COMM. MATH. SCI.
Vol. 2, No. 3, pp. 515–534

TIME SPLITTING FOR THE LIOUVILLE EQUATION IN A RANDOM MEDIUM *

GUILLAUME BAL † AND LENYA RYZHIK ‡

Abstract. We consider the Liouville equations with highly heterogeneous Hamiltonians and their numerical solution by a time splitting algorithm. Such equations model the density of particles evolving according to the corresponding Hamiltonian dynamics as well as the propagation of high frequency waves with a wavelength much smaller than the correlation length of the random Hamiltonian.

Our main results are on the relation between the time step used in the time splitting algorithm and the correlation length of the Hamiltonian. In order to fully resolve the Liouville equation, the time step must be chosen much smaller than the correlation length. However, we show that the time step can be chosen of the same order as the correlation length of the Hamiltonian when one is only interested in suitable statistical properties of the solution to the Liouville equation. We also present a more involved time splitting algorithm that allows us to take a time step independent of the correlation length.

Key words. Random medium, time-splitting methods.

AMS Classification numbers. 65M12, 60F05.

1. Introduction

In many problems of wave propagation the typical wavelength of the propagating fields can be assumed to be smaller than the correlation length of the underlying propagating medium. Applications include light in a turbulent atmosphere, microwaves in wireless communication, acoustic waves in underwater communication, and seismic waves generated by earthquakes [5, 12, 17, 18]. Here we consider propagation of high frequency acoustic waves. Wave propagation in such a regime can be approximated by a Liouville equation with random potential for the acoustic energy density in the phase space (see [1] and references therein).

There exist many works on the numerical simulation of wave equations; see, e.g., [6, 7] for recent monographs. The numerical techniques are usually well adapted to the low-to-moderate frequency regime where the size of the calculation domain is not too large compared to the typical wavelength of the system. High frequency wave propagation in the semi-classical regime, which corresponds to the high frequency regime with slowly varying underlying media, has also been considered [4, 14, 15].

Here we consider the propagation of the energy density of waves over times and distances that are *large* compared to the correlation length of the Hamiltonian. This is modelled by the following Liouville equation

$$\frac{\partial W_\varepsilon}{\partial t} + \mathbf{k}\cdot\nabla_{\mathbf{x}} W_\varepsilon - \frac{1}{\sqrt{\varepsilon}}\nabla_{\mathbf{x}} V\left(\frac{t}{\varepsilon},\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}} W_\varepsilon = 0. \tag{1.1}$$

The function W_ε is the phase space energy density of acoustic waves, that is, the energy density of a wave at a position $\mathbf{x}$ with a wave vector $\mathbf{k}$. It is defined as the limit Wigner distribution of the solution of the underlying wave equation in the

*Received: February 22, 2004; accepted (in revised version): August 5, 2004. Communicated by Weinan E.

†Department of Applied Physics & Applied Mathematics, Columbia University, New York, NY 10027, USA (gb2030@columbia.edu).

‡Department of Mathematics, University of Chicago, Chicago, IL 60637, USA (ryzhik@math.uchicago.edu).

high frequency limit [9, 16]. The parameter $\varepsilon \ll 1$ measures the correlation length and time of the random medium relative to the overall propagation distance and time, respectively. It is small in the applications we have in mind. The strength of the potential is scaled in such a way that the above equation admits a physically interesting limit as $\varepsilon \to 0$: see Section 2.

We consider a time discretization of the above equation by a time splitting method [19]. Our analysis follows the method of study of the time splitting algorithm of a parabolic wave equation that we have used in [3]. Let us denote by

$$A = \mathbf{k} \cdot \nabla_{\mathbf{x}}, \quad B_\varepsilon = -\frac{1}{\sqrt{\varepsilon}} \nabla_{\mathbf{x}} V\left(\frac{t}{\varepsilon}, \frac{\mathbf{x}}{\varepsilon}\right) \cdot \nabla_{\mathbf{k}}$$

the spatial advection and the "scattering" operators, respectively. The time splitting algorithm consists of separating the advection of the particles from the interaction with the underlying medium. So during a time interval $\mathcal{T} = (T_n, T_{n+1})$, we first solve the advection equation

$$\left(\frac{\partial}{\partial t} + A\right) W_\varepsilon = 0$$

on $\mathcal{T}$ assuming that the medium is homogeneous. We then consider the solution at the end of the interval as the initial condition to solve the wave vector advection ("scattering") equation

$$\left(\frac{\partial}{\partial t} + B_\varepsilon\right) W_\varepsilon = 0$$

on the same interval $\mathcal{T}$, this time accounting for interactions with the random medium but without advection in the physical space. The interest of the method is that the first step can easily be solved for instance by the Fourier method since the medium is homogeneous, and the second step can also easily be solved since the problem is now local in space during the time integration on $\mathcal{T}$.

One may, of course, also consider (2.1) as the Liouville equation for the density of particles evolving according to the random Hamiltonian

$$H(t, \mathbf{x}, \mathbf{k}) = \frac{|\mathbf{k}|^2}{2} + \sqrt{\varepsilon} V\left(\frac{t}{\varepsilon}, \frac{\mathbf{x}}{\varepsilon}\right)$$

without any reference to the acoustic waves. The time-splitting algorithm described above corresponds to alternatingly solving the Hamilton equations: first, $\dfrac{d\mathbf{X}}{dt} = \mathbf{K}$ with $\mathbf{K}$ fixed, and then

$$\frac{d\mathbf{K}}{dt} = -\frac{1}{\sqrt{\varepsilon}} \nabla_{\mathbf{x}} V\left(\frac{t}{\varepsilon}, \frac{\mathbf{X}}{\varepsilon}\right)$$

with $\mathbf{X}$ fixed.

The question is then how one should optimally choose the time step $\Theta_\varepsilon = T_{n+1} - T_n$ provided that the correlation length ε of the underlying Hamiltonian is known. It is a classical result [19] that the accuracy (in the strong L^2 sense for instance) of the time splitting scheme is governed by

$$\Theta_\varepsilon \|[A, B_\varepsilon]\| = \Theta_\varepsilon \|AB_\varepsilon - B_\varepsilon A\| = O(\Theta_\varepsilon \varepsilon^{-3/2}),$$

where $\|\cdot\|$ is the L^2 norm. The constraint $\Theta_\varepsilon \ll \varepsilon^{3/2}$ is necessary to fully resolve the solution of the Liouville equation, with an accuracy (in the strong L^2 sense) of order $\Theta_\varepsilon \varepsilon^{-3/2}$. Instead of the classical time splitting scheme mentioned above, we could use the more accurate Strang time splitting for the same computational cost. The latter scheme corresponds to solving $\left(\frac{\partial}{\partial t} + A\right) W_\varepsilon = 0$ for a time $\Theta_\varepsilon/2$ followed by $\left(\frac{\partial}{\partial t} + B_\varepsilon\right) W_\varepsilon = 0$ for a time Θ_ε, and finally by $\left(\frac{\partial}{\partial t} + A\right) W_\varepsilon = 0$ again for a time $\Theta_\varepsilon/2$. The leading term in the error made is then of order

$$\Theta_\varepsilon^2 \|A^2 B_\varepsilon\| = O(\Theta_\varepsilon^2 \varepsilon^{-5/2}).$$

We thus find that a better constraint $\Theta_\varepsilon \ll \varepsilon^{5/4}$ guarantees convergence in the strong L^2 sense when the Strang time splitting scheme is used.

Our main result is that $\Theta_\varepsilon \ll \varepsilon$ is actually sufficient provided one is interested in *statistical properties* of the solution and not in its complete detailed structure. More precisely we show that when $\Theta_\varepsilon \ll \varepsilon$, moments of the solution of the Liouville equation have the same deterministic limit as $\varepsilon \to 0$ as the continuous time dependent solution of the random Liouville equation. We actually show a stronger result, namely that the whole *law* of the moments of the solution of the time splitting scheme agrees in the limit $\varepsilon \to 0$ to the limiting law of the continuous solution of the Liouville equation. The moments are of the form $\langle W_\varepsilon, \lambda \rangle$, where $\lambda(\mathbf{x}, \mathbf{k})$ is a smooth test function.

We also introduce a modified time splitting algorithm for the Liouville equation, which allows us to obtain the same result provided $\Theta \ll 1$ independent of ε. This time splitting scheme is still based on separating convection from scattering. However in the new scheme the random potential $\nabla_{\mathbf{x}} V$ is evaluated at a point that depends on time t on the interval $\mathcal{T}$. This is a hybrid scheme, which no longer enjoys the property of the original scheme that the scattering term is local. The resulting equations are thus more complicated to solve than for the original scheme (though they are less complicated than the initial Liouville equation) but the time step can be chosen much larger and independent of ε.

Our results are related to other results on the commutativity of mesh size convergence and "small parameter" convergence, to a common limiting equation in homogenization problems, such as the homogenization of an elliptic problem in a periodic medium in [11], or the diffusion limit of the linear transport equation with a small mean free path in [10].

The proposed scheme is valid in two different asymptotical regimes, namely the regime of wave propagation in homogeneous media, where the phase information of the wave is important, and the regime of propagation in highly heterogeneous media, where only the energy density is meaningful because of the multiple scattering. At a much lower cost than classical time splitting schemes, we can thus expect to recover the phase information where it matters, as well as the correct energy density of waves in the scattering regions.

The rest of the paper is organized as follows. We present the proof that the time splitting scheme has the correct statistical limit as $\varepsilon \to 0$ in section 2. The proofs are similar to those of the limit theorems in [2] and [3]. Nevertheless we present the important details for the convenience of the reader. The Liouville problem is formulated in Section 2.1, where the assumptions on the random potential are also presented. The time splitting algorithm is introduced and our main result presented in Section 2.2. An outline of the proof of our result is given in Section 2.3. The martingale structure at the core of our demonstration is presented in section 2.4 and the convergence result for the whole law of the time splitting algorithm in Section 2.5.

In section 3 we introduce the time splitting algorithm that allows us to choose $\Theta_\varepsilon \ll 1$ *independent* of ε. We present the modified scheme in Section 3.1 and its convergence properties in Section 3.2.

2. The high-frequency limit of the time-splitting scheme

2.1. The Liouville equation. We start with the following Liouville equation

$$\frac{\partial W_\varepsilon}{\partial t} + \mathbf{k}\cdot\nabla_{\mathbf{x}} W_\varepsilon - \frac{1}{\sqrt{\varepsilon}}\nabla_{\mathbf{x}} V\left(\frac{t}{\varepsilon},\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}} W_\varepsilon = 0, \tag{2.1}$$

with the initial data $W_\varepsilon(0,\mathbf{x},\mathbf{k}) = W_0(\mathbf{x},\mathbf{k}) \in L^2(\mathbb{R}^d\times\mathbb{R}^d)$ independent of the parameter ε.

We assume that the random potential $V(t,\mathbf{x})$ is a Markov process in time taking values in $C^1(\mathbb{R}^d)$. The Markovian hypothesis is not necessary to obtain the results presented below – a sufficiently strong mixing assumption in time is sufficient [8]. It is however crucial to simplify the mathematical analysis because it allows us to treat the process $t\mapsto(V(t/\varepsilon,\mathbf{x}/\varepsilon),W_\varepsilon(t,\mathbf{x},\mathbf{k}))$ as jointly Markov and to apply the martingale method. Here are some additional assumptions on the potential. We assume that there exists a deterministic constant $C_0>0$ so that

$$\|V\|_{C^1(\mathbb{R}^d)} \le C_0 \tag{2.2}$$

with probability one. We define the set $\mathcal{V} = \{V\in C^1(\mathbb{R}^d)\colon \|V\|_{C^1(\mathbb{R}^d)}\le C_0\}$ where the process V takes its values. Furthermore, $V(t,\mathbf{x})$ is assumed to be stationary in $\mathbf{x}$ and t and to have mean zero: $\mathbb{E}\{V(t,\mathbf{x})\}=0$. The correlation function

$$R(t,\mathbf{x}) = \mathbb{E}\{V(s,\mathbf{y})V(t+s,\mathbf{x}+\mathbf{y})\} \tag{2.3}$$

is assumed to be smooth and rapidly decaying in space. We assume that the generator Q of the Markov process $V(t)$ is a bounded operator on $L^\infty(\mathcal{V})$ with a unique invariant measure $\pi(V)$

$$Q^*\pi = 0,$$

and that there exists $\alpha>0$ such that if $\langle g,\pi\rangle = 0$ then

$$\|e^{rQ}g\|_{L^\infty_{\mathcal{V}}} \le C\|g\|_{L^\infty_{\mathcal{V}}} e^{-\alpha r}. \tag{2.4}$$

A simple example of a generator with a gap in the spectrum and invariant measure π is a jump process on $\mathcal{V}$ where

$$Qg(V) = \int_{\mathcal{V}} g(V_1)d\pi(V_1) - g(V), \quad \int_{\mathcal{V}} d\pi(V) = 1.$$

Given (2.4), the Fredholm alternative holds for the Poisson equation

$$Qf = g,$$

provided that g satisfies $\langle\pi,g\rangle = 0$. It has a unique solution f with $\langle\pi,f\rangle=0$ and $\|f\|_{L^\infty_{\mathcal{V}}}\le C\|g\|_{L^\infty_{\mathcal{V}}}$. The solution f is given explicitly by

$$f(V) = -\int_0^\infty e^{rQ}g(V)dr.$$

The integral above converges absolutely because of the spectral gap assumption (2.4). More generally, the mean-zero bounded solution of

$$\frac{\partial f}{\partial \tau} + Qf = g(\tau, V) \tag{2.5}$$

with the right side $g \in L^\infty([0,T] \times \mathcal{V})$ is given by

$$f(\tau, V) = -\int_0^\infty e^{rQ} g(\tau + r, V) dr \tag{2.6}$$

provided that $\langle \pi, g \rangle(\tau) = 0$ for all $\tau > 0$.

One may show under the above assumptions that the expectation of the solution of (2.1) converges to the solution of a Fokker-Planck equation.

THEOREM 2.1. *Let $W_0(\mathbf{x},\mathbf{k}) \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$, and let $\phi \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$ be a test function. Then the process $\int \phi(\mathbf{x},\mathbf{k}) W_\varepsilon(t,\mathbf{x},\mathbf{k}) d\mathbf{x} d\mathbf{k}$ converges in probability to $\int \phi(\mathbf{x},\mathbf{k}) W(t,\mathbf{x},\mathbf{k}) d\mathbf{x} d\mathbf{k}$. Here $W(t,\mathbf{x},\mathbf{k})$ is a solution of the Fokker-Planck equation*

$$\frac{\partial W}{\partial t} + \mathbf{k} \cdot \nabla_{\mathbf{x}} W = \frac{\partial}{\partial k_m} \left(D_{mn}(\mathbf{k}) \frac{\partial W}{\partial k_n} \right) \tag{2.7}$$

with the diffusion matrix

$$D_{mn}(\mathbf{k}) = -\int_0^\infty \frac{\partial^2 R(s, s\mathbf{k})}{\partial z_n \partial z_m} ds. \tag{2.8}$$

Theorem 2.1 was proved in the much harder case of a time-independent random potential V in [13] for $\mathbb{E}\{W_\varepsilon\}$, while convergence in probability for such potentials was established in [1]. We will restrict our analysis in this paper to the case of time-dependent Markovian potentials in order to keep the presentation simple. Nevertheless our results may be generalized to the time independent case under sufficiently strong mixing assumptions on the potential using the techniques of [1]. This, however, is highly technical and lies beyond the scope of the present paper.

2.2. The time-splitting algorithm. In the Liouville equation (2.1), the advection part $\partial_t + A$ and the scattering part $\partial_t + B_\varepsilon$ are easier to solve than the full equation $\partial_t + A + B_\varepsilon$. This justifies the time splitting method to solve (2.1) numerically. The time splitting method consists of two steps. First, given the approximation $W(n\Theta,\mathbf{x},\mathbf{k})$ one solves the pure streaming part

$$\frac{\partial U^{n+1}}{\partial t} + \mathbf{k} \cdot \nabla_{\mathbf{x}} U^{n+1} = 0 \tag{2.9}$$

on a time interval $n\Theta \le t \le (n+1)\Theta$ with the initial data $U^{n+1}(n\Theta,\mathbf{x},\mathbf{k}) = W(n\Theta,\mathbf{x},\mathbf{k})$. The solution of (2.9) may be written explicitly as

$$U^{n+1}(t,\mathbf{x},\mathbf{k}) = W(n\Theta, \mathbf{x} - (t - n\Theta)\mathbf{k}, \mathbf{k}).$$

During the second step one solves

$$\frac{\partial Z^{n+1}}{\partial t} - \frac{1}{\sqrt{\varepsilon}} \nabla_{\mathbf{x}} V\left(\frac{t}{\varepsilon}, \frac{\mathbf{x}}{\varepsilon}\right) \cdot \nabla_{\mathbf{k}} Z^{n+1} = 0, \quad n\Theta \le t < (n+1)\Theta, \tag{2.10}$$

with initial data $Z^{n+1}(n\Theta,\mathbf{x},\mathbf{k}) = U^{n+1}((n+1)\Theta,\mathbf{x},\mathbf{k})$. Equation (2.10) may also be solved explicitly:

$$Z^{n+1}(t,\mathbf{x},\mathbf{k}) = U^{n+1}\left((n+1)\Theta,\mathbf{x},\mathbf{k}+\frac{1}{\sqrt{\varepsilon}}\int_{n\Theta}^{t}\nabla_{\mathbf{x}}V\left(\frac{s}{\varepsilon},\frac{\mathbf{x}}{\varepsilon}\right)ds\right).$$

Then we set $W((n+1)\Theta,\mathbf{x},\mathbf{k}) = Z^{(n+1)}((n+1)\Theta,\mathbf{x},\mathbf{k})$ and iterate the procedure.

It remains to understand how the time step Θ should be chosen. We have seen in the introduction that time steps smaller than ε^{α} with $\alpha = 3/2$ for the above time splitting method and $\alpha = 5/4$ for the Strang time splitting method are necessary to obtain an error converging to 0 strongly in the L^2 sense. In this paper, we are interested in the largest time step Θ so that the solution of the time-splitting algorithm (2.9)-(2.10) has the *statistical behavior* close to that of the exact solution of the Liouville equation. In practice, this is often all we are interested in. We show that in the latter case, the time step Θ can be chosen much larger than what we just mentioned, whereby substantially reducing the computational cost of the Liouville solution.

The time splitting algorithm can also be given the following interpretation. The Liouville equation (2.1) may be solved by the classical method of characteristics:

$$W_{\varepsilon}(t,\mathbf{x},\mathbf{k}) = W_0(\mathbf{X}^{\varepsilon}(s=0;t,\mathbf{x},\mathbf{k}),\mathbf{K}^{\varepsilon}(s=0;t,\mathbf{x},\mathbf{k}))$$

with

$$\frac{d\mathbf{X}^{\varepsilon}}{ds} = \mathbf{K}^{\varepsilon},\quad \frac{d\mathbf{K}^{\varepsilon}}{ds} = -\frac{1}{\sqrt{\varepsilon}}\nabla_{\mathbf{x}}V\left(\frac{t}{\varepsilon},\frac{\mathbf{X}^{\varepsilon}}{\varepsilon}\right),\quad \mathbf{X}^{\varepsilon}(s=t;t,\mathbf{x},\mathbf{k}) = \mathbf{x},\quad \mathbf{K}^{\varepsilon}(s=t;t,\mathbf{x},\mathbf{k}) = \mathbf{k}. \tag{2.11}$$

The time-splitting algorithm (2.9)-(2.10) corresponds to a time-splitting approximation of (2.11) obtained by evolving $\mathbf{X}^{\varepsilon}$ and $\mathbf{K}^{\varepsilon}$ in an alternating manner. The dynamics are then trivially solved with source terms that become constant over the time-splitting interval.

Before stating the main result, it is convenient to reformulate the time-splitting algorithm in a somewhat more general framework as follows. We replace the exact equation (2.1) by

$$\frac{\partial W_{\varepsilon}}{\partial t} + \phi\left(\frac{t}{\varepsilon}\right)\mathbf{k}\cdot\nabla_{\mathbf{x}}W_{\varepsilon} - \frac{1}{\sqrt{\varepsilon}}\psi\left(\frac{t}{\varepsilon}\right)\nabla_{\mathbf{x}}V\left(\frac{t}{\varepsilon},\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}}W_{\varepsilon} = 0. \tag{2.12}$$

The functions $\phi(\tau)$ and $\psi(\tau)$ are periodic in $\tau = t/\varepsilon$ with period $\Theta > 0$, which corresponds to a period $\Theta\varepsilon$ on the large time scale. A generalized time-splitting algorithm corresponds to periodically shutting down the two operators in (2.1) so that

$$\psi(\tau) = \begin{cases} 0,\ \tau\in[0,\tau_0) \\ \dfrac{\Theta}{\Theta-\tau_0},\ \tau\in[\tau_0,\Theta), \end{cases} \qquad \text{and} \qquad \phi(\tau) = \begin{cases} \dfrac{1}{\tau_0},\ \tau\in[0,\tau_0) \\ 0,\ \tau\in[\tau_0,\Theta). \end{cases}$$

The standard time-splitting scheme corresponds to $\psi(\tau) = 1$ and $\phi(\tau) = \sum_{n=-\infty}^{\infty}\delta(\tau - n\Theta)$, that is, the limit $\tau_0 \to 0$ of the above. Note that information is lost if $\tau_0 > 0$ since the random potential is time-dependent. This would not be the case for time-independent potentials. We allow for more general distributions ψ and ϕ in order to investigate other possibilities but impose the constraints

$$\frac{1}{\Theta}\int_0^{\Theta}\phi(\tau)d\tau = 1, \qquad \frac{1}{\Theta}\int_0^{\Theta}\psi(\tau)d\tau = 1. \tag{2.13}$$

This is a natural restriction ensuring that both operators in the time-splitting procedure have equal weight and that time is not re-scaled.

Let us define the diffusion operator

$$\mathcal{L}_\Theta f = -\frac{1}{\Theta}\int_0^\Theta \int_0^\infty \psi(\tau)\psi(\tau+s)\frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\mathbf{k}[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds\ d\tau, \tag{2.14}$$

where $\Phi(s)$ is an anti-derivative of ϕ: $\dfrac{d\Phi}{d\tau}=\phi$. Since only increments of Φ appear in our results the choice of a particular anti-derivative is irrelevant. The main result of this section is the following theorem.

THEOREM 2.2. *Let the initial data $W_0(\mathbf{x},\mathbf{k})$ for (2.12) be bounded in $L^2(\mathbb{R}^{2d})$, the function ψ and ϕ satisfy the normalization (2.13) and be uniformly bounded. Then the solution of (2.12) converges in probability and weakly in $L^2(\mathbb{R}^d)$ to the solution $\overline{W}$ of the modified Fokker-Planck equation*

$$\frac{\partial \overline{W}}{\partial t}+\mathbf{k}\cdot\nabla_{\mathbf{x}}\overline{W}=\mathcal{L}_\Theta \overline{W} \tag{2.15}$$

with initial data $W_0(\mathbf{x},\mathbf{k})$. More precisely, for any test function $\lambda\in L^2(\mathbb{R}^d)$ the random process

$$\langle W_\varepsilon,\lambda(t)\rangle=\int_{\mathbb{R}^{2d}} W_\varepsilon(t,\mathbf{x},\mathbf{k})\lambda(\mathbf{x},\mathbf{k})d\mathbf{x}d\mathbf{k}$$

converges in probability to $\langle\overline{W},\lambda\rangle$ as $\varepsilon\to 0$ uniformly on finite time intervals $t\in[0,T]$.

An important special case arises when $\psi=1$ and $\phi(\tau)=\Theta\sum_{j=-\infty}^{\infty}\delta(\tau-j\Theta)$. Theorem 2.2 does not apply to this case as stated, as the function ϕ is unbounded. However, as only the anti-derivative Φ of the function ϕ enters in most estimates, the only modifications in the proof required to treat this case are in estimates (2.36) and (2.37), and these are straightforward. This corresponds to the time-splitting scheme (2.9)-(2.10) when scattering by the random potential V is accounted for at all times while advection in the spatial variable is accounted for at times $t=j\varepsilon\Theta$ by the correction $W(\varepsilon j\Theta^+,\mathbf{x},\mathbf{k})=W(\varepsilon j\Theta^-,\mathbf{x}-\varepsilon\Theta\mathbf{k},\mathbf{k})$. Then we have $\Phi(s)=\Theta[s/\Theta]:=[s]_\Theta$ and obtain the following expression for $\mathcal{L}_\Theta$:

$$\mathcal{L}_\Theta f = -\frac{1}{\Theta}\int_0^\Theta \int_0^\infty \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\mathbf{k}\{[\tau+s]_\Theta-[\tau]_\Theta\})}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds\ d\tau.$$

However, we have $[\tau]_\Theta=0$ when $0\le\tau<\Theta$ and thus obtain

$$\mathcal{L}_\Theta f = -\frac{1}{\Theta}\int_0^\Theta \int_0^\infty \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\mathbf{k}\{[\tau+s]_\Theta\})}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds\ d\tau. \tag{2.16}$$

It is instructive to consider the limits $\Theta\to 0$ and $\Theta\to\infty$. The former limit corresponds to a time step much smaller than the correlation length. Then we obtain after the change of variables $\eta=\tau/\Theta$:

$$\begin{aligned}
\mathcal{L}_\Theta f &= -\frac{1}{\Theta}\int_0^\Theta \int_0^\infty \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\mathbf{k}\{[\tau+s]_\Theta\})}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds\ d\tau \\
&= -\int_0^1 \int_0^\infty \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\mathbf{k}\{[s+\Theta\eta]_\Theta\})}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds\ d\eta \\
&\to -\int_0^\infty \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,s\mathbf{k})}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds = \frac{\partial}{\partial k_m}\left(D_{mn}(\mathbf{k})\frac{\partial f}{\partial k_n}\right)
\end{aligned} \tag{2.17}$$

since $[s+\Theta\eta]_\Theta = \Theta[\eta+(s/\Theta)] \to s$ as $\Theta \to 0$ point-wise in s and η. Here $D_{mn}(\mathbf{k})$ is the exact limit diffusion matrix (2.8) that arises without the time-splitting approximation.

In order to consider the opposite limit $\Theta \to \infty$, it is convenient to assume that the correlation function $R(s,\mathbf{x})$ has a compact support in s: $R(s,\mathbf{y}) = 0$ for $s \geq T_0$. Then we have for $0 \leq s \leq T_0$ and Θ sufficiently large:

$$[s+\Theta\eta]_\Theta = \Theta[\eta+(s/\Theta)] = \begin{cases} 0, & 0 \leq \eta < 1 - \dfrac{s}{\Theta}, \\ \Theta, & 1 - \dfrac{s}{\Theta} \leq \eta \leq 1. \end{cases}$$

The operator $\mathcal{L}_\Theta$ becomes

$$\begin{aligned}
\mathcal{L}_\Theta f &= -\int_0^1 \int_0^\infty \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\mathbf{k}\{[s+\Theta\eta]_\Theta\})}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds\, d\eta \\
&= -\int_0^\infty \int_0^{1-s/\Theta} \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,0)}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] d\eta ds - \int_0^\infty \int_{1-s/\Theta}^1 \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\Theta\mathbf{k})}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] d\eta ds \\
&= -\int_0^\infty \frac{\partial}{\partial k_m}\left[\left\{\left(1-\frac{s}{\Theta}\right)\frac{\partial^2 R(s,0)}{\partial z_m \partial z_n} + \frac{s}{\Theta}\frac{\partial^2 R(s,\Theta\mathbf{k})}{\partial z_m \partial z_n}\right\}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds \\
&\to -\int_0^\infty \frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,0)}{\partial z_m \partial z_n}\frac{\partial f(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right] ds = \frac{\partial}{\partial k_m}\left(\bar{D}_{mn}(\mathbf{k})\frac{\partial f}{\partial k_n}\right). \qquad (2.18)
\end{aligned}$$

The above diffusion matrix $\bar{D}_{mn}$ corresponds to the case of fluctuations in V that are white noise in time. It means that when the time step is chosen much larger than the correlation length, the limit diffusion operator of the time-splitting scheme is correct only when the medium fluctuations are white in time. Therefore, choosing a time step on the order of or much larger than the wavelength ε usually leads to incorrect statistics for the solution to the Liouville equation. Theorem 2.2 and (2.14) quantify the error made by choosing too big a time step. Let us also remark that it is straightforward to generalize the above calculation to correlation functions that are not compactly supported in time, although we shall not consider this case here.

2.3. Outline of the proof. The proof of Theorem 2.2 follows the idea of the proof of the main result in [2]. Therefore we briefly outline the main steps and concentrate on the necessary modifications in the proof. First, we need to show that the family of measures P_ε generated by the process $W_\varepsilon(t)$ on $C([0,T];L^2(\mathbb{R}^{2d}))$ is tight:

LEMMA 2.3. *The family of measures P_ε is weakly compact.*

The proof of this lemma is very similar to that in [2] and is omitted.

It is straightforward to verify that the L^2-norm of W_ε is preserved by the evolution and hence W_ε takes values in a ball $X = \{W \in L^2 : \|W\|_{L^2} \leq C\}$.

LEMMA 2.4. *The L^2-norm of the solution of (2.12) is preserved:*

$$\|W_\varepsilon(t)\|_{L^2(\mathbb{R}^{2d})} = \|W_\varepsilon(0)\|_{L^2(\mathbb{R}^{2d})}. \qquad (2.19)$$

Let $\lambda(t,\mathbf{x},\mathbf{k})$ be a fixed deterministic function. In order to identify the limit of W_ε, we construct the functional $G_\lambda : C([0,T];X) \to C[0,T]$ defined by

$$G_\lambda[W](t) = \langle W,\lambda\rangle(t) - \int_0^t \langle W, \frac{\partial\lambda}{\partial t} + \mathbf{k}\cdot\nabla_{\mathbf{x}}\lambda + \mathcal{L}_\Theta\lambda\rangle(s)ds \qquad (2.20)$$

and show that it is an approximate martingale. More precisely, we show that the following lemma holds.

LEMMA 2.5. *There exists a constant $C>0$ so that*

$$\left|\mathbb{E}^{P_\varepsilon}\{G_\lambda[W](t)|\mathcal{F}_s\}-G_\lambda[W](s)\right|\le C_{\lambda,T}\sqrt{\varepsilon} \tag{2.21}$$

uniformly for all $W\in C([0,T];X)$ and $0\le s<t\le T$.

The proof of Lemma 2.5 is based on the construction of an exact martingale $G^\varepsilon_\lambda[W]$ that is uniformly close to $G_\lambda[W]$ within $O(\sqrt{\varepsilon})$. Lemma 2.3 implies that there exists a subsequence $\varepsilon_j\to 0$ so that P_{ε_j} converges weakly to a measure P supported on $C([0,T];X)$. The weak convergence of P_ε and the strong convergence (2.21) altogether imply that $G_\lambda[W](t)$ is a P-martingale so that

$$\mathbb{E}^P\{G_\lambda[W](t)|\mathcal{F}_s\}-G_\lambda[W](s)=0. \tag{2.22}$$

Taking $s=0$ in the above equation we obtain the Fokker-Planck equation (2.15) for $\overline{W}=E^P\{W(t)\}$ in its weak formulation. The construction of the martingale G^ε_λ and the proof of Lemma 2.5 are presented in detail in Section 2.4.

The second step is to show that for every test function $\lambda(t,\mathbf{x},\mathbf{k})$ the second functional

$$G_{2,\lambda}[W](t)=\langle W,\lambda\rangle^2(t)-2\int_0^t\langle W,\lambda\rangle(s)\langle W,\frac{\partial\lambda}{\partial s}+\mathbf{k}\cdot\nabla_{\mathbf{x}}\lambda+\mathcal{L}_\Theta\lambda\rangle(s)ds$$

is also an approximate P_ε-martingale. We then obtain that $\mathbb{E}^{P_\varepsilon}\left\{\langle W,\lambda\rangle^2\right\}\to\langle\overline{W},\lambda\rangle^2$, which implies (weak) convergence in probability of $W_\varepsilon(t,\mathbf{x},\mathbf{k})$ to $\overline{W}(t,\mathbf{x},\mathbf{k})$. It follows that the limit measure P is unique and deterministic, and that the whole sequence P_ε converges to P.

2.4. The approximate martingale. To obtain the approximate martingale property (2.21) and prove Lemma 2.5, one has to consider the conditional expectation of functionals $F(W,V)$ with respect to the probability measure $\tilde{P}_\varepsilon$ on the space $C([0,T];\mathcal{V}\times X)$ generated by $V(t/\varepsilon)$ and the Cauchy problem (2.12). The only functions we need to consider are actually of the form $F(W,V)=\langle W,\lambda(V)\rangle$ with $\lambda\in L^\infty(\mathcal{V};C^1([0,T];\mathcal{S}(\mathbb{R}^{2d})))$. Given a function $F(W,V)$ let us define the conditional expectation

$$\mathbb{E}^{\tilde{P}_\varepsilon}_{W,V,t}\{F(W,V)\}(\tau)=\mathbb{E}^{\tilde{P}_\varepsilon}\{F(W(\tau),V(\tau))|\ W(t)=W,V(t)=V\},\quad \tau\ge t.$$

The weak form of the infinitesimal generator of the Markov process generated by $\tilde{P}_\varepsilon$ is given by

$$\begin{aligned}&\frac{d}{dh}\mathbb{E}^{\tilde{P}_\varepsilon}_{W,V,t}\{\langle W,\lambda(V)\rangle\}(t+h)\Big|_{h=0}\\&=\frac{1}{\varepsilon}\langle W,Q\lambda\rangle+\left\langle W,\left(\frac{\partial}{\partial t}+\phi\left(\frac{t}{\varepsilon}\right)\mathbf{k}\cdot\nabla_{\mathbf{x}}-\frac{1}{\sqrt{\varepsilon}}\psi\left(\frac{t}{\varepsilon}\right)\nabla_{\mathbf{x}}V\left(\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}}\right)\lambda\right\rangle,\end{aligned} \tag{2.23}$$

hence

$$\begin{aligned}&G^\varepsilon_\lambda=\langle W,\lambda(V)\rangle(t)-\\&\int_0^t\left\langle W(s),\left(\frac{1}{\varepsilon}Q+\frac{\partial}{\partial s}+\phi\left(\frac{s}{\varepsilon}\right)\mathbf{k}\cdot\nabla_{\mathbf{x}}-\frac{1}{\sqrt{\varepsilon}}\psi\left(\frac{s}{\varepsilon}\right)\nabla_{\mathbf{x}}V\left(\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}}\right)\lambda(s)\right\rangle ds,\end{aligned} \tag{2.24}$$

is a $\tilde{P}_\varepsilon$-martingale. The generator (2.23) comes from equation (2.12).

Given a test function $\lambda(t,\mathbf{x},\mathbf{k})\in C^1([0,L];\mathcal{S})$ we will construct the function

$$\lambda_\varepsilon(t,\mathbf{x},\mathbf{k},V)=\lambda(t,\mathbf{x},\mathbf{k})+\sqrt{\varepsilon}\lambda_1^\varepsilon(t,\mathbf{x},\mathbf{k},V)+\varepsilon\lambda_2^\varepsilon(t,\mathbf{x},\mathbf{k},V), \tag{2.25}$$

with the correctors $\lambda_{1,2}^\varepsilon(t)$ bounded in $L^\infty(\mathcal{V};L^2(\mathbb{R}^{2d}))$ uniformly in $t\in[0,T]$. The functions $\lambda_{1,2}^\varepsilon$ will be chosen so that

$$\|G^\varepsilon_{\lambda_\varepsilon}(t)-G_\lambda(t)\|_{L^\infty(\mathcal{V})}\le C_\lambda\sqrt{\varepsilon}, \tag{2.26}$$

for all $t\in[0,T]$. Here $G^\varepsilon_{\lambda_\varepsilon}$ is defined by (2.24) with λ replaced by λ_ε, and G_λ is defined by (2.20). The approximate martingale property (2.21) follows from this.

As is usual in homogenization theory we need to account for the behavior both at the fast and the slow scales and define $\lambda_j^\varepsilon(t,\mathbf{x},\mathbf{k},V)=\lambda_j(t,\frac{t}{\varepsilon},\mathbf{x},\frac{\mathbf{x}}{\varepsilon},\mathbf{k},V)$ for $j=1,2$. The function $\lambda_1(t,\tau,\mathbf{x},\mathbf{z},\mathbf{k},V)$ is the mean-zero solution of

$$\frac{\partial\lambda_1}{\partial\tau}+\phi(\tau)\mathbf{k}\cdot\nabla_{\mathbf{z}}\lambda_1+Q\lambda_1=\psi(\tau)\nabla V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda, \tag{2.27}$$

with $V(\mathbf{z})$ fixed and independent of τ. Solution of this equation is given by

$$\lambda_1(t,\tau,\mathbf{x},\mathbf{z},\mathbf{k},V)=-\int_0^\infty e^{sQ}\frac{\partial V}{\partial z_n}(\mathbf{z}+\mathbf{k}[\Phi(\tau+s)-\Phi(\tau)])\psi(\tau+s)ds\frac{\partial\lambda(t,\mathbf{x},\mathbf{k})}{\partial k_n}. \tag{2.28}$$

The equation for λ_2 is

$$\frac{\partial\lambda_2}{\partial\tau}+\phi(\tau)\mathbf{k}\cdot\nabla_{\mathbf{z}}\lambda_2+Q\lambda_2=[\mathcal{L}_\Theta\lambda+\psi(\tau)\nabla_{\mathbf{z}}V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda_1]+[1-\phi(\tau)]\mathbf{k}\cdot\nabla_{\mathbf{x}}\lambda. \tag{2.29}$$

The first term on the right may be further decomposed as

$$\Big(\mathcal{L}_\Theta\lambda-\mathcal{L}^\tau_\Theta\lambda\Big)+\Big(\mathcal{L}^\tau_\Theta\lambda+\psi(\tau)\nabla_{\mathbf{z}}V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda_1).$$

We decompose λ_2 as

$$\lambda_2=\lambda_{21}+\lambda_{22}+\lambda_{23}, \tag{2.30}$$

corresponding to the three arising source terms, respectively. The operator $\mathcal{L}^\tau_\Theta$ is defined by

$$\mathcal{L}^\tau_\Theta\lambda=-\mathbb{E}\{\psi(\tau)\nabla_{\mathbf{z}}V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda_1\},$$

and

$$\mathcal{L}_\Theta\lambda=\frac{1}{\Theta}\int_0^\Theta\mathcal{L}^\tau_\Theta\lambda d\tau, \tag{2.31}$$

where Θ is the period of ϕ and ψ. The explicit form of $\mathcal{L}^\tau_\Theta$ is:

$$\begin{aligned}
&\mathcal{L}^\tau_\Theta\lambda(t,\tau,\mathbf{x},\mathbf{z},\mathbf{k})=-\mathbb{E}\{\psi(\tau)\nabla_{\mathbf{z}}V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda_1\}=-\psi(\tau)\mathbb{E}\left\{\frac{\partial}{\partial k_m}\left(\frac{\partial V(\mathbf{z})}{\partial z_m}\lambda_1\right)\right\}\\
&=\psi(\tau)\frac{\partial}{\partial k_m}\left[\mathbb{E}\left\{\int_0^\infty e^{sQ}\frac{\partial V(\mathbf{z})}{\partial z_m}\frac{\partial V}{\partial z_n}(\mathbf{z}+\mathbf{k}[\Phi(\tau+s)-\Phi(\tau)])\psi(\tau+s)ds\frac{\partial\lambda(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right\}\right]\\
&=-\psi(\tau)\frac{\partial}{\partial k_m}\left[\int_0^\infty\frac{\partial^2R(s,\mathbf{k}[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_m\partial z_n}\psi(\tau+s)ds\frac{\partial\lambda(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right].
\end{aligned}$$

Therefore we have

$$\mathcal{L}_\Theta\lambda=-\frac{1}{\Theta}\int_0^\Theta\int_0^\infty\psi(\tau)\psi(\tau+s)\frac{\partial}{\partial k_m}\left[\frac{\partial^2 R(s,\mathbf{k}[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_m\partial z_n}\frac{\partial\lambda(t,\mathbf{x},\mathbf{k})}{\partial k_n}\right]ds\ d\tau.$$

We observe that the operators $\mathcal{L}_\Theta$ and $\mathcal{L}_\Theta^\tau$ are independent of V and $\mathbf{z}$ and therefore the function $\lambda_{21}=\lambda_{21}(t,\tau,\mathbf{x},\mathbf{k})$ is also independent of these variables. It is given explicitly by

$$\lambda_{21}(t,\tau,\mathbf{x},\mathbf{k})=\int_0^\tau[\mathcal{L}_\Theta(t,\mathbf{x},\mathbf{k})-\mathcal{L}_\Theta^\zeta(t,\mathbf{x},\mathbf{k})]\lambda(t,\mathbf{x},\mathbf{k})d\zeta \tag{2.32}$$

and is periodic in the fast variable τ. Similarly the function λ_{23} is also independent of V and $\mathbf{z}$ and is given by

$$\lambda_{23}(t,\tau,\mathbf{x},\mathbf{k})=[\tau-\Phi(\tau)]\mathbf{k}\cdot\nabla_{\mathbf{x}}\lambda(t,\mathbf{x},\mathbf{k}). \tag{2.33}$$

The function λ_{22} satisfies

$$\frac{\partial\lambda_{22}}{\partial\tau}+\phi(\tau)\mathbf{k}\cdot\nabla_{\mathbf{z}}\lambda_{22}+Q\lambda_{22}=\mathcal{L}_\Theta^\tau\lambda-g,\quad g=\psi(\tau)\nabla_{\mathbf{z}}V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda_1,$$

and is thus given explicitly by

$$\begin{aligned}&\lambda_{22}(t,\tau,\mathbf{x},\mathbf{z},\mathbf{k},V)\\&=-\int_0^\infty e^{sQ}[\mathcal{L}_\Theta^{\tau+s}\lambda(t,\mathbf{x},\mathbf{k})-g(t,\tau+s,\mathbf{x},\mathbf{z}+(\Phi(\tau+s)-\Phi(\tau))\mathbf{k},\mathbf{k},V)]ds.\end{aligned} \tag{2.34}$$

Using (2.27) and (2.29) we have

$$\begin{aligned}&\frac{d}{dh}\mathbb{E}^{\tilde{P}_\varepsilon}_{W,V,t}\{\langle W,\lambda_\varepsilon\rangle\}(t+h)\Big|_{h=0}\\&=\left\langle W,\left(\frac{\partial}{\partial t}+\phi\left(\frac{t}{\varepsilon}\right)\mathbf{k}\cdot\nabla_{\mathbf{x}}-\frac{1}{\sqrt{\varepsilon}}\psi\left(\frac{t}{\varepsilon}\right)\nabla V\left(\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}}+\frac{1}{\varepsilon}Q\right)\left(\lambda+\sqrt{\varepsilon}\lambda_1^\varepsilon+\varepsilon\lambda_2^\varepsilon\right)\right\rangle\\&=\left\langle W,\left(\frac{\partial}{\partial t}+\mathbf{k}\cdot\nabla_{\mathbf{x}}\right)\lambda+\mathcal{L}_\Theta\lambda\right\rangle\\&+\left\langle W,\left(\frac{\partial}{\partial t}+\phi\left(\frac{t}{\varepsilon}\right)\mathbf{k}\cdot\nabla_{\mathbf{x}}\right)\left(\sqrt{\varepsilon}\lambda_1^\varepsilon+\varepsilon\lambda_2^\varepsilon\right)+\sqrt{\varepsilon}\psi\left(\frac{t}{\varepsilon}\right)\nabla V\left(\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}}\lambda_2^\varepsilon\right\rangle\\&=\left\langle W,\left(\frac{\partial}{\partial t}+\mathbf{k}\cdot\nabla_{\mathbf{x}}\right)\lambda+\mathcal{L}_\Theta\lambda\right\rangle+\sqrt{\varepsilon}\langle W,\zeta_\varepsilon^\lambda\rangle,\end{aligned}$$

with

$$\begin{aligned}\zeta_\varepsilon^\lambda=&\left(\frac{\partial}{\partial t}+\phi\left(\frac{t}{\varepsilon}\right)\mathbf{k}\cdot\nabla_{\mathbf{x}}\right)\lambda_1\left(\mathbf{x},\frac{\mathbf{x}}{\varepsilon}\right)+\psi\left(\frac{t}{\varepsilon}\right)\nabla V\left(\frac{\mathbf{x}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}}\lambda_2\left(\mathbf{x},\frac{\mathbf{x}}{\varepsilon}\right)\\&+\sqrt{\varepsilon}\left(\frac{\partial}{\partial t}+\phi\left(\frac{t}{\varepsilon}\right)\mathbf{k}\cdot\nabla_{\mathbf{x}}\right)\lambda_2\left(\mathbf{x},\frac{\mathbf{x}}{\varepsilon}\right).\end{aligned}$$

Note that the terms $\mathbf{k}\cdot\nabla_{\mathbf{x}}\lambda_1$ above are understood as differentiation with respect to the slow variable $\mathbf{x}$ only and that the gradients are evaluated at $\mathbf{z}=\mathbf{x}/\varepsilon$. It follows that $G^\varepsilon_{\lambda_\varepsilon}$ is given by

$$G^\varepsilon_{\lambda_\varepsilon}(t)=\langle W(t),\lambda_\varepsilon\rangle-\int_0^t\left\langle W,\left(\frac{\partial}{\partial t}+\mathbf{k}\cdot\nabla_{\mathbf{x}}+\mathcal{L}_\Theta\right)\lambda\right\rangle(s)ds-\sqrt{\varepsilon}\int_0^t\langle W,\zeta_\varepsilon^\lambda\rangle(s)ds \tag{2.35}$$

and is a martingale with respect to the measure $\tilde{P}_\varepsilon$ defined on $C([0,T];X\times\mathcal{V})$. Lemma 2.5 and the estimate (2.21) follow from the following lemma.

LEMMA 2.6. *Let $\lambda\in C^1([0,T];\mathcal{S}(\mathbb{R}^{2d}))$. Then there exists a constant $C_\lambda>0$ independent of $t\in[0,T]$ so that the correctors $\lambda_1^\varepsilon(t)$ and $\lambda_2^\varepsilon(t)$ satisfy the uniform bounds*

$$\|\lambda_1^\varepsilon(t)\|_{L^\infty(\mathcal{V};L^2)}+\|\lambda_2^\varepsilon(t)\|_{L^\infty(\mathcal{V};L^2)}\le C_\lambda \tag{2.36}$$

and

$$\begin{aligned}&\Big\|\frac{\partial\lambda_1^\varepsilon(t)}{\partial t}+\phi\Big(\frac{t}{\varepsilon}\Big)\mathbf{k}\cdot\nabla_{\mathbf{x}}\lambda_1^\varepsilon(t)\Big\|_{L^\infty(\mathcal{V};L^2)}\\ &+\Big\|\frac{\partial\lambda_2^\varepsilon(t)}{\partial t}+\phi\Big(\frac{t}{\varepsilon}\Big)\mathbf{k}\cdot\nabla_{\mathbf{x}}\lambda_2^\varepsilon(t)\Big\|_{L^\infty(\mathcal{V};L^2)}+\|\nabla_{\mathbf{k}}\lambda_2^\varepsilon\|_{L^\infty(\mathcal{V};L^2)}\le C_\lambda.\end{aligned} \tag{2.37}$$

Proof of Lemma 2.6. The estimates (2.36) and (2.37) follow immediately from the explicit expressions (2.28) for λ_1 and (2.32), (2.33), (2.34) for λ_2, the gap property (2.4) and the almost sure a priori bounds (2.2). For example, we have

$$|\lambda_1(t,\mathbf{x},\mathbf{k},\mathbf{z},V)|\le\int_0^\infty e^{-\alpha s}\|\nabla V\|_{L^\infty}ds\,|\nabla_{\mathbf{k}}\lambda(t,\mathbf{x},\mathbf{k})|\le\frac{C_0}{\alpha}\,|\nabla_{\mathbf{k}}\lambda(t,\mathbf{x},\mathbf{k})|,$$

from which the first bound in (2.36) follows. The other estimates in Lemma 2.6 are shown in a similar if slightly more tedious fashion.

Proof of Lemma 2.5. Observe that (2.36) implies $|\langle W,\lambda\rangle-\langle W,\lambda_\varepsilon\rangle|\le C\sqrt{\varepsilon}$ for all $W\in X$ and $V\in\mathcal{V}$, while (2.37) implies that for all $t\in[0,T]$,

$$\|\zeta_\varepsilon^\lambda(t)\|_{L^2}\le C \tag{2.38}$$

for all $V\in\mathcal{V}$. So (2.21) follows from the fact that (2.35) is a martingale.

As explained in Section 2.3 the tightness of measures P_ε given by Lemma 2.3 implies that the expectation $\mathbb{E}\{W_\varepsilon(t,\mathbf{x},\mathbf{k})\}$ converges weakly in $L^2(\mathbb{R}^{2d})$ to the solution $\overline{W}(t,\mathbf{x},\mathbf{k})$ of the transport equation for each $t\in[0,T]$.

2.5. Convergence in probability. We now prove that for any test function λ the second moment $\mathbb{E}\{\langle W_\varepsilon,\lambda\rangle^2\}$ converges to $\langle\overline{W},\lambda\rangle^2$. This will imply the convergence in probability claimed in Theorem 2.2. The proof is based on constructing an appropriate approximate martingale for the quadratic functional $\langle W\otimes W,\mu\rangle$, where $\mu(t,\mathbf{x}_1,\mathbf{k}_1,\mathbf{x}_2,\mathbf{k}_2)$ is a test function, and $W\otimes W(t,\mathbf{x}_1,\mathbf{k}_1,\mathbf{x}_2,\mathbf{k}_2)=W(t,\mathbf{x}_1,\mathbf{k}_1)W(t,\mathbf{x}_2,\mathbf{k}_2)$. We need to consider the action of the infinitesimal generator on functions of W and V of the form

$$F(W,V)=\langle W(\mathbf{x}_1,\mathbf{k}_1)W(\mathbf{x}_2,\mathbf{k}_2),\mu(t,\mathbf{x}_1,\mathbf{k}_1,\mathbf{x}_2,\mathbf{k}_2,\hat{V})\rangle=\langle W\otimes W,\mu(V)\rangle,$$

where μ is a given function. The infinitesimal generator acts on such functions as

$$\frac{d}{dh}\mathbb{E}^{\tilde{P}_\varepsilon}_{W,V,t}\{\langle W\otimes W,\mu(V)\rangle\}(t+h)\Big|_{h=0}=\frac{1}{\varepsilon}\langle W\otimes W,Q\lambda\rangle+\langle W\otimes W,\mathcal{H}_2^\varepsilon\mu\rangle, \tag{2.39}$$

where

$$\mathcal{H}_2^\varepsilon\mu=\sum_{j=1}^2\phi\Big(\frac{t}{\varepsilon}\Big)\mathbf{k}^j\cdot\nabla_{\mathbf{x}^j}\mu-\frac{1}{\sqrt{\varepsilon}}\psi\Big(\frac{t}{\varepsilon}\Big)\nabla V\Big(\frac{\mathbf{x}^j}{\varepsilon}\Big)\cdot\nabla_{\mathbf{k}^j}\mu. \tag{2.40}$$

Therefore the functional

$$G_\mu^{2,\varepsilon} = \langle W\otimes W, \mu(V)\rangle(t) - \int_0^t \left\langle W\otimes W, \Big(\frac{1}{\varepsilon}Q + \frac{\partial}{\partial s} + \phi\Big(\frac{s}{\varepsilon}\Big)[\mathbf{k}_1\cdot\nabla_{\mathbf{x}_1} + \mathbf{k}_2\cdot\nabla_{\mathbf{x}_2}]\right\rangle ds$$
$$+ \int_0^t \left\langle W\otimes W, \frac{1}{\sqrt{\varepsilon}}\psi\Big(\frac{s}{\varepsilon}\Big)\left[\nabla V\left(\frac{\mathbf{x}^1}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}^1}\mu + \nabla V\left(\frac{\mathbf{x}^2}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}^2}\mu\right]\right\rangle ds, \tag{2.41}$$

is a $\tilde{P}^\varepsilon$ martingale. We let $\mu(t,\mathbf{X},\mathbf{K})\in\mathcal{S}(\mathbb{R}^{2d}\times\mathbb{R}^{2d})$ be a test function independent of V, where $\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2)$, and $\mathbf{K}=(\mathbf{k}_1,\mathbf{k}_2)$. We define an approximation

$$\mu_\varepsilon(t,\mathbf{X},\mathbf{K}) = \mu(t,\mathbf{X},\mathbf{K}) + \sqrt{\varepsilon}\mu_1(t,\mathbf{X},\mathbf{X}/\varepsilon,\mathbf{K}) + \varepsilon\mu_2(t,\mathbf{X},\mathbf{X}/\varepsilon,\mathbf{K}).$$

We will use the notation $\mu_1^\varepsilon(t,\mathbf{X},\mathbf{K}) = \mu_1(t,\mathbf{X},\mathbf{X}/\varepsilon,\mathbf{K})$ and $\mu_2^\varepsilon(t,\mathbf{X},\mathbf{K}) = \mu_2(t,\mathbf{X},\mathbf{X}/\varepsilon,\mathbf{K})$. The functions μ_1 and μ_2 are to be determined. We now use (2.39) to get

$$D_\varepsilon := \frac{d}{dh}\Big|_{h=0}\mathbb{E}_{W,V,t}(\langle W\otimes W,\mu_\varepsilon(V)\rangle)(t+h) = \frac{1}{\varepsilon}\left\langle W\otimes W, \left(\frac{\partial}{\partial\tau} + Q + \phi(\tau)\sum_{j=1}^2 \mathbf{k}^j\cdot\nabla_{\mathbf{z}^j}\right)\mu\right\rangle$$
$$+\frac{1}{\sqrt{\varepsilon}}\left\langle W\otimes W, \left(\frac{\partial}{\partial\tau} + Q + \phi(\tau)\sum_{j=1}^2 \mathbf{k}^j\cdot\nabla_{\mathbf{z}^j}\right)\mu_1 - \psi(\tau)\sum_{j=1}^2 \nabla V(\mathbf{z}^j)\cdot\nabla_{\mathbf{k}^j}\mu\right\rangle$$
$$+\left\langle W\otimes W, \left(\frac{\partial}{\partial\tau} + Q + \phi(\tau)\sum_{j=1}^2 \mathbf{k}^j\cdot\nabla_{\mathbf{z}^j}\right)\mu_2 - \psi(\tau)\sum_{j=1}^2 \nabla V(\mathbf{z}^j)\cdot\nabla_{\mathbf{k}^j}\mu_1\right\rangle$$
$$+\left\langle W\otimes W, \frac{\partial\mu}{\partial t} + \phi(\tau)\sum_{j=1}^2 \mathbf{k}^j\cdot\nabla_{\mathbf{x}^j}\mu\right\rangle$$
$$+\sqrt{\varepsilon}\left\langle W\otimes W, \left(\frac{\partial}{\partial t} + \phi(\tau)\sum_{j=1}^2 \mathbf{k}^j\cdot\nabla_{\mathbf{x}^j}\right)(\mu_1+\sqrt{\varepsilon}\mu_2) - \psi(\tau)\sum_{j=1}^2 \nabla V(\mathbf{z}^j)\cdot\nabla_{\mathbf{k}^j}\mu_2\right\rangle.$$

The above expression is evaluated at $\mathbf{z}_j = \mathbf{x}_j/\varepsilon$ and $\tau = t/\varepsilon$. The term of order ε^{-1} in D_ε vanishes since μ is independent of V and of the fast variables τ and $\mathbf{z}$. We cancel the term of order $\varepsilon^{-1/2}$ as before by defining μ_1 as the unique mean-zero (in the variables V, τ and $\mathbf{Z}=(\mathbf{z}_1,\mathbf{z}_2)$) solution of

$$\Big(\frac{\partial}{\partial\tau} + Q + \phi(\tau)\sum_{j=1}^2 \mathbf{k}^j\cdot\nabla_{\mathbf{z}^j}\Big)\mu_1 - \psi(\tau)\sum_{j=1}^2 \nabla V(\mathbf{z}^j)\cdot\nabla_{\mathbf{k}^j}\mu = 0. \tag{2.42}$$

It is given explicitly by

$$\mu_1(\tau,\mathbf{X},\mathbf{Z},\mathbf{K},V) = -\int_0^\infty e^{sQ}\psi(\tau+s)\big[\nabla V(\mathbf{z}^1 + \mathbf{k}^1[\Phi(\tau+s)-\Phi(\tau)])\cdot\nabla_{\mathbf{k}^1}\mu$$
$$+\nabla V(\mathbf{z}^2 + \mathbf{k}^2[\Phi(\tau+s)-\Phi(\tau)])\cdot\nabla_{\mathbf{k}^2}\mu\big]\,ds.$$

When μ has the form $\mu=\lambda\otimes\lambda$, then μ_1 has the form $\mu_1=\lambda_1\otimes\lambda+\lambda\otimes\lambda_1$ with the corrector λ_1 given by (2.28). Let us also define μ_2 as a sum of three terms, as in (2.30),

$$\mu_2 = \mu_{21} + \mu_{22} + \mu_{23}.$$

The function μ_{23} is the solution of

$$\frac{\partial \mu_{23}}{\partial \tau} + Q\mu_{23} + \phi(\tau)\sum_{j=1}^{2} \mathbf{k}^j \cdot \nabla_{\mathbf{z}^j} \mu_{23} = [1-\phi(\tau)][\mathbf{k}_1 \cdot \nabla_{\mathbf{z}^1}\mu + \mathbf{k}_2 \cdot \nabla_{\mathbf{z}^2}\mu].$$

It is given explicitly by

$$\mu_{23}(t,\tau,\mathbf{x},\mathbf{k}) = [\tau - \Phi(\tau)][\mathbf{k}_1 \cdot \nabla_{\mathbf{z}^1}\mu + \mathbf{k}_2 \cdot \nabla_{\mathbf{z}^2}\mu]. \tag{2.43}$$

The function μ_{21} is the mean zero solution, with respect to the invariant measure $\pi(V)$, of

$$\left(\frac{\partial}{\partial \tau} + Q + \phi(\tau)\sum_{j=1}^{2} \mathbf{k}^j \cdot \nabla_{\mathbf{z}^j}\right)\mu_{21} = \psi(\tau)\sum_{j=1}^{2} \nabla V(\mathbf{z}^j)\cdot \nabla_{\mathbf{k}^j}\mu_1 + \mathcal{L}^{\tau}_{2,\Theta}\mu, \tag{2.44}$$

where $\mathcal{L}^{\tau}_{2,\Theta}\mu$ is given by

$$\mathcal{L}^{\tau}_{2,\Theta}\mu = \mathbb{E}\Big\{-\psi(\tau)\sum_{j=1}^{2} \nabla V(\mathbf{z}^j)\cdot \nabla_{\mathbf{k}^j}\mu_1\Big\}.$$

We now compute this operator explicitly:

$$\begin{aligned}
\mathcal{L}^{\tau}_{2,\Theta}\mu &= \mathbb{E}\Big\{-\psi(\tau)\sum_{j=1}^{2} \nabla V(\mathbf{z}^j)\cdot \nabla_{\mathbf{k}^j}\mu_1\Big\} \\
&= \mathbb{E}\Big\{\psi(\tau)\sum_{j,l=1}^{2}\sum_{n,m=1}^{d} \frac{\partial V(\mathbf{z}^j)}{\partial z_n^j}\frac{\partial}{\partial k_n^j}\left(\int_0^{\infty} e^{sQ}\psi(\tau+s)\frac{\partial V(\mathbf{z}^l + \mathbf{k}^l[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_m^l}\frac{\partial \mu}{\partial k_m^{\,l}}ds\right)\Big\} \\
&= \sum_{j,l=1}^{2}\sum_{n,m=1}^{d} \frac{\partial}{\partial k_n^j}\left(\int_0^{\infty}\psi(\tau+s)\psi(\tau)\frac{\partial^2 R(s,\mathbf{z}^l - \mathbf{z}^j + \mathbf{k}^l[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_n^j \partial z_m^l}\frac{\partial \mu}{\partial k_m^{\,l}}ds\right).
\end{aligned}$$

Note that unlike the first moment calculation now the operator $\mathcal{L}^{\tau}_{2,\Theta}$ depends also on the fast spatial variable $\mathbf{Z}$ and not only on the fast time τ. The difference is that while the dependence on $\tau \in [0,\Theta]$ may be averaged out by integration over the period, the dependence on $\mathbf{Z} \in \mathbb{R}^d$ has to be treated differently. The function μ_{21} may be written explicitly as

$$\mu_{21}(t,\tau,\mathbf{X},\mathbf{Z},\mathbf{k},V) = -\int_0^{\infty} e^{sQ} g_{21}(t,\tau+s,\mathbf{X},\mathbf{Z}+(\Phi(\tau+s)-\Phi(\tau))\mathbf{K},\mathbf{K},V)ds \tag{2.45}$$

with

$$g_{21}(t,\tau,\mathbf{X},\mathbf{Z},\mathbf{k},V) = \psi(\tau)\sum_{j=1}^{2} \nabla V(\mathbf{z}^j)\cdot \nabla_{\mathbf{k}^j}\mu_1 + \mathcal{L}^{\tau}_{2,\Theta}\mu. \tag{2.46}$$

Finally, μ_{22} is the solution of

$$\left(\frac{\partial}{\partial \tau} + Q + \phi(\tau)\sum_{j=1}^{2} \mathbf{k}^j \cdot \nabla_{\mathbf{z}^j}\right)\mu_{22} = \mathcal{L}_{2,\Theta}\mu - \mathcal{L}^{\tau}_{2,\Theta}\mu$$

with

$$\mathcal{L}_{2,\Theta}\mu(t,\mathbf{X},\mathbf{Z},\mathbf{K}) = \frac{1}{\Theta}\int_0^\Theta \mathcal{L}_{2,\Theta}^\tau \mu d\tau$$
$$= \frac{1}{\Theta}\int_0^\Theta \sum_{j,l=1}^{2}\sum_{n,m=1}^{d} \frac{\partial}{\partial k_n^j}\left(\int_0^\infty \psi(\tau+s)\psi(\tau)\frac{\partial^2 R(s,\mathbf{z}^l-\mathbf{z}^j+\mathbf{k}^l[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_n^j \partial z_m^l}\frac{\partial \mu}{\partial k_m^{\,l}} ds d\tau\right). \tag{2.47}$$

The function μ_{22} may be written explicitly as

$$\mu_{22}(t,\tau,\mathbf{X},\mathbf{Z},\mathbf{K}) = \int_0^\tau [\mathcal{L}_{2,\Theta}^\varepsilon \mu - \mathcal{L}_{2,\Theta}^{\varepsilon,\zeta}]\mu(t,\mathbf{X},\mathbf{K})d\zeta. \tag{2.48}$$

The $\tilde{P}^\varepsilon$-martingale $G_{\mu_\varepsilon}^{2;\varepsilon}$ is given by

$$G_\mu^{2,\varepsilon} = \langle W\otimes W, \mu(V)\rangle(t)$$
$$-\int_0^t \left\langle W\otimes W, \left(\frac{\partial}{\partial t}+\mathbf{K}\cdot\nabla_{\mathbf{X}}+\mathcal{L}_{2,\Theta}^\varepsilon\right)\mu\right\rangle(s)ds - \sqrt{\varepsilon}\int_0^t \langle W\otimes W, \zeta_\varepsilon^\mu\rangle(s)ds, \tag{2.49}$$

where ζ_ε^μ is given by

$$\zeta_\mu^\varepsilon = \left(\frac{\partial}{\partial t}+\phi(\tau)\sum_{j=1}^{2}\mathbf{k}^j\cdot\nabla_{\mathbf{x}^j}\right)(\mu_1+\sqrt{\varepsilon}\mu_2) - \psi(\tau)\sum_{j=1}^{2}\nabla V(\mathbf{z}^j)\cdot\nabla_{\mathbf{k}^j}\mu_2$$

and the operator $\mathcal{L}_{2,\Theta}^\varepsilon$ is defined by (2.47) with $\mathbf{Z}=\mathbf{X}/\varepsilon$:

$$\mathcal{L}_{2,\Theta}^\varepsilon \mu(t,\mathbf{X},\mathbf{K})$$
$$= \sum_{j,l=1}^{2}\sum_{n,m=1}^{d}\frac{\partial}{\partial k_n^j}\left(\int_0^\Theta\int_0^\infty \psi(\tau+s)\psi(\tau)\frac{\partial^2 R(s,\frac{\mathbf{x}^l-\mathbf{x}^j}{\varepsilon}+\mathbf{k}^l[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_n^j\partial z_m^l}\frac{\partial\mu}{\partial k_m^{\,l}}\frac{ds d\tau}{\Theta}\right). \tag{2.50}$$

The bound on ζ_ε^μ is similar to that on $\zeta_\varepsilon^\lambda$ obtained previously as the correctors μ_j^ε satisfy the same type of estimates as the correctors λ_j:

LEMMA 2.7. *There exists a constant $C_\mu>0$ so that the functions $\mu_{1,2}^\varepsilon$ obey the uniform bounds*

$$\|\mu_1^\varepsilon(t)\|_{L^2(\mathbb{R}^{2d})} + \|\mu_2^\varepsilon(t)\|_{L^2(\mathbb{R}^{2d})} \le C_\mu \tag{2.51}$$

and

$$\left\|\frac{\partial\mu_1^\varepsilon(t)}{\partial t}+\mathbf{K}\cdot\nabla_{\mathbf{X}}\mu_1^\varepsilon(t)\right\|_{L^2(\mathbb{R}^{2d})} + \|\nabla_{\mathbf{K}}\mu_1^\varepsilon(t)\|_{L^2(\mathbb{R}^{2d})}$$
$$+\left\|\frac{\partial\mu_2^\varepsilon(t)}{\partial t}+\mathbf{K}\cdot\nabla_{\mathbf{X}}\mu_2^\varepsilon(t)\right\|_{L^2(\mathbb{R}^{2d})} + \|\nabla_{\mathbf{K}}\mu_2^\varepsilon(t)\|_{L^2(\mathbb{R}^{2d})} \le C_\mu \tag{2.52}$$

for all $t\in[0,T]$ and $V\in\mathcal{V}$. The proof of this lemma is very similar to that of Lemma 2.6 and is therefore omitted.

Unlike the first moment case, the averaged operator $\mathcal{L}_2^\varepsilon$ still depends on ε. We cannot claim yet strong convergence of the $\tilde{P}^\varepsilon$-martingale $G_{\mu_\varepsilon}^{2,\varepsilon}$ to its limit. However,

the a priori bound on W_ε in L^2 allows us to characterize the limit of $G^{2,\varepsilon}_{\mu_\varepsilon}$ and to show strong convergence. This is done as follows. The terms in (2.50) with $l=j$ are independent of ε and give the contribution:

$$\begin{aligned}&\mathcal{L}^0_{2,\Theta}\mu(t,\mathbf{X},\mathbf{K})\\&=\sum_{j=1}^{2}\sum_{n,m=1}^{d}\frac{\partial}{\partial k_n^j}\left(\int_0^\Theta\int_0^\infty\psi(\tau+s)\psi(\tau)\frac{\partial^2 R(s,\mathbf{k}^j[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_n^j\partial z_m^j}\frac{\partial\mu}{\partial k_m^{\,j}}\frac{dsd\tau}{\Theta}\right).\end{aligned}$$

The two remaining terms give a contribution that tend to 0 as $\varepsilon\to 0$ for sufficiently smooth test functions. They are given by

$$\begin{aligned}&\mathcal{L}^{12,\varepsilon}_{2,\Theta}\mu(t,\mathbf{X},\mathbf{K})\\&=\sum_{n,m=1}^{d}\frac{\partial}{\partial k_n^2}\left(\int_0^\Theta\int_0^\infty\psi(\tau+s)\psi(\tau)\frac{\partial^2 R(s,\frac{\mathbf{x}^1}{\varepsilon}-\frac{\mathbf{x}^2}{\varepsilon}+\mathbf{k}^1[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_n^2\partial z_m^1}\frac{\partial\mu}{\partial k_m^{\,1}}\frac{dsd\tau}{\Theta}\right)\\&+\sum_{n,m=1}^{d}\frac{\partial}{\partial k_n^1}\left(\int_0^\Theta\int_0^\infty\psi(\tau+s)\psi(\tau)\frac{\partial^2 R(s,\frac{\mathbf{x}^2}{\varepsilon}-\frac{\mathbf{x}^1}{\varepsilon}+\mathbf{k}^2[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_n^1\partial z_m^2}\frac{\partial\mu}{\partial k_m^{\,2}}\frac{dsd\tau}{\Theta}\right)\\&=I+II.\end{aligned}\tag{2.53}$$

The first term above may be written as

$$I_\varepsilon(t,\mathbf{X},\mathbf{K})=q\left(\frac{\mathbf{x}^1-\mathbf{x}^2}{\varepsilon},\mathbf{k}^2\right)\frac{\partial^2\mu(t,\mathbf{X},\mathbf{K})}{\partial k_n^2\partial k_m^1},$$

with

$$q(\mathbf{x},\mathbf{k})=\int_0^\Theta\int_0^\infty\psi(\tau+s)\psi(\tau)\frac{\partial^2 R(s,\mathbf{x}+\mathbf{k}[\Phi(\tau+s)-\Phi(\tau)])}{\partial z_n^1\partial z_m^2}\frac{dsd\tau}{\Theta}.$$

We observe that

$$\begin{aligned}&\int|q^2(\mathbf{x},\mathbf{k})|^2d\mathbf{x}=\int_{\mathbb{R}^d}\int_0^\Theta\int_0^\Theta\int_0^\infty\int_0^\infty\psi(\tau_1+s_1)\psi(\tau_1)\psi(\tau_2+s_2)\psi(\tau_2)\\&\times g(s_1,\mathbf{x}+\mathbf{k}[\Phi(\tau_1+s_1)-\Phi(\tau_1)])g(s_2,\mathbf{x}+\mathbf{k}[\Phi(\tau_2+s_2)-\Phi(\tau_2)])\frac{ds_1ds_2d\tau_1d\tau_2}{\Theta^2}d\mathbf{x}\\&\le C\left(\int_0^\infty\|g(s)\|_2ds\right)^2,\end{aligned}$$

where

$$g(s,\mathbf{x})=\frac{\partial^2 R(s,\mathbf{x})}{\partial x_n^1\partial x_m^2}.$$

We may assume without loss of generality, using classical density arguments, that $\dfrac{\partial^2\mu(t,\mathbf{X},\mathbf{K})}{\partial k_n^2\partial k_m^1}=\eta(t,\mathbf{x}_1-\mathbf{x}_2,\mathbf{k}_1)\eta'(t,(\mathbf{x}_1+\mathbf{x}_2)/2,\mathbf{k}_2)$ with sufficiently regular functions η and η'. Then we obtain

$$\begin{aligned}&\|I_\varepsilon\|_{L^2}=\int\left|q\left(\frac{\mathbf{x}^1-\mathbf{x}^2}{\varepsilon}\right)\right|^2|\eta(t,\mathbf{x}_1-\mathbf{x}_2,\mathbf{k}_1)|^2|\eta'(t,(\mathbf{x}_1+\mathbf{x}_2)/2,\mathbf{k}_2)|^2d\mathbf{x}^1d\mathbf{x}^2d\mathbf{k}^1d\mathbf{k}^2\\&\le\varepsilon^d\|q\|_{L^2_{\mathbf{x}}}\|\eta'\|_{L^2_{\mathbf{x},\mathbf{k}}}\int\sup_{\mathbf{x}}|\eta(t,\mathbf{x},\mathbf{k})|^2d\mathbf{k}.\end{aligned}$$

A similar bound holds also for the second term in (2.53). This proves that $\|(\mathcal{L}^{\varepsilon}_{2,\Theta} - \mathcal{L}^{0}_{2,\Theta})\mu\|_{L^2} \to 0$ as $\varepsilon \to 0$.

We therefore deduce that

$$G^2_\mu = \langle W\otimes W, \mu(\hat{V})\rangle(z) - \int_0^z \left\langle W\otimes W, \Big(\frac{\partial}{\partial z} + \mathbf{k}_1\cdot\nabla_{\mathbf{x}_1} + \mathbf{k}_2\cdot\nabla_{\mathbf{x}_2} + \mathcal{L}^0_{2,\Theta}\Big)\mu\right\rangle(s)ds$$

is an approximate $\tilde{P}_\varepsilon$ martingale. The limit of the second moment

$$W_2(t,\mathbf{x}_1,\mathbf{k}_1,\mathbf{x}_2,\mathbf{k}_2) = \mathbb{E}^P\{W(t,\mathbf{x}_1,\mathbf{k}_1)W(t,\mathbf{x}_2,\mathbf{k}_2)\}$$

thus satisfies (weakly) the transport equation

$$\frac{\partial W_2}{\partial t} + (\mathbf{k}_1\cdot\nabla_{\mathbf{x}_1} + \mathbf{k}_2\cdot\nabla_{\mathbf{x}_2})W_2 = \mathcal{L}^0_{2,\Theta}W_2,$$

with initial data $W_2(0,\mathbf{x}_1,\mathbf{k}_1,\mathbf{x}_2,\mathbf{k}_2) = W_0(\mathbf{x}_1,\mathbf{k}_1)W_0(\mathbf{x}_2,\mathbf{k}_2)$. Moreover, the operator $\mathcal{L}^0_{2,\Theta}$ acting on a tensor product $\lambda\otimes\lambda$ has the form

$$\mathcal{L}^0_{2,\Theta}[\lambda\otimes\lambda] = \mathcal{L}_\Theta\lambda\otimes\lambda + \lambda\otimes\mathcal{L}_\Theta\lambda.$$

This implies that

$$\mathbb{E}^P\{W(t,\mathbf{x}_1,\mathbf{k}_1)W(t,\mathbf{x}_2,\mathbf{k}_2)\} = \mathbb{E}^P\{W(t,\mathbf{x}_1,\mathbf{k}_1)\}\mathbb{E}^P\{W(t,\mathbf{x}_2,\mathbf{k}_2)\}$$

by uniqueness of the solution to the above transport equation with initial conditions given by $W_0(\mathbf{x}_1,\mathbf{k}_1)W_0(\mathbf{x}_2,\mathbf{k}_2)$. This proves that the limiting measure P is deterministic and unique (because characterized by the transport equation) and that the sequence $W_\varepsilon(t,\mathbf{x},\mathbf{k})$ converges in probability to $W(t,\mathbf{x},\mathbf{k})$.

3. An efficient time splitting algorithm

3.1. A modified time splitting scheme. The results of the preceding sections imply that $\Theta_\varepsilon = \varepsilon/N$ with $N \gg 1$ is necessary to obtain the correct statistics of the wave energy density by solving the time splitting algorithm (2.9)-(2.10). We now show that the interval Θ_ε can be chosen substantially larger if one modifies the treatment of the scattering term in an appropriate manner in the time splitting scheme. However, unlike (2.9)-(2.10) the modified scheme can no longer be solved explicitly and requires a more complicated numerical scheme to handle the scattering part.

Let us consider a time splitting algorithm for (2.1), which differs from (2.9)-(2.10) as follows. The advection operator $\mathbf{k}\cdot\nabla_{\mathbf{x}}$ in (2.1) is treated in the same way as in (2.9): given the approximation $\tilde{W}_\varepsilon(n\Theta,\mathbf{x},\mathbf{k})$ we first solve

$$\frac{\partial U^{n+1}}{\partial t} + \mathbf{k}\cdot\nabla_{\mathbf{x}}U^{n+1} = 0, \quad n\Theta \le t \le (n+1)\Theta, \tag{3.1}$$

with the initial data $U^{n+1}(n\Theta,\mathbf{x},\mathbf{k}) = \tilde{W}_\varepsilon(n\Theta,\mathbf{x},\mathbf{k})$. We have explicitly as before

$$U^{n+1}(t,\mathbf{x},\mathbf{k}) = \tilde{W}_\varepsilon(n\Theta,\mathbf{x}-(t-n\Theta)\mathbf{k},\mathbf{k}), \quad n\Theta \le t \le (n+1)\Theta.$$

The scattering term is now accounted for in a different fashion from (2.10). The new scattering equation is

$$\frac{\partial Z^{n+1}}{\partial t} = \frac{1}{\sqrt{\varepsilon}}\nabla_{\mathbf{x}}V\left(\frac{t}{\varepsilon}, \frac{\mathbf{x}-(t-n\Theta)\mathbf{k}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}}Z^{n+1}, \quad n\Theta \le t \le (n+1)\Theta, \tag{3.2}$$

with the initial data $Z^{n+1}(n\Theta,\mathbf{x},\mathbf{k}) = U^{n+1}((n+1)\Theta,\mathbf{x},\mathbf{k})$. We then reset the approximation as $\tilde{W}_\varepsilon((n+1)\Theta,\mathbf{x},\mathbf{k}) = Z^{n+1}((n+1)\Theta,\mathbf{x},\mathbf{k})$ and repeat the above scheme. Note that (3.2) is different from (2.10) as the random potential V is evaluated at the point $\mathbf{x}-(t-n\Theta_\varepsilon)\mathbf{k}$ instead of $\mathbf{x}$. This means, roughly speaking, that we account for the advection of the rapidly varying part of the right side in (3.2), which depends on the oscillatory potential V, but we do not advect W_ε, which we hope to be statistically slowly varying in the limit $\varepsilon \to 0$. This modification allows us to obtain the right dynamics with much larger Θ_ε (actually independent of ε) than in previous sections because indeed W_ε has a slowly varying limit as $\varepsilon \to 0$.

However, as we mentioned above the possibility to take a large time step comes at a price: the modified equation (3.2) has no explicit solution, unlike (2.10). The main advantage of the modification (2.10) is that it still allows us to bypass the advection part in the $\mathbf{x}$-variable, which has to be performed much less often than in the time-splitting scheme (2.9)-(2.10).

3.2. Convergence of the time-splitting scheme for the Wigner equation. We analyze now convergence of the time-splitting algorithm (3.1)-(3.2) in the small ε limit with a time-step Θ independent of ε. As in the analysis in Section 2 it is convenient to introduce a somewhat more general set-up including (3.1)-(3.2) as a particular example. We modify (2.1) as follows:

$$\frac{\partial W_\varepsilon}{\partial t} + \phi(t)\mathbf{k}\cdot\nabla_{\mathbf{x}} W_\varepsilon = \frac{\psi(t)}{\sqrt{\varepsilon}}\nabla_{\mathbf{x}} V\left(\frac{t}{\varepsilon}, \frac{\mathbf{x}-\eta(t)\mathbf{k}}{\varepsilon}\right)\cdot\nabla_{\mathbf{k}} W_\varepsilon. \tag{3.3}$$

This is the analog to the modified Wigner equation (2.12). Once again, choosing the functions ϕ and ψ equal to zero on alternating time intervals in (3.3) leads to a genuine time-splitting scheme. In general the periodic functions ϕ and ψ are as in Section 2 with period Θ and with mean value equal to one. However, there is an important difference between the general time-splitting in (3.3) and that in (2.12): now the functions ϕ and ψ vary on the macroscopic rather than the microscopic time-scale. This corresponds to taking a time-step Θ independent of ε in the time-splitting scheme. However, in order to allow for such a large time step one has to modify the oscillatory phase in the operator on the right side of (3.3) by means of a function $\eta(t)$ also varying on the macroscopic scale. We choose

$$\eta(t) = \int_0^t [\phi(s)-1]ds = \Phi(t) - t. \tag{3.4}$$

This allows us to compensate for the large time-step by appropriately adjusting the potential V accounting indirectly for advection during the long time-step. This modification need not be made in the argument of W_ε since the latter has a macroscopic limit. The main result is as follows.

THEOREM 3.1. *Let the initial data $W_\varepsilon^0(\mathbf{x},\mathbf{k})$ for (2.12) converge to $W_0(\mathbf{x},\mathbf{k})$ strongly in $L^2(\mathbb{R}^{2d})$ and the functions ϕ and ψ satisfy the normalization (2.13) and be uniformly bounded. Then the modified Wigner distribution W_ε, solution of (3.3) converges in probability and weakly in $L^2(\mathbb{R}^d)$ to the solution $\overline{W}$ of the modified Fokker-Planck equation*

$$\frac{\partial \overline{W}}{\partial t} + \phi(t)\mathbf{k}\cdot\nabla\overline{W} = \psi^2(t)\frac{\partial}{\partial k_m}\left(D_{mn}(\mathbf{k})\frac{\partial W}{\partial k_n}\right), \tag{3.5}$$

with diffusion matrix

$$D_{mn}(\mathbf{k}) = -\int_0^\infty \frac{\partial^2 R(s,s\mathbf{k})}{\partial z_n \partial z_m} ds, \tag{3.6}$$

and initial data $W_0(\mathbf{x},\mathbf{k})$. *More precisely, for any test function* $\lambda \in L^2(\mathbb{R}^d)$ *the random process*

$$\langle W, \lambda(t)\rangle = \int_{\mathbb{R}^{2d}} W_\varepsilon(t,\mathbf{x},\mathbf{k})\lambda(\mathbf{x},\mathbf{k}) d\mathbf{x} d\mathbf{k}$$

converges in probability to $\langle \overline{W}, \lambda\rangle$ *as* $\varepsilon \to 0$ *uniformly on finite time intervals* $t \in [0,T]$.

Equation (3.5) is nothing but a time-splitting approximation with time-step Θ of the correct limiting Fokker-Planck equation (2.7), which has the form

$$\frac{\partial \overline{W}}{\partial t} + \mathbf{k}\cdot\nabla \overline{W} = \frac{\partial}{\partial k_m}\left(D_{mn}(\mathbf{k}) \frac{\partial W}{\partial k_n}\right). \tag{3.7}$$

In particular we obtain the correct diffusion matrix for all Θ.

The proof of Theorem 3.1 is very similar to that of Theorem 2.2. We only explain the necessary modifications. One no longer needs to introduce separately the fast spatial and temporal variables $\mathbf{z} = \mathbf{x}/\varepsilon$ and $\tau = t/\varepsilon$ in the construction of the correctors. The new fast variable is $\mathbf{z} = (\mathbf{x} - \eta(t)\mathbf{k})/\varepsilon$ so that one formally has

$$\nabla_{\mathbf{x}} \to \nabla_{\mathbf{x}} + \frac{1}{\varepsilon}\nabla_{\mathbf{z}}, \quad \frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - \frac{1}{\varepsilon}\dot\eta(t)\mathbf{k}\cdot\nabla_{\mathbf{z}}.$$

With our choice (3.4) of $\eta(t)$, the equation for the corrector λ_1 takes a particularly simple form:

$$\mathbf{k}\cdot\nabla_{\mathbf{z}}\lambda_1 + Q\lambda_1 = \nabla_{\mathbf{x}} V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda(\mathbf{x},\mathbf{k})\psi(t), \tag{3.8}$$

since the function $\phi(t) - \dot\eta(t)$ that would multiply the $\mathbf{k}\cdot\nabla_{\mathbf{z}}$ term on the left side is equal to one identically. The function λ_1 is given explicitly by

$$\lambda_1(\mathbf{x},\mathbf{k},\mathbf{z},V) = -\int_0^\infty e^{sQ}\nabla_{\mathbf{x}} V(\mathbf{z}+s\mathbf{k})\cdot\nabla_{\mathbf{k}}\lambda(\mathbf{x},\mathbf{k})\psi(t) ds.$$

The integral above is convergent because of the gap property (2.4). Then as in the proof of Theorem 2.2 the right side of the limit equation is given by:

$$\mathcal{L}\lambda = \psi(t)\mathbb{E}\{\nabla_{\mathbf{x}} V(\mathbf{z})\cdot\nabla_{\mathbf{k}}\lambda_1\} = \frac{\partial}{\partial k_m}\left(\mathbb{E}\left\{-\frac{\partial V(\mathbf{z})}{\partial z_m}\int_0^\infty e^{sQ}\frac{\partial V(\mathbf{z}+s\mathbf{k})}{\partial z_n}\frac{\partial \lambda(\mathbf{x},\mathbf{k})}{\partial k_n}\psi^2(t) ds\right\}\right)$$
$$= -\frac{\partial}{\partial k_m}\left(\int_0^\infty \frac{\partial^2 R(s,s\mathbf{k})}{\partial z_n \partial z_m} ds \frac{\partial \lambda}{\partial k_n}\right) = \frac{\partial}{\partial k_m}\left(D_{mn}(\mathbf{k})\frac{\partial \lambda}{\partial k_n}\right).$$

The rest of the proof of Theorem 3.1 is very similar to that of Theorem 2.2 and we omit the details.

Acknowledgment. This work was supported by the ONR grant N00014-02-1-0089, DARPA-ONR grant N00014-04-1-0224, NSF Grants DMS-0239097 (GB) and DMS-0203537 (LR), and two Alfred P. Sloan Fellowships.

REFERENCES

[1] G. Bal, T. Komorowski and L. Ryzhik, *Self-averaging of Wigner transforms in random media*, Comm. Math. Phys., 242, 81–135, 2003.
[2] G. Bal, G. Papanicolaou, and L. Ryzhik, *Self-averaging in time reversal for the parabolic wave equation*, Stochastics and Dynamics, 4 , 507–531, 2002.
[3] G. Bal and L. Ryzhik, *Time Splitting for Wave Equations in Random Media*, Preprint, 2004, to appear in M2AN Math. Model. Numer. Anal., 2005.
[4] W. Bao, S. Jin, and P. A. Markowich, *On Time-Splitting spectral approximations for the Schrödinger equation in the semiclassical regime*, J. Comp. Phys., 175, 487–524, 2002.
[5] S. Chandrasekhar, *Radiative Transfer*, Dover Publications, New York, 1960.
[6] G. C. Cohen, *Higher-Order Numerical Methods for Transient Wave Equations*, Scientific Computation, Springer Verlag, Berlin, 2002.
[7] D. R. Durran, *Nunerical Methods for Wave Equations in Geophysical Fluid Dynamics*, Springer, New York, 1999.
[8] A. Fannjiang, *Self-averaging in scaling Limits for random high-frequency parabolic waves*, Preprint, 2003.
[9] P. Gérard, P. A. Markowich, N. J. Mauser, and F. Poupaud, *Homogenization limits and Wigner transforms*, Comm. Pure Appl. Math., 50, 323–380, 1997.
[10] F. Golse, S. Jin, and C. D. Levermore, *The convergence of numerical transfer schemes in diffusive regimes. I. Discrete-ordinate method.*, SIAM J. Numer. Anal., 36, 1333–1369, 1999.
[11] T. Y. Hou, X. Wu, and Z. Cai, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp., 227, 913–943, 1999.
[12] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, New York, Academics, 1978.
[13] H. Kesten and G. Papanicolaou, *A limit theorem for stochastic acceleration*, Comm. Math. Phys., 78, 19–63, 1980.
[14] P. Markowich, P. Pietra, and C. Pohl, *Numerical approximation of quadratic observables of Schrödinger-type equations in the semi-classical limit*, Numer. Math., 81, 595–630, 1999.
[15] P. Markowich, P. Pietra, C. Pohl, and H. P. Stimming, *A Wigner-measure analysis of the Dufort-Frankel scheme for the Schrödinger equation*, SIAM J. Numer. Anal., 40, 1281–1310, 2002.
[16] L. Ryzhik, G. Papanicolaou, and J. B. Keller, *Transport equations for elastic and other waves in random media*, Wave Motion, 24, 327–370, 1996.
[17] H. Sato and M. C. Fehler, *Seismic wave propagation and scattering in the heterogeneous earth*, AIP Series in Modern Acoustics and Signal Processing, AIP Press, Springer, New York, 1998.
[18] P. Sheng, *Introduction to Wave Scattering, Localization and Mesoscopic Phenomena*, Academic Press, New York, 1995.
[19] G. Strang, *On the construction and comparison of difference schemes*, SIAM J. Numer. Anal., 5, 507–517, 1968.