sin(2\pi\omega n\delta)$ in (6.15).

Numerical results are illustrated in Figure 6.12 (also available as a movie at http://www.cds.caltech.edu/~mtao/InvertedPendulum.avi and http://www.acm.caltech.edu/~owhadi/). Notice in this example that θ, being the only degree of free-dom, contains a combination of slow and fast dynamics. FLAVOR could capture only the fast dynamics in the sense of measures, and this is why dents appear as modulation on the slow oscillation of θ. On the other hand, although this forced system does not

admit a conserved energy, the value of the Hamiltonian should oscillate periodically due to the periodic external driving force. While a nonmechanics-based method such as forward Euler often produces an unbounded growth or a decrease in the energy, FLAVORS do not have this drawback.

REMARK 6.2. *Consider the case of a rapid potential of the form $\Omega^2(q_1)q_2^2/\epsilon^2$ (where q_1 is the slow variable and q_2 the fast variable). In the limit of a vanishing ϵ, it is known that the term contributes to the effective Hamiltonian with a contribution $V(q_1)$ (the so-called Fixman term). One may suppose that FLAVOR would be consistent only with a term of the form $\gamma V(q_1)$, where $0 < \gamma < 1$, because the rapid force is accounted for only over a time $\tau < \delta$. This intuition is not correct because the effect of FLAVOR is not to account for the rapid force over a time $\tau < \delta$ but to slow down the rapid force by a fraction τ/ϵ. This effect can also be seen in the algorithm* (6.15), *where the force term $\omega^2\sin(2\pi\omega n\tau)$ has been slowed down by a factor τ/δ (Kapitza's inverted pendulum illustrates a similar phenomenon where rapid oscillations contribute a stabilizing term to the effective Hamiltonian; nevertheless, FLAVORS remain accurate).*

6.7. Nonautonomous SDE system with hidden slow variables. Consider the following artificial nonautonomous SDE system:

(6.16)
$$\begin{cases} du = \frac{4}{3(u+v)^2}\left(-\frac{1}{2}\left(\frac{v-u}{2}\right)^2 + 5\sin(2\pi t)\right)dt - \frac{1}{\epsilon}\left(\left(\frac{u+v}{2}\right)^3 + c - \frac{v-u}{2}\right)dt - \sqrt{\frac{2}{\epsilon}}dW_t, \\ dv = \frac{4}{3(u+v)^2}\left(-\frac{1}{2}\left(\frac{v-u}{2}\right)^2 + 5\sin(2\pi t)\right)dt + \frac{1}{\epsilon}\left(\left(\frac{u+v}{2}\right)^3 + c - \frac{v-u}{2}\right)dt + \sqrt{\frac{2}{\epsilon}}dW_t, \end{cases}$$

where c is a positive constant and the two dW_t terms refer to the same Brownian motion. The system (6.16) can be converted via the local diffeomorphism

(6.17)
$$\begin{cases} u = (x-c)^{1/3} - y, \\ v = (x-c)^{1/3} + y \end{cases}$$

into the following hidden system separating slow and fast variables:

(6.18)
$$\begin{cases} dx = -\frac{1}{2}y^2 dt + 5\sin(2\pi t)dW_t, \\ dy = \frac{1}{\epsilon}(x-y)dt + \sqrt{\frac{2}{\epsilon}}dW_t. \end{cases}$$

Nonintrusive FLAVOR (3.7) can be directly applied to (6.16) using a time step $\delta \gg \epsilon$ without prior identification of the slow and fast variables, i.e., without prior identification of the slow variable x or of the system (6.18). The expected values of solutions of (6.16) integrated by FLAVORS with mesostep δ and Euler–Maruyama with a small time step τ are presented in Figure 6.13. FLAVOR has accelerated the computation by 100x.

6.8. Langevin equations with slow noise and friction. In this subsection, we consider the one-dimensional, two degrees of freedom system modeled by the SDEs (now both springs are quartic rather than harmonic)

(6.19)
$$\begin{cases} dy = p_y dt, \\ dx = p_x dt, \\ dp_y = -\epsilon^{-1}y^3 dt - 4(y-x)^3 dt - cp_y dt + \sigma dW_t^1, \\ dp_x = -4(x-y)^3 dt - cp_x dt + \sigma dW_t^2. \end{cases}$$
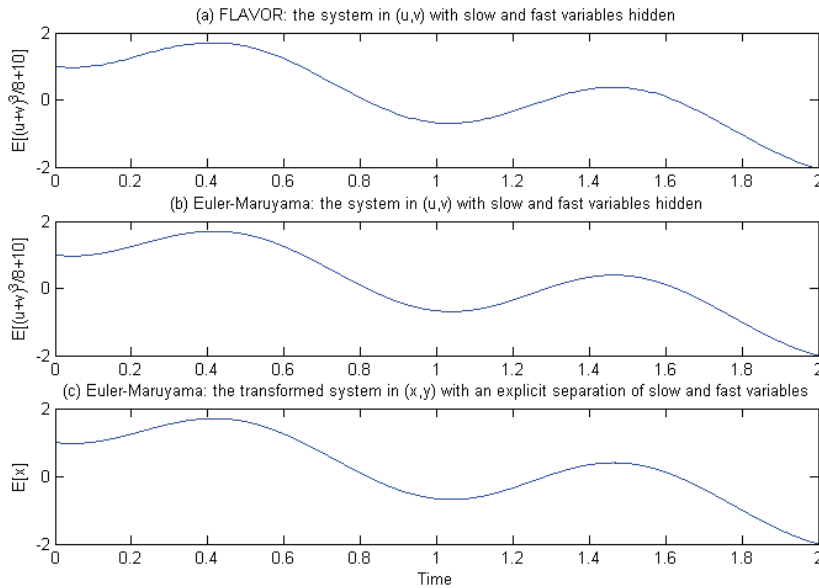
FIG. 6.13. (a) *Integration of* (6.16) *by nonintrusive FLAVOR* (3.7) *using mesostep* $\delta = 0.01$. (b) *Integration of* (6.16) *by Euler–Maruyama using fine time step* $h = 10^{-4}$. (c) *Integration of* (6.18) *by Euler–Maruyama using the same small step* $h = 10^{-4}$. *Expectations of the slow variable (whether or not hidden) are obtained by empirically averaging over an ensemble of* 100 *independent sample trajectories.* $\epsilon = 10^{-4}$, $x(0) = 1+\epsilon$, $y(0) = 1$, *and* $T = 2$ *(the expectation of the real solution will blow up around* $T = 3$*). We have chosen* $c = 10$ *so that the transformation is a diffeomorphism.*

We compare several autocorrelation functions and time-dependent moments of this stochastic process integrated by a quasi-symplectic FLAVOR ((4.7) and (2.9)) and the geometric Langevin algorithm (GLA) [17]. FLAVOR and GLA gave results in agreement (Figures 6.14, 6.15(a), and 6.15(b)). Since GLA is weakly convergent and Boltzmann–Gibbs preserving, this is numerical evidence that the quasi-symplectic FLAVOR is too.

Expectations are empirically calculated by averaging over an ensemble of 100 sample trajectories with $T = 30$, $\epsilon = 10^{-8}$, $\tau = 0.001$, and $\delta = 0.01$. $y(0) = 2.1/\omega$ (with $\omega := 1/\sqrt{\epsilon}$), $x(0) = y(0) + 1.8$, $c = 0.1$, and $\sigma = 0.5$. GLA uses time step $h = 0.001$. Noise and friction are slow here in the sense that they are not of the order $\mathcal{O}(\omega)$ or larger.

As shown in the plots, in the regime dominated by deterministic dynamics (roughly from $t = 0$ to $t = 8$) various moments calculated empirically by FLAVORS and GLA are in agreement, indicating that the same rate of convergence towards the Boltzmann–Gibbs distribution is obtained. And in that regime, autocorrelation functions of the slow variables agree, serving as numerical evidence that FLAVORS are weakly converging towards the SDE solution, whereas autocorrelation functions of the fast variables agree only in the sense of measures (after time averaging over a mesoscopic ($o(1)$) time span). The fluctuations between FLAVORS and GLA for large times are an effect of the finite number of samples (100) used to compute sample averages.

Recall that if the noise is applied to slow variables, FLAVORS do not converge strongly but only in the sense of distributions.
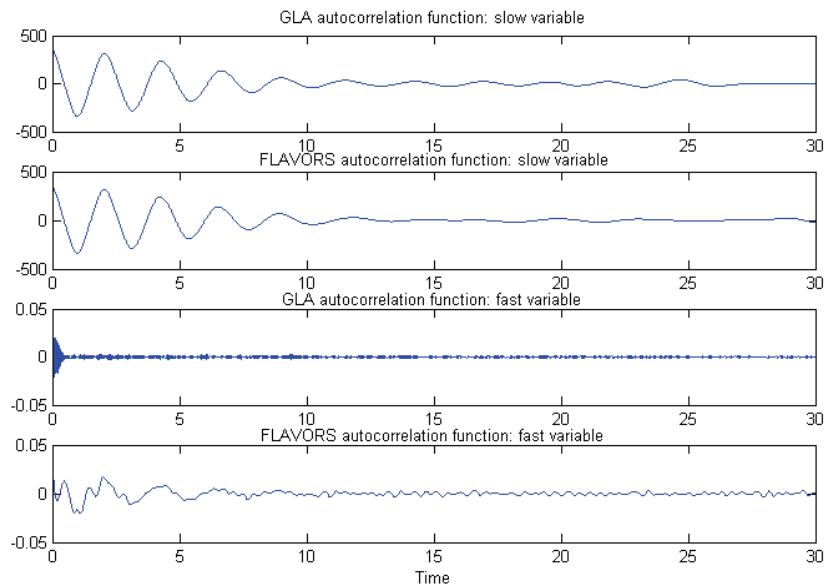
FIG. 6.14. *SDE* (6.19): *autocorrelation functions of* $\mathbb{E}[y(t)y(0)]$ *(dominantly fast) and of* $\mathbb{E}[(x(t) - y(t))(x(0) - y(0))]$ *(dominantly slow), empirically obtained by GLA and FLAVORS.*
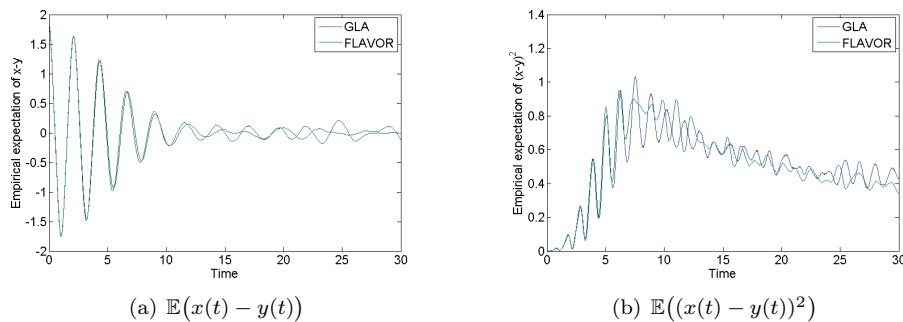

(a) $\mathbb{E}\big(x(t) - y(t)\big)$

(b) $\mathbb{E}\big((x(t) - y(t))^2\big)$

FIG. 6.15. *SDE* (6.19): *empirical moments obtained from simulations of ensembles of* (6.19) *with GLA and a quasi-symplectic FLAVOR (subsection* 4.2.1*).*

6.9. Langevin equations with fast noise and friction. Consider a system with the same configuration as above. The difference is that the soft spring oscillates at a frequency nonlinearly dependent on the stiff spring's length, and the left mass experiences strong friction and noise while the right mass does not. The Hamiltonian is

$$(6.20) \qquad H(y, x, p_y, p_x) = \frac{1}{2}p_y^2 + \frac{1}{2}p_x^2 + \frac{1}{4}\omega^4 y^4 + e^y(x - y)^2,$$

and the governing SDEs are

$$(6.21) \qquad \begin{cases} dy = p_y dt, \\ dx = p_x dt, \\ dp_y = -\omega^4 y^3 dt - (2 + y - x)(y - x)e^y dt - \omega^2 c p_y dt + \omega\sigma dW^t, \\ dp_x = -2(x - y)e^y dt. \end{cases}$$
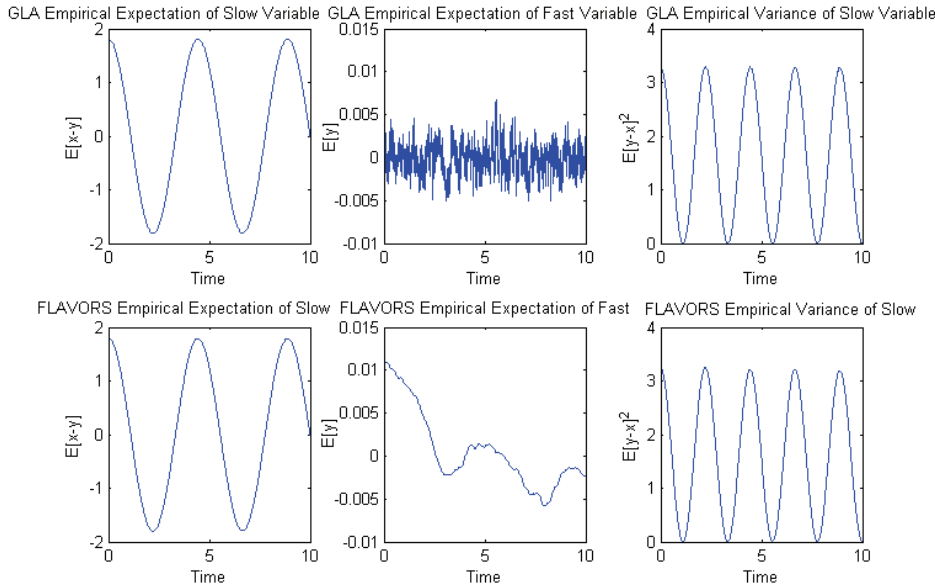
FIG. 6.16. $\mathbb{E}[x(t) - y(t)]$, $\mathbb{E}[y(t)]$, and $\mathbb{E}[x(t) - y(t)]^2$ obtained by GLA and a quasi-symplectic FLAVOR (subsection 4.2.1). Expectations are empirically calculated by averaging over an ensemble of 50 sample trajectories with $T = 10$, $\omega = 100$, $\tau = 10^{-4}$, $\delta = 0.01$, $y(0) = 1.1/\omega$, $x(0) = y(0) + 1.8$, $c = 0.1$, and $\sigma = 1$. GLA uses time step $h = 10^{-4}$.

In this system, the deterministic dynamics and the effects of noise and friction both involve a $\mathcal{O}(1/\omega^2)$ time scale. We have implemented the fast noise and friction version of FLAVORS ((4.8) and (2.9)).

In Figure 6.16, we have plotted the first and second moments of the dominantly slow variable $x(t) - y(t)$ as well as the first moment of the dominantly fast variable $y(t)$ as functions of time. Moments of the dominantly slow variable integrated by a quasi-symplectic FLAVOR (subsection 4.2.1) and GLA [17] concur, numerically suggesting weak convergence and preservation of Boltzmann–Gibbs. 100x computational acceleration is achieved.

Appendix.

A.1. Proof of Theorems 1.2 and 1.3. Define the process $t \mapsto (\bar{x}_t, \bar{y}_t)$ by

(A.1)
$$(\bar{x}_t, \bar{y}_t) := \eta(\bar{u}_t).$$

It follows from the regularity of η that it is sufficient to prove the F-convergence of $(\bar{x}_t, \bar{y}_t)$ towards $\delta_{X_t} \otimes \mu(X_t, dy)$. Moreover, it is also sufficient to prove inequalities (A.2) and (A.3) in order to obtain inequalities (1.24) and (1.25):

(A.2)
$$|x_t^\epsilon - \bar{x}_t| \le C e^{Ct} \psi_1(u_0, \epsilon, \delta, \tau)$$

and
(A.3)
$$\left| \frac{1}{T} \int_t^{t+T} \varphi(\bar{x}_s, \bar{y}_s) \, ds - \int_{\mathbb{R}^p} \varphi(X_t, y) \mu(X_t, dy) \right| \le \psi_2(u_0, \epsilon, \delta, \tau, T, t)(\|\varphi\|_{L^\infty} + \|\nabla\varphi\|_{L^\infty}).$$

Now define ψ_τ^ϵ by

$$(A.4) \qquad \psi_\tau^\epsilon(x,y) := \eta \circ \theta_\tau^\epsilon \circ \eta^{-1}(x,y).$$

Define ψ_h^g by

$$(A.5) \qquad \psi_h^g(x,y) := \eta \circ \theta_h^G \circ \eta^{-1}(x,y).$$

PROPOSITION A.1. *The vector fields f and g associated with the system of equations (1.2) are Lipschitz continuous. We also have*

$$(A.6) \qquad (\bar{x}_t, \bar{y}_t) = \left(\psi_{\delta-\tau}^g \circ \psi_\tau^\epsilon\right)^k(x_0, y_0) \quad for \quad k\delta \le t < (k+1)\delta.$$

Moreover, there exists $C > 0$ such that for $h \le h_0$ and $\frac{\tau}{\epsilon} \le \tau_0$ we have

$$(A.7) \qquad \left|\psi_\tau^\epsilon(x,y) - (x,y) - \tau\left(g(x,y),0\right) - \frac{\tau}{\epsilon}\left(0, f(x,y)\right)\right| \le C\left(\frac{\tau}{\epsilon}\right)^2$$

and

$$(A.8) \qquad \left|\psi_h^g(x,y) - (x,y) - h\left(g(x,y),0\right)\right| \le Ch^2.$$

Furthermore, given x_0, y_0, the trajectories of $(x_t^\epsilon, y_t^\epsilon)$ and $(\bar{x}_t, \bar{y}_t)$ are uniformly bounded in ϵ, $\delta \le h_0$, and $\tau \le \min(\tau_0\epsilon, \delta)$.

Proof. Since $(x,y) = \eta(u)$, we have

$$(A.9) \qquad \dot{x} = \left(G + \frac{1}{\epsilon}F\right)\nabla\eta^x \circ \eta^{-1}(x,y),$$

$$(A.10) \qquad \dot{y} = \left(G + \frac{1}{\epsilon}F\right)\nabla\eta^y \circ \eta^{-1}(x,y).$$

Hence, we deduce from (1.2) of Condition 1.1 that

$$(A.11) \qquad g(x,y) = G\nabla\eta^x \circ \eta^{-1}(x,y),$$

$$(A.12) \qquad f(x,y) = F\nabla\eta^y \circ \eta^{-1}(x,y).$$

We deduce the regularity of f and g from the regularity of G, F, and η. Equation (A.6) is a direct consequence of the definition of ψ_τ^ϵ and ψ_h^g and (1.27) (we write $(x_0, y_0) := \eta(u_0)$). Observe that (1.2) of Condition 1.1 also requires that

$$(A.13) \qquad F\nabla\eta^x = 0, \qquad G\nabla\eta^y = 0.$$

Now observe that

$$(A.14) \qquad \begin{aligned} &\psi_\tau^\epsilon(x,y) - (x,y) - \left(g(x,y),0\right)\tau - \left(0, f(x,y)\right)\frac{\tau}{\epsilon} \\ &= \left(\eta \circ \theta_\tau^\epsilon - \eta - \tau\left(G\nabla\eta^x, 0\right) - \frac{\tau}{\epsilon}\left(0, F\nabla\eta^y\right)\right) \circ \eta^{-1}(x,y). \end{aligned}$$

Using (A.13), (1.29), Taylor expansion, and the regularity of η, we obtain (A.7). Similarly,

$$(A.15) \quad \psi_h^g(x,y) - (x,y) - h\left(g(x,y),0\right) := \left(\eta \circ \theta_h^G - \eta(x,y) - h\left(G\nabla\eta^x, 0\right)\right) \circ \eta^{-1}(x,y).$$

Using (A.13), (1.28), Taylor expansion, and the regularity of η, we obtain (A.8). The uniform bound (depending on x_0, y_0) on the trajectories of $(x_t^\epsilon, y_t^\epsilon)$ and $(\bar{x}_t, \bar{y}_t)$ is a consequence of the uniform bound (given u_0) on the trajectories of u_t^ϵ and $\bar{u}_t$. $\quad\square$

It follows from Proposition A.1 that it is sufficient to prove Theorems 1.2 and 1.3 in the situation where η is the identity diffeomorphism. More precisely, the F-convergence of $\bar{u}_t$ is a consequence of the F-convergence of $(\bar{x}_t, \bar{y}_t)$ and the regularity of η. Furthermore, from the uniform bound (depending on (x_0, y_0)) on the trajectories of $(x_t^\epsilon, y_t^\epsilon)$ and $(\bar{x}_t, \bar{y}_t)$, we deduce that g and f are uniformly bounded and Lipschitz continuous (in ϵ, $\delta \leq h_0$, $\tau \leq \min(\tau_0\epsilon, \delta)$) over those trajectories.

Define

$$\bar{g} := \int g(x, y) \, \mu(x, dy),$$

where μ is the family of measures introduced in Condition 1.2. Let us prove the following lemma.

LEMMA A.2.

$$(A.16) \qquad |x_{n\delta}^\epsilon - \bar{x}_{n\delta}| \leq Ce^{Cn\delta} \left(\delta + \left(\frac{\tau}{\epsilon} \right)^2 \frac{1}{\delta} + \sup_{1 \leq l \leq n} |J(l)| \right)$$

with $J(k) = J_1(k) + J_2(k)$,

$$(A.17) \qquad J_1(k) := \sum_{n=0}^{k-1} \left(\int_{n\delta}^{(n+1)\delta} g(x_{n\delta}^\epsilon, y_s^\epsilon) \, ds - \delta\bar{g}(x_{n\delta}^\epsilon) \right),$$

and

$$(A.18) \qquad J_2(k) := \sum_{n=0}^{k-1} \delta\big(\bar{g}(\bar{x}_{n\delta}) - g(\bar{x}_{n\delta}, \bar{y}_{n\delta})\big).$$

Proof. Observe that

$$(A.19) \quad x_{(n+1)\delta}^\epsilon = x_{n\delta}^\epsilon + \int_{n\delta}^{(n+1)\delta} g(x_{n\delta}^\epsilon, y_s^\epsilon) \, ds + \int_{n\delta}^{(n+1)\delta} (g(x_s^\epsilon, y_s^\epsilon) - g(x_{n\delta}^\epsilon, y_s^\epsilon)) \, ds.$$

Hence,

$$(A.20) \qquad x_{(n+1)\delta}^\epsilon - \bar{x}_{(n+1)\delta} = x_{n\delta}^\epsilon - \bar{x}_{n\delta} + I_1 + I_2(n) + I_3 + I_4(n) + I_5$$

with

$$(A.21) \qquad I_1 := \int_{n\delta}^{(n+1)\delta} (g(x_s^\epsilon, y_s^\epsilon) - g(x_{n\delta}^\epsilon, y_s^\epsilon)) ds,$$

$$(A.22) \qquad I_2(n) := \int_{n\delta}^{(n+1)\delta} g(x_{n\delta}^\epsilon, y_s^\epsilon) \, ds - \delta\bar{g}(x_{n\delta}^\epsilon),$$

$$(A.23) \qquad I_3 := \delta\big(\bar{g}(x_{n\delta}^\epsilon) - \bar{g}(\bar{x}_{n\delta})\big),$$

$$(A.24) \qquad I_4(n) := \delta\big(\bar{g}(\bar{x}_{n\delta}) - g(\bar{x}_{n\delta}, \bar{y}_{n\delta})\big),$$

$$(A.25) \qquad I_5 := \delta g(\bar{x}_{n\delta}, \bar{y}_{n\delta}) - (\bar{x}_{(n+1)\delta} - \bar{x}_{n\delta}).$$

Now observe that

$$(A.26) \qquad |I_1| \leq \|\nabla_x g\|_{L^\infty} \delta^2$$

and

$$(A.27) \qquad |I_3| \leq \delta \|\nabla_x g\|_{L^\infty} |x_{n\delta}^\epsilon - \bar{x}_{n\delta}|.$$

Using (A.7) and (A.8), we obtain

$$(A.28) \qquad |I_5| \leq C \left(\delta^2 + \left(\frac{\tau}{\epsilon} \right)^2 \right).$$

Combining the previous equations, we obtain

$$(A.29) \quad x_{(n+1)\delta}^\epsilon - \bar{x}_{(n+1)\delta} \leq x_{n\delta}^\epsilon - \bar{x}_{n\delta} + C \left(\delta^2 + \left(\frac{\tau}{\epsilon} \right)^2 \right) + C\delta |x_{n\delta}^\epsilon - \bar{x}_{n\delta}| + (I_2 + I_4)(n)$$

and

$$(A.30) \quad x_{(n+1)\delta}^\epsilon - \bar{x}_{(n+1)\delta} \geq x_{n\delta}^\epsilon - \bar{x}_{n\delta} - C \left(\delta^2 + \left(\frac{\tau}{\epsilon} \right)^2 \right) - C\delta |x_{n\delta}^\epsilon - \bar{x}_{n\delta}| + (I_2 + I_4)(n).$$

Write

$$(A.31) \qquad J(n) := \sum_{k=0}^{n-1} (I_2 + I_4)(k).$$

Summing the first n inequalities (A.29) and (A.30), we obtain

$$(A.32) \qquad x_{n\delta}^\epsilon - \bar{x}_{n\delta} \leq C \left(\delta^2 + \left(\frac{\tau}{\epsilon} \right)^2 \right) n + C\delta \sum_{k=0}^{n-1} |x_{k\delta}^\epsilon - \bar{x}_{k\delta}| + J(n)$$

and

$$(A.33) \qquad x_{n\delta}^\epsilon - \bar{x}_{n\delta} \geq -C \left(\delta^2 + \left(\frac{\tau}{\epsilon} \right)^2 \right) n - C\delta \sum_{k=0}^{n-1} |x_{k\delta}^\epsilon - \bar{x}_{k\delta}| + J(n).$$

Hence

$$(A.34) \qquad |x_{n\delta}^\epsilon - \bar{x}_{n\delta}| \leq C \left(\delta^2 + \left(\frac{\tau}{\epsilon} \right)^2 \right) n + C\delta \sum_{k=0}^{n-1} |x_{k\delta}^\epsilon - \bar{x}_{k\delta}| + |J(n)|.$$

And we obtain by induction

$$(A.35) \qquad |x_{n\delta}^\epsilon - \bar{x}_{n\delta}| \leq C \left(\delta^2 + \left(\frac{\tau}{\epsilon} \right)^2 \right) \left(n + C\delta \sum_{k=1}^{n} (n-k)(1+C\delta)^{k-1} \right)$$
$$+ |J(n)| + C\delta \sum_{l=2}^{n} (1+C\delta)^{l-2} |J(n-l+1)|.$$

Equation (A.35) concludes the proof of Lemma A.2. ☐
     We now need to control $J_1(k)$ and $J_2(k)$. First, let us prove the following lemma.

LEMMA A.3. *For $N \in \mathbb{N}^*$, we have*

$$(A.36) \qquad |J_1(k)| \leq (\delta k) C \left(\delta e^{C \frac{\delta}{N\epsilon}} + E \left(\frac{\delta}{N\epsilon} \right) \right).$$

Proof. Define $\hat{y}_t^\epsilon$ such that $\hat{y}_t^\epsilon = y_t^\epsilon$ for $t = (n + j/N)\delta$, $j \in \mathbb{N}^*$, and

$$(A.37) \qquad \frac{d\hat{y}_t^\epsilon}{dt} = \frac{1}{\epsilon} f(x_{n\delta}^\epsilon, \hat{y}_t^\epsilon) \quad \text{for} \quad (n + j/N)\delta \leq t < (n + (j+1)/N)\delta.$$

Using the regularity of f and g, we obtain

$$(A.38) \qquad |\hat{y}_t^\epsilon - y_t^\epsilon| \leq C\delta e^{C \frac{\delta}{N\epsilon}}.$$

First, observe that

$$(A.39) \qquad \frac{1}{\delta} \int_{n\delta}^{(n+1)\delta} g(x_{n\delta}^\epsilon, y_s^\epsilon)\, ds - \bar{g}(x_{n\delta}^\epsilon) = K_1 + K_2$$

with

$$(A.40) \qquad K_1 := \frac{1}{\delta} \sum_{j=0}^{N-1} \int_{(n+j/N)\delta}^{(n+(j+1)/N)\delta} \left(g(x_{n\delta}^\epsilon, y_s^\epsilon) - g(x_{n\delta}^\epsilon, \hat{y}_s^\epsilon) \right) ds$$

and

$$(A.41) \qquad K_2 := \frac{1}{N} \sum_{j=0}^{N-1} \left(\frac{N}{\delta} \int_{(n+j/N)\delta}^{(n+(j+1)/N)\delta} g(x_{n\delta}^\epsilon, \hat{y}_s^\epsilon)\, ds - \bar{g}(x_{n\delta}^\epsilon) \right).$$

We have

$$(A.42) \qquad |K_1| \leq \|\nabla_y g\|_{L^\infty} \frac{1}{N} \sum_{j=0}^{N-1} \sup_{(n+j/N)\delta \leq s \leq (n+(j+1)/N)\delta} |y_s^\epsilon - \hat{y}_s^\epsilon|.$$

Hence, we obtain from (A.38)

$$(A.43) \qquad |K_1| \leq C\delta e^{C \frac{\delta}{N\epsilon}}.$$

Moreover, we obtain from Conditions 1.2 and 1.3

$$(A.44) \qquad |K_2| \leq C E \left(\frac{\delta}{N\epsilon} \right).$$

This concludes the proof of Lemma A.3. $\quad \square$

LEMMA A.4. *We have for $m \in \mathbb{N}^*$*

$$(A.45) \qquad \left| J_2(k) \right| \leq C\delta k \left(m\delta + E\left(\frac{m\tau}{\epsilon} \right) + \left(\frac{\tau}{\epsilon} + m\delta + m \left(\frac{\tau}{\epsilon} \right)^2 \right) e^{C \frac{m\tau}{\epsilon}} \right).$$

Proof. Let $m \in \mathbb{N}^*$. Define $(\tilde{x}_s, \tilde{y}_s)$ such that for $j \in \mathbb{N}^*$, $n \in \mathbb{N}^*$

$$(A.46) \qquad \begin{cases} \frac{d\tilde{x}_s}{dt} = g(\tilde{x}_s, \tilde{y}_s) & \text{for} \quad jm\delta \leq s < (j+1)m\delta, \\ \frac{d\tilde{y}_s}{dt} = \frac{1}{\epsilon} f(\tilde{x}_s, \tilde{y}_s) & \text{for} \quad n\delta \leq s < n\delta + \tau, \\ \tilde{y}_s = \tilde{y}_{n\delta+\tau} & \text{for} \quad n\delta + \tau \leq s < (n+1)\delta, \\ \tilde{y}_{(n+1)\delta} = \tilde{y}_{n\delta+\tau} & \text{for} \quad n+1 \neq jm, \\ (\tilde{x}_{jm}, \tilde{y}_{jm}) = (\bar{x}_{jm\delta}, \bar{y}_{jm\delta}). \end{cases}$$

Define $\tilde{y}_s^a$ by

(A.47)
$$\begin{cases} \frac{d\tilde{y}_t^a}{dt} = \frac{1}{\epsilon} f(\bar{x}_{jm\delta}, \tilde{y}_t^a) & \text{for} \quad jm\tau \le t < (j+1)m\tau, \\ \tilde{y}_{jm\tau}^a = \bar{y}_{jm\delta} \end{cases}$$

and define $\tilde{x}_n^a$ by

(A.48)
$$\tilde{x}_n^a = \bar{x}_{jm\delta} \quad \text{for} \quad jm \le n < (j+1)m.$$

Observe that

(A.49)
$$J_2(k) = K_3 + K_4 + K_5 + K_6 + K_7$$

with

(A.50)
$$K_3 := \sum_{n=0}^{k-1} \left(\int_{n\delta}^{(n+1)\delta} g(\tilde{x}_s, \tilde{y}_s) \, ds - \delta g(\bar{x}_{n\delta}, \bar{y}_{n\delta}) \right),$$

(A.51)
$$K_4 := \sum_{n=0}^{k-1} \delta \left(\frac{1}{\tau} \int_{n\tau}^{(n+1)\tau} g(\tilde{x}_n^a, \tilde{y}_s^a) \, ds - \frac{1}{\delta} \int_{n\delta}^{(n+1)\delta} g(\tilde{x}_s, \tilde{y}_s) \, ds \right),$$

(A.52)
$$K_5 := \frac{\delta}{\tau} \sum_{n=0}^{k-1} \left(\tau \bar{g}(\tilde{x}_n^a) - \int_{n\tau}^{(n+1)\tau} g(\tilde{x}_n^a, \tilde{y}_s^a) \, ds \right),$$

(A.53)
$$K_6 := \delta \sum_{n=0}^{k-1} \left(\bar{g}(\bar{x}_{n\delta}) - \bar{g}(\tilde{x}_n^a) \right).$$

Using the regularity of g, we obtain

(A.54)
$$|K_6| \le \delta k C \delta m.$$

Arranging the right-hand side of (A.51) into groups of m terms corresponding to the intervals of (A.47), we obtain, from Conditions 1.2 and 1.3,

(A.55)
$$|K_5| \le Ck\delta E \left(\frac{m\tau}{\epsilon} \right).$$

Using (A.48) and the regularity of f and g, we obtain

(A.56)
$$|\tilde{y}_{\frac{\delta}{\tau}t}^a - \tilde{y}_t| \le Cm\delta e^{C\frac{m\tau}{\epsilon}}.$$

It follows that

(A.57)
$$|K_4| \le C\delta km\delta e^{C\frac{m\tau}{\epsilon}}.$$

Similarly, using (A.7) and (A.8), we obtain

(A.58)
$$|\tilde{y}_{n\delta} - \bar{y}_{n\delta}| \le C \left(\frac{\tau}{\epsilon} + m\delta + m \left(\frac{\tau}{\epsilon} \right)^2 \right) \frac{m\tau}{\epsilon} e^{C\frac{m\tau}{\epsilon}},$$

(A.59)
$$|\tilde{x}_{n\delta} - \bar{x}_{n\delta}| \le Cm \left(\delta + \left(\frac{\tau}{\epsilon} \right)^2 \right).$$

It follows that

(A.60) $$|K_3| \leq C\delta k \left(\frac{\tau}{\epsilon} + m\delta + m\left(\frac{\tau}{\epsilon}\right)^2\right) e^{C\frac{m\tau}{\epsilon}}.$$

This concludes the proof of Lemma A.4. $\square$

Combining Lemmas A.2, A.3, and A.4, we obtain

(A.61)
$$|x_{n\delta}^\epsilon - \bar{x}_{n\delta}| \leq Ce^{C\delta n}\left(\delta + \left(\frac{\tau}{\epsilon}\right)^2\frac{1}{\delta} + \delta e^{C\frac{\delta}{N\epsilon}} + E\left(\frac{\delta}{N\epsilon}\right) + E\left(\frac{m\tau}{\epsilon}\right)\right.$$
$$\left. + \left(\frac{\tau}{\epsilon} + m\delta + m\left(\frac{\tau}{\epsilon}\right)^2\right)e^{C\frac{m\tau}{\epsilon}}\right).$$

Choosing N such that $e^{C\frac{\delta}{N\epsilon}} \sim \delta^{-\frac{1}{2}}$ (observe that we need $\epsilon \leq \delta/(-C\ln\delta)$) and m such that $\frac{m\tau}{\epsilon}e^{C\frac{m\tau}{\epsilon}} \sim \left(\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon}\right)^{-\frac{1}{2}}$, we obtain for $\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon} \leq 1$

(A.62)
$$|x_{n\delta}^\epsilon - \bar{x}_{n\delta}| \leq Ce^{C\delta n}\left(\sqrt{\delta} + \left(\frac{\tau}{\epsilon}\right)^2\frac{1}{\delta} + E\left(\frac{1}{C}\ln\frac{1}{\delta}\right)\right.$$
$$\left. + \left(\frac{\delta\epsilon}{\tau}\right)^{\frac{1}{2}} + \left(\frac{\tau}{\epsilon}\right)^{\frac{1}{2}} + E\left(\frac{1}{C}\ln\left(\left(\frac{\delta\epsilon}{\tau} + \frac{\tau}{\epsilon}\right)^{-1}\right)\right)\right).$$

This concludes the proof of inequality (A.2). The proof of (A.3) is similar and is also a consequence of (A.2).

A.2. Proof of Theorem 3.2. Define the process $t \mapsto (\bar{x}_t, \bar{y}_t)$ by

(A.63) $$(\bar{x}_t, \bar{y}_t) := \eta(\bar{u}_t).$$

It follows from the regularity of η that it is sufficient to prove the F-convergence of $(\bar{x}_t, \bar{y}_t)$ towards $\delta_{X_t} \otimes \mu(X_t, dy)$. Now define ψ_τ^ϵ by

(A.64) $$\psi_\tau^\epsilon(x, y, \omega) := \eta \circ \theta_\tau^\epsilon(., \omega) \circ \eta^{-1}(x, y).$$

Define ψ_h^g by

(A.65) $$\psi_h^g(x, y, \omega) := \eta \circ \theta_h^G(., \omega) \circ \eta^{-1}(x, y).$$

PROPOSITION A.5. *The vector fields f, g and matrix fields σ, Q associated with the system of equations (3.2) are uniformly bounded and Lipschitz continuous. We also have*

(A.66)
$$\begin{cases} (\bar{x}_0, \bar{y}_0) = \eta(u_0), \\ (\bar{x}_{(k+1)\delta}, \bar{y}_{(k+1)\delta}) = \psi_{\delta-\tau}^g(., \omega_k') \circ \psi_\tau^\epsilon((\bar{x}_{k\delta}, \bar{y}_{k\delta}), \omega_k), \\ (\bar{x}_t, \bar{y}_t) = (\bar{x}_{k\delta}, \bar{y}_{k\delta}) \quad for \quad k\delta \leq t < (k+1)\delta, \end{cases}$$

where ω_k, ω_k' are i.i.d. samples from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, there exist $C > 0$ and d-dimensional centered Gaussian vectors $\xi'(\omega)$, $\xi''(\omega)$ with identity covariance matrices such that for $h \leq h_0$ and $\frac{\tau}{\epsilon} \leq \tau_0$ we have

(A.67) $$\left(\mathbb{E}\left[\left|\psi_h^g(x, y, \omega) - (x, y) - h(g(x, y), 0) - \sqrt{h}(\sigma(x, y)\xi'(\omega), 0)\right|^2\right]\right)^{\frac{1}{2}} \leq Ch^{\frac{3}{2}},$$

(A.68)
$$\left(\mathbb{E}\left[\left|\psi_\tau^\epsilon(x,y,\omega) - (x,y) - \tau\big(g(x,y),0\big) - \frac{\tau}{\epsilon}\big(0,f(x,y)\big) - \sqrt{\tau}\big(\sigma(x,y)\xi''(\omega),0\big)\right.\right.\right.$$
$$\left.\left.\left. - \sqrt{\frac{\tau}{\epsilon}}\big(0,Q(x,y)\xi''(\omega)\big)\right|^2\right]\right)^{\frac{1}{2}} \le C\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}.$$

Proof. Since $(x,y) = \eta(u)$, we obtain from (3.1) and Itô's formula

(A.69)
$$dx = \left(\left(G + \frac{1}{\epsilon}F\right)\nabla\eta^x \circ \eta^{-1}(x,y)\right)dt + \left(\nabla\eta^x\left(H + \frac{1}{\sqrt{\epsilon}}K\right)\right)\circ \eta^{-1}(x,y)\,dW_t$$
$$+ \frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^x\left(\left(H + \frac{1}{\sqrt{\epsilon}}K\right)\left(H + \frac{1}{\sqrt{\epsilon}}K\right)^T\right)_{ij}dts,$$

(A.70)
$$dy = \left(\left(G + \frac{1}{\epsilon}F\right)\nabla\eta^y \circ \eta^{-1}(x,y)\right)dt + \left(\nabla\eta^y\left(H + \frac{1}{\sqrt{\epsilon}}K\right)\right)\circ \eta^{-1}(x,y)\,dW_t$$
$$+ \left(\frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^y\left(\left(H + \frac{1}{\sqrt{\epsilon}}K\right)\left(H + \frac{1}{\sqrt{\epsilon}}K\right)^T\right)_{ij}\right)\circ \eta^{-1}\,dt.$$

Hence we deduce from (3.2) of Condition 3.1 that

(A.71)
$$g(x,y) = \left(G\nabla\eta^x + \frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^x(HH^T)_{ij}\right)\circ \eta^{-1}(x,y),$$

(A.72)
$$\sigma(x,y) = \big(\nabla\eta^x H\big)\circ \eta^{-1}(x,y),$$

(A.73)
$$f(x,y) = \left(F\nabla\eta^y + \frac{1}{2}\sum_{ij}\partial_i\partial_j\eta^y(KK^T)_{ij}\right)\circ \eta^{-1}(x,y),$$

(A.74)
$$Q(x,y) = \big(\nabla\eta^y K\big)\circ \eta^{-1}(x,y).$$

REMARK A.1. *Observe that (3.2) of Condition 3.1 requires that*

(A.75)
$$F\nabla\eta^x = 0, \qquad G\nabla\eta^y = 0,$$

(A.76)
$$\sum_{ij}\partial_i\partial_j\eta^x\big(KK^T\big)_{ij} = 0,$$

(A.77)
$$\sum_{ij}\partial_i\partial_j\eta^y\big(HH^T\big)_{ij} = 0,$$

(A.78)
$$\sum_{ij}\partial_i\partial_j\eta^x\big(KH^T + HK^T\big)_{ij} = 0,$$

and

(A.79)
$$\sum_{ij}\partial_i\partial_j\eta^y\big(KH^T + HK^T\big)_{ij} = 0.$$

Equations (A.78) *and* (A.79) *are satisfied if* KH^T *is skew-symmetric. One particular case could be, of course,* $KH^T = 0$, *which translates into the fact that for all u the ranges of* $H(u)$ *and* $K(u)$ *are orthogonal; i.e., the noise with amplitude* $1/\sqrt{\epsilon}$ *is applied to degrees of freedom orthogonal to those with* $\mathcal{O}(1)$ *noise.*

We deduce the regularity of f, g, σ, and Q from the regularity of G, F, H, K, and η. Equation (A.6) is a direct consequence of the definition of ψ_τ^ϵ and ψ_h^g and (A.66). Now observe that

(A.80)

$$
\psi_\tau^\epsilon(x, y, \omega) - (x, y) - \tau\big(g(x,y), 0\big) - \frac{\tau}{\epsilon}\big(0, f(x,y)\big) - \sqrt{\tau}\big(\sigma(x,y)\xi'(\omega), 0\big)
$$

$$
- \sqrt{\frac{\tau}{\epsilon}}\big(0, Q(x,y)\xi'(\omega)\big) = \left(\eta \circ \theta_\tau^\epsilon - \eta - \tau\left(G\nabla\eta^x + \frac{1}{2}\sum_{ij} \partial_i\partial_j\eta^x (HH^T)_{ij}, 0 \right) \right.
$$

$$
- \frac{\tau}{\epsilon}\left(0, F\nabla\eta^y + \frac{1}{2}\sum_{ij} \partial_i\partial_j\eta^y (KK^T)_{ij} \right) - \sqrt{\tau}\big(\nabla\eta^x H\xi'(\omega), 0\big)
$$

$$
\left. - \sqrt{\frac{\tau}{\epsilon}}\big(0, \nabla\eta^y K\xi'(\omega)\big) \right) \circ \eta^{-1}(x, y).
$$

Using (A.75), (A.76), (A.77), (A.78), and (A.79), the Taylor–Itô expansion of $\eta \circ \theta_\tau^\epsilon$, and the regularity of η, and setting ξ' equal to ξ defined in (3.14), we obtain (A.68). The proof of (A.67) is similar. $\square$

It follows from Proposition A.5 that it is sufficient to prove Theorem 3.2 in the situation where η is the identity diffeomorphism. More precisely, the F-convergence of $\bar{u}_t$ is a consequence of the F-convergence of $(\bar{x}_t, \bar{y}_t)$ and the regularity of η.

Let $x \mapsto \varphi(x)$ be a function with continuous and bounded derivatives up to order 3. Let us prove the following lemma.

LEMMA A.6. *We have*

(A.81)
$$
\mathbb{E}\big[\varphi(\bar{x}_{(n+1)\delta})\big] - \mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big]
$$
$$
= \delta\mathbb{E}\Big[g(\bar{x}_{n\delta}, \bar{y}_{n\delta})\nabla\varphi(\bar{x}_{n\delta}) + \sigma\sigma^T(\bar{x}_{n\delta}, \bar{y}_{n\delta}) : \mathrm{Hess}\,\varphi(\bar{x}_{n\delta})\Big] + I_0
$$

with

(A.82)
$$
|I_0| \leq C\left(\delta^{\frac{3}{2}} + \left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}} \right).
$$

Proof. Write $(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau}) := \psi_\tau^\epsilon(\bar{x}_{n\delta}, \bar{y}_{n\delta}, \omega_n)$. Using (A.68), we obtain that there exists an $\mathcal{N}(0,1)$ random vector ξ_n, independent from $(\bar{x}_{n\delta}, \bar{y}_{n\delta})$, such that

(A.83)
$$
\bar{x}_{n\delta+\tau} - \bar{x}_{n\delta} = g(\bar{x}_{n\delta})\tau + \sqrt{\tau}\sigma(\bar{x}_{n\delta}, \bar{y}_{n\delta})\xi_n + I_1
$$

with

(A.84)
$$
\big(\mathbb{E}[(I_1)^2]\big)^{\frac{1}{2}} \leq C\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}.
$$

Hence

(A.85)
$$
\left| \mathbb{E}\big[\varphi(\bar{x}_{n\delta+\tau})\big] - \mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] - \tau\mathbb{E}\Big[g(\bar{x}_{n\delta}, \bar{y}_{n\delta})\nabla\varphi(\bar{x}_{n\delta}) \right.
$$
$$
\left. + \sigma\sigma^T(\bar{x}_{n\delta}, \bar{y}_{n\delta}) : \mathrm{Hess}\,\varphi(\bar{x}_{n\delta})\Big] \right| \leq C\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}.
$$

Similarly, using (A.67), we obtain that there exists an $\mathcal{N}(0,1)$ random vector ξ'_n, independent from $(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau})$, such that

$$(A.86) \quad \bar{x}_{(n+1)\delta} - \bar{x}_{n\delta+\tau} = g(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau})(\delta - \tau) + \sigma(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau})\sqrt{\delta - \tau}\xi'_n + I_2$$

with

$$(A.87) \qquad \qquad \left(\mathbb{E}[(I_2)^2]\right)^{\frac{1}{2}} \leq C(\delta - \tau)^{\frac{3}{2}}.$$

Whence

$$(A.88) \quad \left| \mathbb{E}\big[\varphi(\bar{x}_{(n+1)\delta})\big] - \mathbb{E}\big[\varphi(\bar{x}_{n\delta+\tau})\big] - (\delta - \tau)\mathbb{E}\Big[g(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau})\nabla\varphi(\bar{x}_{n\delta+\tau}) \right.$$
$$\left. + \sigma\sigma^T(\bar{x}_{n\delta+\tau}, \bar{y}_{n\delta+\tau}) : \operatorname{Hess}\varphi(\bar{x}_{n\delta+\tau})\Big]\right| \leq C(\delta - \tau)^{\frac{3}{2}}.$$

Using the regularity of σ, we obtain

$$(A.89) \qquad \left(\mathbb{E}\Big[\big|\sigma(\bar{x}_{n\delta+\tau}, \bar{y}_{(n+1)\delta}) - \sigma(\bar{x}_{n\delta}, \bar{y}_{n\delta})\big|^2\Big]\right)^{\frac{1}{2}} \leq C\left(\delta^{\frac{1}{2}} + \sqrt{\frac{\tau}{\epsilon}}\right).$$

The proof of (A.81) follows from (A.68), (A.85), (A.88), (A.89), and the regularity of g and φ. $\quad\square$

LEMMA A.7. *We have*

$$(A.90) \qquad \left|\frac{\mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] - \varphi(x_0)}{n\delta} - L\varphi(x_0)\right| \leq J_5$$

with (for $\delta \leq C\tau/\epsilon$)

$$(A.91) \qquad |J_5| \leq C\left(\left(\frac{\delta\epsilon}{\tau}\right)^{\frac{1}{4}} + \left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\frac{1}{\delta} + \sqrt{\frac{\tau}{\epsilon}}\right) + CE\left(\frac{1}{C}\ln\frac{\tau}{\delta\epsilon}\right).$$

Proof. Define $\hat{B}_t$ by $\hat{B}_0 = 0$ and

$$(A.92) \qquad \hat{B}_t - \hat{B}_{n\tau} = B_{n\delta+t} - B_{n\delta} \quad \text{for} \quad n\tau \leq t \leq (n+1)\tau.$$

Define $\tilde{y}_s$ by $\tilde{y}_0 = y_0$ and

$$(A.93) \qquad d\tilde{y}_t = \frac{1}{\epsilon}f(x_0, \tilde{y}_t)\,dt + \frac{1}{\sqrt{\epsilon}}Q(x_0, \tilde{y}_t)d\hat{B}_t.$$

Write

$$(A.94) \qquad \bar{g}(x_0) := \int g(x_0, y)\,\mu(x_0, dy).$$

Using Lemma A.6, we obtain

$$(A.95) \qquad \frac{\mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] - \varphi(x_0)}{n\delta} = L\varphi(x_0) + J_1 + J_2 + J_3 + J_4$$

with

$$(A.96) \qquad L\varphi(x_0) := \bar{g}(x_0)\nabla\varphi(x_0) + \bar{\sigma}\bar{\sigma}^T(x_0) : \text{Hess}\,\varphi(x_0),$$

$$(A.97) \qquad \begin{aligned} J_1 &= \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\Big[g(\bar{x}_{k\delta},\bar{y}_{k\delta})\nabla\varphi(\bar{x}_{k\delta}) + \sigma\sigma^T(\bar{x}_{k\delta},\bar{y}_{k\delta}) : \text{Hess}\,\varphi(\bar{x}_{k\delta})\Big] \\ &\quad - \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\Big[g(\bar{x}_0,\bar{y}_{k\delta})\nabla\varphi(\bar{x}_0) + \sigma\sigma^T(\bar{x}_0,\bar{y}_{k\delta}) : \text{Hess}\,\varphi(\bar{x}_0)\Big], \end{aligned}$$

$$(A.98) \qquad \begin{aligned} J_2 &= \frac{1}{n}\sum_{k=0}^{n-1}\bigg(\mathbb{E}\Big[g(\bar{x}_0,\bar{y}_{k\delta})\nabla\varphi(\bar{x}_0) + \sigma\sigma^T(\bar{x}_0,\bar{y}_{k\delta}) : \text{Hess}\,\varphi(\bar{x}_0)\Big] \\ &\quad - \frac{1}{\tau}\int_{k\tau}^{(k+1)\tau}\mathbb{E}\Big[g(x_0,\tilde{y}_s)\nabla\varphi(x_0) + \sigma\sigma^T(x_0,\tilde{y}_s) : \text{Hess}\,\varphi(x_0)\Big]\,ds\bigg), \end{aligned}$$

$$(A.99) \qquad J_3 = \frac{1}{n\tau}\int_0^{n\tau}\mathbb{E}\Big[g(x_0,\tilde{y}_s)\nabla\varphi(x_0) + \sigma\sigma^T(x_0,\tilde{y}_s) : \text{Hess}\,\varphi(x_0)\Big]\,ds - L\varphi(x_0),$$

$$(A.100) \qquad |J_4| \le C\left(\delta^{\frac{1}{2}} + \left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\frac{1}{\delta}\right).$$

Using the regularity of σ, g, φ, (A.6), and (A.7), we obtain

$$(A.101) \qquad |J_1| \le C\left((n\delta)^{\frac{1}{2}} + n\delta + n\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\right).$$

Using property 3 of Condition 3.1 and property 3 of Condition 3.2, we obtain

$$(A.102) \qquad |J_3| \le CE\left(\frac{n\tau}{\epsilon}\right).$$

Using (A.67) and (A.68), we obtain

$$(A.103) \qquad \left(\mathbb{E}\Big[\big|\bar{y}_{n\delta} - \tilde{y}_{n\tau}\big|^2\Big]\right)^{\frac{1}{2}} \le C\left(\sqrt{\frac{\tau}{\epsilon}} + (n\delta)^{\frac{1}{2}} + n\delta + n\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\right)\frac{n\tau}{\epsilon}e^{C\frac{n\tau}{\epsilon}},$$

which leads to

$$(A.104) \qquad |J_2| \le C\left(\sqrt{\frac{\tau}{\epsilon}} + (n\delta)^{\frac{1}{2}} + n\delta + n\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\right)e^{C\frac{n\tau}{\epsilon}}.$$

Hence, we obtain

$$(A.105) \qquad \left|\frac{\mathbb{E}\big[\varphi(\bar{x}_{n\delta})\big] - \varphi(x_0)}{n\delta} - L\varphi(x_0)\right| \le J_5$$

with

$$(A.106) \qquad |J_5| \le C\left(\sqrt{\frac{\tau}{\epsilon}} + (n\delta)^{\frac{1}{2}} + n\delta + n\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\right)e^{C\frac{n\tau}{\epsilon}} + E\left(\frac{n\tau}{\epsilon}\right) + C\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}}\frac{1}{\delta}.$$

Choosing n such that $\sqrt{\frac{n\tau}{\epsilon}}e^{C\frac{n\tau}{\epsilon}} \sim \left(\frac{\tau}{\epsilon\delta}\right)^{\frac{1}{4}}$, we obtain (A.91) for $\delta \le C\tau/\epsilon$. $\qquad \square$

We now combine Lemma A.7 with Theorem 1 of Chapter 2 of [108], which states that the uniform convergence (in x_0, y_0) of $\frac{\mathbb{E}[\varphi(\bar{x}_{n\delta})] - \varphi(x_0)}{n\delta}$ to $L\varphi(x_0)$ as $\epsilon \downarrow 0$, $\tau \leq \delta$, $\frac{\tau}{\epsilon} \downarrow 0$, $\frac{\delta\epsilon}{\tau} \downarrow 0$, and $\left(\frac{\tau}{\epsilon}\right)^{\frac{3}{2}} \frac{1}{\delta} \downarrow 0$ implies the convergence in distribution of $\bar{x}_{n\delta}$ to the Markov process generated by L.

The F-convergence of $(\bar{x}_t, \bar{y}_t)$ can be deduced from the convergence in distribution of $\bar{x}_t$ and (3.4) of Condition 3.1. The proof follows the same lines as above, which will not be repeated here.

Acknowledgments. We are grateful to S. Flach, C. Lebris, J. M. Sanz-Serna, E. S. Titi, R. Tsai, and E. Vanden-Eijnden for useful comments and providing references. We would also like to thank two anonymous referees for precise and detailed comments and suggestions.

REFERENCES

[1] A. ABDULLE, *Fourth order Chebyshev methods with recurrence relation*, SIAM J. Sci. Comput., 23 (2002), pp. 2041–2054.

[2] A. ABDULLE AND S. CIRILLI, *S-ROCK: Chebyshev methods for stiff stochastic differential equations*, SIAM J. Sci. Comput., 30 (2008), pp. 997–1014.

[3] G. ALLAIRE, *Homogenization and two-scale convergence*, SIAM J. Math. Anal., 23 (1992), pp. 1482–1518.

[4] H. ANDERSON, *RATTLE: A velocity version of the SHAKE algorithm for molecular dynamics calculations*, J. Comput. Phys., 52 (1983), pp. 24–34.

[5] G. ARIEL, B. ENGQUIST, AND Y.-H. TSAI, *A multiscale method for highly oscillatory ordinary differential equations with resonance*, Math. Comp., 78 (2009), pp. 929–956.

[6] G. ARIEL, B. ENGQUIST, AND Y.-H. TSAI, *A reversible multiscale integration method*, Commun. Math. Sci., 7 (2009), pp. 595–610.

[7] A. ARMAOU AND I. KEVREKIDIS, *Equation-free optimal switching policies for bistable reacting systems*, Internat. J. Robust Nonlinear Control, 15 (2005), pp. 713–726.

[8] S. ARTEM′EV AND K. SHURTS, *Zhestkie sistemy stokhasticheskikh differentsial'nykh uravneniĭ s malym shumom i ikh chislennoe reshenie*, Preprint 1039, Ross. Akad. Nauk Sibirsk. Otdel., Vychisl. Tsentr, Novosibirsk, 1995.

[9] Z. ARTSTEIN, I. G. KEVREKIDIS, M. SLEMROD, AND E. S. TITI, *Slow observables of singularly perturbed differential equations*, Nonlinearity, 20 (2007), pp. 2463–2481.

[10] Z. ARTSTEIN, J. LINSHIZ, AND E. S. TITI, *Young measure approach to computing slowly advancing fast oscillations*, Multiscale Model. Simul., 6 (2007), pp. 1085–1097.

[11] A. BENSOUSSAN, J.-L. LIONS, AND G. PAPANICOLAOU, *Asymptotic Analysis for Periodic Structures*, North–Holland, Amsterdam, 1978.

[12] J. BISMUT, *Mécanique aléatoire*, Springer-Verlag, Berlin, New York, 1981.

[13] N. BOGOLYUBOV, *Problemy dinamičeskoĭ teorii v statističeskoĭ fizike*, Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow, Leningrad, 1946.

[14] F. A. BORNEMANN AND C. SCHÜTTE, *A Mathematical Approach to Smoothed Molecular Dynamics: Correcting Potentials for Freezing Bond Angles*, Preprint SC 9-30, Zuse Institute Berlin, Berlin, 1995.

[15] F. A. BORNEMANN AND C. SCHÜTTE, *Homogenization of Hamiltonian systems with a strong constraining potential*, Phys. D, 102 (1997), pp. 57–77.

[16] N. BOU-RABEE AND J. MARSDEN, *Hamilton-Pontryagin integrators on Lie groups*. I. *Introduction and structure-preserving properties*, Found. Comput. Math., 9 (2009), pp. 197–219.

[17] N. BOU-RABEE AND H. OWHADI, *Long-run accuracy of variational integrators in the stochastic context*, SIAM J. Numer. Anal., 48 (2010), pp. 278–297.

[18] N. BOU-RABEE AND H. OWHADI, *Stochastic variational integrators*, IMA J. Numer. Anal., 29 (2009), pp. 421–443.

[19] N. BOU-RABEE AND E. VANDEN-EIJNDEN, *Pathwise accuracy and ergodicity of metropolized integrators for SDEs*, Comm. Pure Appl. Math., 63 (2010), pp. 655–696.

[20] C. LE BRIS AND F. LEGOLL, *Integrators for highly oscillatory Hamiltonian systems: An homogenization approach*, Discrete Contin. Dyn. Syst. Ser. B, 13 (2010), pp. 347–373.

[21] B. BROOKS, R. BRUCCOLERI, B. OLAFSON, D. STATES, S. SWAMINATHAN, AND M. KARPLUS, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*, J. Comput. Chem., 4 (1983), pp. 187–217.

[22] K. Burrage and T. Tian, *Stiffly accurate Runge-Kutta methods for stiff stochastic differential equations*, Comput. Phys. Comm., 142 (2001), pp. 186–190.

[23] K. Burrage and T. Tian, *The composite Euler method for stiff stochastic differential equations*, J. Comput. Appl. Math., 131 (2001), pp. 407–426.

[24] M. P. Calvo and J. M. Sanz-Serna, *Heterogeneous multiscale methods for mechanical systems with vibrations*, SIAM J. Sci. Comput., 32 (2010), pp. 2029–2046.

[25] M. P. Calvo and J. M. Sanz-Serna, *Instabilities and inaccuracies in the integration of highly oscillatory problems*, SIAM J. Sci. Comput., 31 (2009), pp. 1653–1677.

[26] F. Castella, P. Chartier, and E. Faou, *An averaging technique for highly oscillatory Hamiltonian problems*, SIAM J. Numer. Anal., 47 (2009), pp. 2808–2837.

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

[28] A. Chorin, O. Hald, and R. Kupferman, *Optimal prediction with memory*, Phys. D, 166 (2002), pp. 239–257.

[29] A. Chorin, A. Kast, and R. Kupferman, *Optimal prediction of underresolved dynamics*, Proc. Natl. Acad. Sci. USA, 95 (1998), pp. 4094–4098.

[30] G. Ciccotti, T. Lelievre, and E. Vanden-Eijnden, *Projections of diffusions on submanifolds: Application to mean force computation*, Comm. Pure Appl. Math., 61 (2008), pp. 371–408.

[31] D. Cohen, T. Jahnke, K. Lorenz, and C. Lubich, *Numerical integrators for highly oscillatory Hamiltonian systems: A review*, in Analysis, Modeling and Simulation of Multiscale Problems, Springer-Verlag, Berlin, 2006, pp. 553–576.

[32] M. Condon, A. Deaño, and A. Iserles, *On highly oscillatory problems arising in electronic engineering*, M2AN Math. Model. Numer. Anal., 43 (2009), pp. 785–804.

[33] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, and P. Kollman, *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*, J. Am. Chem. Soc., 117 (1995), pp. 5179–5197.

[34] F. Cucker and A. G. Corbalan, *An alternate proof of the continuity of the roots of a polynomial*, Amer. Math. Monthly, 96 (1989), pp. 342–345.

[35] G. Dahlquist, *Convergence and stability in the numerical integration of ordinary differential equations*, Math. Scand., 4 (1956), pp. 33–53.

[36] W. E, *Analysis of the heterogeneous multiscale method for ordinary differential equations*, Commun. Math. Sci., 1 (2003), pp. 423–436.

[37] 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.

[38] W. E, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.

[39] W. E, D. Liu, and E. Vanden-Eijnden, *Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales*, J. Comput. Phys., 221 (2007), pp. 158–180.

[40] W. E, W. Ren, and E. Vanden-Eijnden, *A general strategy for designing seamless multiscale methods*, J. Comput. Phys., 228 (2009), pp. 5437–5453.

[41] B. Engquist and Y.-H. Tsai, *Heterogeneous multiscale methods for stiff ordinary differential equations*, Math. Comp., 74 (2005), pp. 1707–1742.

[42] L. Evans, *A survey of partial differential equations methods in weak KAM theory*, Comm. Pure Appl. Math., 57 (2004), pp. 445–480.

[43] E. Fermi, J. Pasta, and S. Ulam, *Studies of Nonlinear Problems*, Tech. report LA-1940, Los Alamos Scientific Laboratory, Los Alamos, NM, 1955.

[44] M. Fixman, *Classical statistical mechanics of constraints: A theorem and application to polymers*, Proc. Nat. Acad. Sci. U.S.A., 71 (1974), pp. 3050–3053.

[45] S. Flach, M. V. Ivanchenko, and O. I. Kanakov, *q-breathers and the Fermi-Pasta-Ulam problem*, Phys. Rev. Lett., 95 (2005), 064102.

[46] J. Ford, *The Fermi-Pasta-Ulam problem: Paradox turns discovery*, Phys. Rep., 213 (1992), pp. 271–310.

[47] B. García-Archilla, J. M. Sanz-Serna, and R. D. Skeel, *Long-time-step methods for oscillatory differential equations*, SIAM J. Sci. Comput., 20 (1998), pp. 930–963.

[48] W. Gautschi, *Numerical integration of ordinary differential equations based on trigonometric polynomials*, Numer. Math., 3 (1961), pp. 381–397.

[49] C. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice–Hall, Englewood Cliffs, NJ, 1971.

[50] C. Gear and K. Gallivan, *Automatic methods for highly oscillatory ordinary differential equations*, in Numerical Analysis (Dundee, 1981), Lecture Notes in Math. 912, Springer-Verlag, Berlin, 1982, pp. 115–124.

[51] C. W. GEAR AND I. G. KEVREKIDIS, *Projective methods for stiff differential equations: Problems with gaps in their eigenvalue spectrum*, SIAM J. Sci. Comput., 24 (2003), pp. 1091–1106.

[52] I. GIHMAN, *On the theory of differential equations of stochastic processes. I, II*, Amer. Math. Soc. Transl. (2), 1 (1955), pp. 111–137, 139–161.

[53] D. GIVON, I. G. KEVREKIDIS, AND R. KUPFERMAN, *Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems*, Commun. Math. Sci., 4 (2006), pp. 707–729.

[54] H. GRUBMULLER, H. HELLER, A. WINDEMUTH, AND K. SCHULTEN, *Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions*, Mol. Simul., 6 (1991), pp. 121–142.

[55] S. GUSEV, *Algoritm peremennogo shaga dlya chislennogo resheniya zhestkikh sistem stokhasticheskikh differentsial'nykh uravneniĭ*, Preprint 1094, Rossiĭskaya Akademiya Nauk Sibirskoe Otdelenie, Institut Vychislitel'noĭ Matematiki i Matematicheskoĭ Geofiziki, Novosibirsk, 1997.

[56] E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric numerical integration illustrated by the Störmer-Verlet method*, Acta Numer., 12 (2003), pp. 399–450.

[57] E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd ed., Springer-Verlag, Heidelberg, 2004.

[58] E. HAIRER, S. NØRSETT, AND G. WANNER, *Solving Ordinary Differential Equations. I. Nonstiff Problems*, 2nd ed., Springer Ser. Comput. Math. 8, Springer-Verlag, Berlin, 1993.

[59] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations. II. Stiff and Differential-Algebraic Problems*, 2nd ed., Springer Ser. Comput. Math. 14, Springer-Verlag, Berlin, 1996.

[60] C. HARTMANN, *An ergodic sampling scheme for constrained Hamiltonian systems with applications to molecular dynamics*, J. Stat. Phys., 130 (2008), pp. 687–711.

[61] A. ISERLES, *On the global error of discretization methods for highly-oscillatory ordinary differential equations*, BIT, 42 (2002), pp. 561–599.

[62] A. ISERLES, *On the numerical analysis of rapid oscillation*, in Group Theory and Numerical Analysis, CRM Proc. Lecture Notes 39, AMS, Providence, RI, 2005, pp. 149–163.

[63] A. ISERLES, *Think globally, act locally: Solving highly-oscillatory ordinary differential equations*, Appl. Numer. Math., 43 (2002), pp. 145–160.

[64] A. ISERLES, S. P. NØRSETT, AND S. OLVER, *Highly oscillatory quadrature: The story so far*, in Numerical Mathematics and Advanced Applications, Springer-Verlag, Berlin, 2006, pp. 97–118.

[65] V. JIKOV, S. KOZLOV, AND O. OLEINĬK, *Homogenization of Differential Operators and Integral Functionals*, Springer-Verlag, Berlin, 1994.

[66] P. KAPITZA, *Collected Papers of P.L. Kapitza*, Vol. II, D. Ter Haar, ed., Pergamon Press, Oxford, UK, 1965.

[67] I. KEVREKIDIS, C. GEAR, J. HYMAN, P. KEVREKIDIS, O. RUNBORG, AND C. THEODOROPOULOS, *Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis*, Commun. Math. Sci., 1 (2003), pp. 715–762.

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

[69] H.-O. KREISS, *Problems with different time scales*, in Acta Numerica, 1992, Cambridge University Press, Cambridge, UK, 1992, pp. 101–139.

[70] B. KRYLOFF AND N. BOGOLIOUBOFF, *La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire*, Ann. of Math. (2), 38 (1937), pp. 65–113.

[71] N. KRYLOFF AND N. BOGOLIOUBOFF, *On some problems in the ergodic theory of stochastic systems*, Zap. Kafedr. Mat. Fiz. Inst. Budivel. Mat. Akad. Nauk Ukrain. SSR, 4 (1939), pp. 243–287.

[72] J.-A. LÁZARO-CAMÍ AND J. ORTEGA, *Stochastic Hamiltonian dynamical systems*, Rep. Math. Phys., 61 (2008), pp. 65–122.

[73] V. LEBEDEV AND S. FINOGENOV, *The use of ordered Čebyšev parameters in iteration methods*, Ž. Vyčisl. Mat. i Mat. Fiz., 16 (1976), pp. 895–907, 1084.

[74] B. LEIMKUHLER AND S. REICH, *A reversible averaging integrator for multiple time-scale dynamics*, J. Comput. Phys., 171 (2001), pp. 95–114.

[75] B. LEIMKUHLER AND S. REICH, *Simulating Hamiltonian Dynamics*, Cambridge Monogr. Appl. Comput. Math. 14, Cambridge University Press, Cambridge, UK, 2004.

[76] B. J. LEIMKUHLER, S. REICH, AND R. D. SKEEL, *Integration methods for molecular dynamics*,

in Mathematical Approaches to Biomolecular Structure and Dynamics (Minneapolis, MN, 1994), IMA Vol. Math. Appl. 82, Springer-Verlag, New York, 1996, pp. 161–185.

[77] A. Lew, J. Marsden, M. Ortiz, and M. West, *Asynchronous variational integrators*, Arch. Ration. Mech. Anal., 167 (2003), pp. 85–146.

[78] A. Lew, J. Marsden, M. Ortiz, and M. West, *Variational time integrators*, Internat. J. Numer. Methods Engrg., 60 (2004), pp. 153–212.

[79] T. Li, A. Abdulle, and W. E, *Effectiveness of implicit methods for stiff stochastic differential equations*, Commun. Comput. Phys., 3 (2008), pp. 295–307.

[80] R. S. MacKay and S. Aubry, *Proof of existence of breathers for time-reversible or Hamiltonian networks of weakly coupled oscillators*, Nonlinearity, 7 (1994), pp. 1623–1643.

[81] S. J. A. Malham and A. Wiese, *Stochastic Lie group integrators*, SIAM J. Sci. Comput., 30 (2008), pp. 597–617.

[82] J. Marsden and M. West, *Discrete mechanics and variational integrators*, Acta Numer., 10 (2001), pp. 357–514.

[83] R. McLachlan and D. O'Neale, *Comparison of Integrators for the Fermi-Pasta-Ulam Problem*, Preprint NI07052-HOP, Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, 2007; also available online from http://www.newton.ac.uk/preprints/NI07052.pdf.

[84] R. McLachlan and M. Perlmutter, *Conformal Hamiltonian systems*, J. Geom. Phys., 39 (2001), pp. 276–300.

[85] R. McLachlan, G. Reinout, and W. Quispel, *Splitting methods*, Acta Numer., 11 (2002), pp. 341–434.

[86] G. N. Milstein, Yu. M. Repin, and M. V. Tretyakov, *Numerical methods for stochastic systems preserving symplectic structure*, SIAM J. Numer. Anal., 40 (2002), pp. 1583–1604.

[87] G. N. Milstein, Yu. M. Repin, and M. V. Tretyakov, *Symplectic integration of Hamiltonian systems with additive noise*, SIAM J. Numer. Anal., 39 (2002), pp. 2066–2088.

[88] G. Milstein and M. Tretyakov, *Quasi-symplectic methods for Langevin-type equations*, IMA J. Numer. Anal., 23 (2003), pp. 593–626.

[89] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer-Verlag, Berlin, 2004.

[90] G. Nguetseng, *A general convergence result for a functional related to the theory of homogenization*, SIAM J. Math. Anal., 20 (1989), pp. 608–623.

[91] G. Papanicolaou and W. Kohler, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math., 27 (1974), pp. 641–668.

[92] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Texts Appl. Math. 53, Springer-Verlag, New York, 2008.

[93] D. Perchak, J. Skolnick, and R. Yaris, *Dynamics of rigid and flexible constraints for polymers. Effect of the Fixman potential*, Macromolecules, 18 (1985), pp. 519–525.

[94] L. Petzold, L. Jay, and J. Yen, *Problems with different time scales*, in Acta Numerica, 1997, Cambridge University Press, Cambridge, UK, 1997, pp. 437–483.

[95] S. Reich, *Smoothed Langevin dynamics of highly oscillatory systems*, Phys. D, 138 (2000), pp. 210–224.

[96] R. Rosenberg, B. Berne, and D. Chandler, *Isomerization dynamics in liquids by molecular dynamics*, Chem. Phys. Lett., 75 (1980), pp. 162–168.

[97] J. Ryckaert and A. Bellemans, *Molecular dynamics of liquid n-butane near its boiling point*, Chem. Phys. Lett., 30 (1975), pp. 123–125.

[98] J. Ryckaert, G. Ciccotti, and H. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes*, J. Comput. Phys., 23 (1977), pp. 327–341.

[99] J. A. Sanders and F. Verhulst, *Averaging Methods in Nonlinear Dynamical Systems*, Appl. Math. Sci. 59, Springer-Verlag, New York, 1985.

[100] J. M. Sanz-Serna, *Modulated Fourier expansions and heterogeneous multiscale methods*, IMA J. Numer. Anal., 29 (2009), pp. 595–605.

[101] J. M. Sanz-Serna, *Mollified impulse methods for highly oscillatory differential equations*, SIAM J. Numer. Anal., 46 (2008), pp. 1040–1059.

[102] J. M. Sanz-Serna, *Stabilizing with a hammer*, Stoch. Dyn., 8 (2008), pp. 47–57.

[103] J. M. Sanz-Serna, G. Ariel, and Y.-H. Tsai, *Multiscale Methods for Stiff and Constrained Mechanical Systems*, preprint, 2009.

[104] R. Scheid, *The accurate numerical solution of highly oscillatory ordinary differential equations*, Math. Comp., 41 (1983), pp. 487–509.

[105] C. Schütte and F. A. Bornemann, *Homogenization approach to smoothed molecular dynamics*, Nonlinear Anal., 30 (1997), pp. 1805–1814.

[106] R. Sharp, Y.-H. Tsai, and B. Engquist, *Multiple time scale numerical methods for the inverted pendulum problem*, in Multiscale Methods in Science and Engineering, Lect. Notes Comput. Sci. Eng. 44, Springer-Verlag, Berlin, 2005, pp. 241–261.

[107] R. Skeel and J. A. Izaguirre, *An impulse integrator for Langevin dynamics*, Mol. Phys., 100 (2002), pp. 3885–3891.

[108] A. Skorokhod, *Asymptotic Methods in the Theory of Stochastic Differential Equations*, Transl. Math. Monogr. 78, AMS, Providence, RI, 1989.

[109] A. Stern and E. Grinspun, *Implicit-explicit variational integration of highly oscillatory problems*, Multiscale Model. Simul., 7 (2009), pp. 1779–1794.

[110] W. B. Streett, D. J. Tildesley, and G. Saville, *Multiple time-step methods in molecular dynamics*, Mol. Phys., 35 (1978), pp. 639–648.

[111] F. Takens, *Motion under the influence of a strong constraining force*, in Global Theory of Dynamical Systems, Z. Nitecki and C. Robinson, eds., Springer-Verlag, Berlin, Heidelberg, 1980, pp. 425–445.

[112] T. Tian and K. Burrage, *Implicit Taylor methods for stiff stochastic differential equations*, Appl. Numer. Math., 38 (2001), pp. 167–185.

[113] M. Tuckerman, B. J. Berne, and G. J. Martyna, *Reversible multiple time scale molecular dynamics*, J. Chem. Phys., 97 (1992), pp. 1990–2001.

[114] E. Vanden-Eijnden, *On HMM-like integrators and projective integration methods for systems with multiple time scales*, Commun. Math. Sci., 5 (2007), pp. 495–505.

[115] E. Vanden-Eijnden and G. Ciccotti, *Second-order integrators for Langevin equations with holonomic constraints*, Chem. Phys. Lett., 429 (2006), pp. 310–316.

[116] F. Verhulst, *Nonlinear Differential Equations and Dynamical Systems*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, 1996.

[117] L. Verlet, *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*, Phys. Rev., 159 (1967), pp. 98–103.

[118] D. Viswanath, *Global errors of numerical ODE solvers and Lyapunov's theory of stability*, IMA J. Numer. Anal., 21 (2001), pp. 387–406.

[119] T. Yanao, W. S. Koon, and J. E. Marsden, *Intramolecular energy transfer and the driving mechanisms for large-amplitude collective motions of clusters*, J. Chem. Phys., 130 (2009), 144111.

[120] G. Zhang and T. Schlick, *LIN: A new algorithm to simulate the dynamics of biomolecules by combining implicit-integration and normal mode techniques*, J. Comput. Chem., 14 (1993), pp. 1212–1233.