The Annals of Applied Probability
1999, Vol. 9, No. 4, 951–1020

MULTISCALE DIFFUSION PROCESSES WITH PERIODIC COEFFICIENTS AND AN APPLICATION TO SOLUTE TRANSPORT IN POROUS MEDIA[1]

BY RABI BHATTACHARYA

Indiana University

Consider diffusions on $\mathbb{R}^k$, $k > 1$, governed by the Itô equation $dX(t) = \{b(X(t)) + \beta(X(t)/a)\}\,dt + \sigma dB(t)$, where b, β are periodic with the same period and are divergence free, σ is nonsingular and a is a large integer. Two distinct Gaussian phases occur as time progresses. The initial phase is exhibited over times $1 \ll t \ll a^{2/3}$. Under a geometric condition on the velocity field β, the final Gaussian phase occurs for times $t \gg a^2(\log a)^2$, and the dispersion grows quadratically with a. Under a complementary condition, the final phase shows up at times $t \gg a^4(\log a)^2$, or $t \gg a^2 \log a$ under additional conditions, with no unbounded growth in dispersion as a function of scale. Examples show the existence of non-Gaussian intermediate phases. These probabilisitic results are applied to analyze a multiscale Fokker–Planck equation governing solute transport in periodic porous media. In case b, β are not divergence free, some insight is provided by the analysis of one-dimensional multiscale diffusions with periodic coefficients.

1. Introduction. In this article we consider phase changes with time for diffusions on $\mathbb{R}^k$ with multiple scale periodic drifts $b(x) + \beta(x/a)$,

$$X(t) = X(0) + \int_0^t \{b(X(s)) + \beta(X(s)/a)\}\,ds + \int_0^t \sigma(X(s))\,dB(s), \tag{1.1}$$

with $\sigma(\cdot)$ a nonsingular matrix–valued function, a being a large spatial scale parameter. Computations of these phase change and their time scales are carried out directly for some examples in Section 6, without requiring the machinery needed for the general case, and the reader may perhaps take a look at these first.

It may be shown that for times $t \ll a^{2/3}$ the large scale fluctuations may be ignored, that is, the function $\beta(x/a)$ in (1.1) may be replaced by the constant drift $\beta(X(0)/a)$. This holds generally, without the assumptions of periodicity of b, β (Theorem 2.1). As a consequence, if b is periodic and β is arbitrary Lipschitz, then for times $1 \ll t \ll a^{2/3}$ the process $X(t)$ is asymptotically a Brownian motion (Theorem 2.2). This *first phase* analysis is carried out in Section 2.

If $b(\cdot)$, $\beta(\cdot)$ are both periodic with the same period lattice, say $\mathbb{Z}^k$, $\sigma(\cdot) = \sigma$ is a constant matrix, and a is a positive integer, then, for a fixed a, $\dot{X}(t) := X(t) \bmod a$ is a diffusion on the *big torus* $\mathscr{T}_a := \{x \bmod a \colon x \in \mathbb{R}^k\}$, and a central

Received August 1998; revised January 1999.
[1]Supported in part by NSF Grant DMS-95-04557.
AMS 1991 *subject classifications*. Primary 60F60, 60J05; secondary 60H10, 60J70.
Key words and phrases. Diffusion on a big torus, speed of convergence to equilibrium, initial and final Gaussian phases, growth in dispersion.

limit theorem holds for $X(t)$ as $t \to \infty$ [Bensoussan, Lions and Papanicolaou (1978), Chapter 3; Bhattacharya (1985)]. For large a, that is, as $a \to \infty$, how large must t be for this approximation to take hold? This Gaussian law is referred to as the *final phase* in this article. Under the *divergence–free* condition $\operatorname{div} b(\cdot) = 0 = \operatorname{div} \beta(\cdot)$ (*incompressibility*), the Gaussian approximation for a set of k_1 coordinates $X_j(t)$, $1 \le j \le k_1$, holds at times $t \gg a^2(\log a)^2$ provided an appropriate geometric condition holds on $\beta(\cdot)$ (Theorem 5.2). Under a different geometric condition the time scales for this final phase of Gaussian approximation are $t \gg a^4(\log a)^4$ in Theorem 5.3 and $t \ge a^2 \log a$ in Theorem 5.4.

Two crucial ingredients for this final phase analysis are (1) the speed at which $\dot{X}(t)$ approaches the uniform (equilibrium) distribution on $\mathscr{T}_a$ (as $a \to \infty$), and (2) the asymptotic relation between a and the dispersion matrix of the limiting Gaussian in the final phase. By spectral methods analogous to those of Diaconis and Stroock (1991) and Fill (1991), the L^1-distance between the distributions of $\dot{X}(t)$, with arbitrary $\dot{X}(0)$, and the equilibrium distribution is bounded above by $ca^{k/2} \exp\{-c' t/a^2\}$ for some positive constants c and c' (Theorem 4.5). For the analysis of final phase dispersion as a function of the scale parameter, it is convenient to look at the related process $Y(t) = X(a^2 t)/a$. Then $\dot{Y}(t) := Y(t) \bmod 1$ is a diffusion on the unit torus $\mathscr{T}_1$ with generator $A_a := \mathscr{D} + a(b(a\cdot) + \beta(\cdot)) \cdot \nabla$, with $\mathscr{D} = (1/2)\sum_{j,\,j'} D_{jj'} \partial^2/(\partial x_j \partial x_{j'})$ $(((D_{jj'})) := \sigma\sigma')$ and $\nabla = \text{grad}$. Since $b(a\cdot)$ is rapidly oscillating, one may approximate A_a by $\overline{A} := \mathscr{D} + a(\overline{b} + \beta(\cdot)) \cdot \nabla$, where $\overline{b} = (\overline{b}_1, \dots, \overline{b}_k)$ is the mean of $b(\cdot)$ w.r.t. the uniform distribution on $\mathscr{T}_1$. According to the central limit theorem for $X(t)$, with a fixed, the asymptotic dispersion (or variance) per unit time of $Y_j(t)$ is given by $D_{jj} - 2\|g_j\|_1^2$, where g_j is the mean-zero solution of $A_a g_j(x) = b_j(ax) + \beta_j(x) - \overline{b}_j - \overline{\beta}_j$. Here $\|g_j\|_1$ is the norm in the complex Hilbert space $H^1 = \{h$ mean-zero, periodic: $|h|^2$ and $|\nabla h|^2$ integrable w.r.t. uniform distribution on $\mathscr{T}_1\}$ endowed with the inner product $\langle g, f\rangle_1 = \int_{[0,1)^k} (\nabla g(x))' D \nabla f(x)^- \, dx$, f^- being the complex conjugate of f. One may replace g_j by the solution h_j to $\overline{A} h_j = \beta_j - \overline{\beta}_j$. The last equation may be expressed as $(\mathscr{I} + a\mathscr{D}^{-1}(\overline{b} + \beta(\cdot)) \cdot \nabla\} h_j = \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)$, or $(\mathscr{I} + a\overline{S}) h_j = \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)$, with $\mathscr{I}$ as the identity operator, and $\overline{S} = \mathscr{D}^{-1}(\overline{b} + \beta(\cdot)) \cdot \nabla$. Since $\overline{S}$ is a skew-symmetric compact operator on H^1 (Proposition 3.2), one may now use the spectral decomposition of $\overline{S}$ to express h_j in an eigenfunction expansion, arriving at $\|h_j\|_1^2 \simeq \|g_j\|_1^2$. This gives an asymptotic relation between a and the dispersion of $X_j(t)$ as that of a^2 times that of $Y_j(t)$. The dominant term in this expansion of the dispersion is $2a^2 \|(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\|_1^2$ where $f_{\underline{N}}$ is the projection of f in H^1 onto the null space $\underline{N}$ of $\overline{S}$. Thus, if $(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}} \neq 0$, the dispersion of $X_j(t)$ per unit time grows with a quadratically and is asymptotically bounded away from 0 and ∞ if $(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}} = 0$ (Theorems 3.7, 3.8).

The Gaussian approximations derived in this paper may be readily strengthened to their functional forms (see Remark 5.2.1). In other words, under appropriate scaling, the diffusion process $X(\cdot)$ in (1.1) has different Brownian motion approximations in the first and final phases.

Multiscale phenomena occur commonly in nature. The present study was motivated in part by the so-called *scale effect* in the dispersion of solute matter such as a chemical pollutant injected at a point in an underground water system, called an aquifer, saturated with water. It has been widely observed that for the solute concentration profile different Gaussian approximations with increasing dispersivity, or variance per unit time, hold at successively larger time scales [Fried and Combarnous (1971); Garabedian, LeBlanc, Gelhar and Colin (1991); Gelhar and Axness (1983); Guven and Molz (1986); LeBlanc, Garabedian, Hess, Gelhar, Quadri, Stollenwerk and Wood (1991); Sauty (1980); Sudicky (1986)]. The concentration $c(t, y)$ is governed at a local scale by a second-order linear parabolic (Fokker–Planck) equation with a drift term $v(\cdot)$ given by the velocity of water and diffusion coefficients which are of a somewhat larger order than the molecular diffusion coefficient of the solute. Since $v(\cdot)$ does not depend on time in a saturated aquifer under isothermal conditions, the root cause for the observed increase in dispersivity is the existence of multiscale heterogeneities in the medium [Bhattacharya and Gupta (1983) and Sposito, Jury and Gupta (1986)]. For the understanding of this it is enough to consider only two such scales of heterogeneity, reflected in the flow velocity as

$$v(y) = b(y) + \beta(y/a), \tag{1.2}$$

with a large. Here b and β are functions whose derivatives are of the same order, so that the derivatives of $\beta(\cdot/a)$ are small, namely, $O(1/a)$. Thus the fluctuations of $\beta(\cdot/a)$—the *large scale fluctuations*—are manifested only over large distances. Note that the solute concentration $c(t, y)$ corresponding to a unit local initial injection at x is simply the transition probability density $p(t; x, y)$ of a diffusion process $X(t)$ governed by the Itô equation (1.1).

It follows that the asymptotics of $t \to c(t, y)$ are given by the asymptotic distribution of $X(t)$. The proper way to look at this, when a is very large compared to the local scale, is to let $a \to \infty$ and let $t \to \infty$ at slower to higher rates relative to a. Initially, for a period of time $t \ll a^{2/3}$ (i.e., $t/a^{2/3} \to 0$), the fluctuations of $\beta(\cdot/a)$ may be ignored and $\beta(X(s)/a)$ may be replaced by its initial value $\beta(X(0)/a)$. Theorem 2.1 says that this new process, say $Y(\cdot)$, approximates the $X(\cdot)$ process up to such times t well in total variation distance. In particular, if $Y(\cdot)$ is asymptotically Gaussian, then so is $X(\cdot)$ for times $1 \ll t \ll a^{2/3}$. This holds, for example, if $b(\cdot)$ is periodic (Theorem 2.2).

The preceding analysis of dispersion of $X_j(t)$ as a function of the distance scale parameter a is formally the same as that for the dispersion of a diffusion $\hat{X}(t)$ with drift $a(\overline{b} + \beta(\cdot))$ [or $a\beta(\cdot)$, absorbing $\overline{b}$ in $\beta(\cdot)$] and diffusion matrix $D = \sigma\sigma'$. For the latter, one may regard $a = u_0$ as the *velocity parameter*. This enables one to study dispersion at a single scale as a function of u_0 (see Proposition 3.1). This latter analysis is also of importance in hydrology, and has been studied experimentally at the laboratory (or Darcy) scale extensively

[Fried and Combarnous (1971)]. This is discussed, along with the scale effect, in greater detail in Section 7.

Although the major emphasis in this article is on the case of divergence-free velocity fields, we also consider general one-dimensional multiscale diffusions with periodic coefficients. Here the speed of convergence to equilibrium may be either of the same order as in the divergence-free case ($k > 1$), or may be exponentially slow in a, requiring times $t \gg \exp\{ca\}$ to approach equilibrium (Theorems 4.6, 4.7, 4.9). The dispersion per unit time in the final phase is always asymptotically constant in a. In the time-reversible case this dispersion actually goes to zero exponentially fast with a. This study throws some light on the general *nondivergence-free case*.

It would be interesting and challenging to extend this study to the case of multiscale diffusions whose coefficients constitute an ergodic random field, or are almost periodic. For central limit theorems with such coefficients see Papanicolaou and Varadhan (1979), Kozlov (1979, 1980), Bhattacharya and Ramasubramanian (1988).

The present article provides a synthesis as well as an exposition of earlier work, often done in collaboration with Vijay Gupta, Homer Walker, and Friedrich Götze [Bhattacharya and Götze (1995); Bhattacharya and Gupta (1979, 1983); Bhattacharya, Gupta and Walker (1989)], although a number of results are either modified versions of earlier results or new. To facilitate exposition, detailed proofs are given for the most part. They also serve to remove some lacunae in Bhattacharya and Götze (1995).

A word on notation. The constants c, c' appearing in this article, with or without subscripts or superscripts, are all independent of the parameter a. The process $Y(\cdot)$ in Section 2 is different from the process $Y(\cdot)$ in Sections 3, 4, 5.

2. First phase of asymptotics. Consider a k-dimensional diffusion ($k \geq 1$) governed by the stochastic integral equation,

$$X(t) = X(0) + \int_0^t \{b(X(s)) + \beta(X(s)/a)\}\, ds + \int_0^t \sigma(X(s))\, dB(s), \quad t \geq 0. \tag{2.1}$$

Here $b(\cdot) \equiv (b_1(\cdot), \dots, b_k(\cdot))$, $\beta(\cdot) \equiv (\beta_1(\cdot), \dots, \beta_k(\cdot))$ are Lipschitzian functions on $\mathbb{R}^k$ to $\mathbb{R}^k$, $\sigma(\cdot)$ is a $(k \times k)$-matrix valued Lipschitzian function on $\mathbb{R}^k$, $B(\cdot)$ is a standard k-dimensional Brownian motion and the initial state $X(0)$ is independent of $B(\cdot)$. The *spatial scale parameter* a is assumed to be "large." One may think of $b(\cdot)$ as the drift velocity at the *local scale,* while $\beta(\cdot/a)$ is the *large scale* drift velocity. Since the vector field $x \to \beta(x/a)$ changes very slowly, the large scale fluctuations are manifested only at large distances and, therefore, not experienced by the process $X(\cdot)$ over an initial stretch of time. Over this time period one would then expect $X(\cdot)$ to behave like the process governed by the Itô equation

$$Y(t) = X(0) + \int_0^t \{b(Y(s)) + \beta(X(0)/a)\}\,ds + \int_0^t \sigma(Y(s))\,dB(s), \quad t \geq 0. \tag{2.2}$$

Note that the large-scale drift velocity $\beta(\cdot/a)$ in (2.1) is replaced by its initial value $\beta(X(0)/a)$ in (2.2). As an appropriate initial condition we will scale the initial state $X(0)$ as

$$X(0) = ax_0, \qquad x_0 \in \mathbb{R}^k. \tag{2.3}$$

This is merely to avoid the artificial importance of the origin that would arise from the assignment $X(0) = x_0$, since in the latter case $\beta(X(0)/a) \to \beta(0)$ as $a \to \infty$.

Our first result identifies the time period over which $Y(\cdot)$ is a good approximation to $X(\cdot)$. In order to state it, let $\mathscr{F}_t$ denote the Borel sigma-field of $\mathscr{C}[0, t]$—the set of all continuous functions on $[0, t]$ into $\mathbb{R}^k$, and let $P_{0,t}$ and $P_{1,t}$ denote the distributions of $Y_0^t := \{Y(s)\colon 0 \leq s \leq t\}$ and $X_0^t := \{X(s)\colon 0 \leq s \leq t\}$, respectively, on $\mathscr{F}_t$. The *total variation distance* between two measures μ and ν is denoted $\|\mu - \nu\|_{\mathrm{TV}}$.

THEOREM 2.1. *Assume that $b(\cdot)$ and its first-order derivatives are bounded, $\beta(\cdot)$ is bounded and has continuous and bounded derivatives of orders one and two, $\sigma(\cdot)$ is Lipschitzian, and the eigenvalues of $\sigma(\cdot)\sigma(\cdot)'$ are bounded away from zero and infinity.*

(a) *Then there exist constants c_i $(i = 1, 2, 3)$ which do not depend on "a" or t such that, uniformly for all x_0,*

$$\|P_{0,t} - P_{1,t}\|_{\mathrm{TV}} \leq c_1 \frac{t^{3/2}}{a} + c_2 \frac{t}{a} + c_3 \frac{t^{3/2}}{a^2}. \tag{2.4}$$

(b) *If $b_j(x) \equiv 0$ and $\beta_j(x_0) = 0$ for $2 \leq j \leq k$, and*

$$\frac{\partial \beta_j(x)}{\partial x_1} \equiv 0 \quad \text{for } 1 \leq j \leq k, \tag{2.5}$$

then one may take $c_1 = 0$ in (2.4).

(c) *If, in addition to the hypothesis in* (b), *one has*

$$\left(\frac{\partial \beta_j}{\partial x_i}\right)(x_0) = 0 \quad \text{for } 1 \leq j \leq k, 2 \leq i \leq k, \tag{2.6}$$

then one may take $c_1 = c_2 = 0$ in (2.4).

Before proving the theorem we make a few remarks on the time scales under (a)–(c) for the validity of the approximation of $X(\cdot)$ by $Y(\cdot)$, and on the physical significance of the conditions (2.5), (2.6).

REMARK 2.1.1. Condition (2.5) means that the large-scale velocity does not depend on the first coordinate x_1. This condition is satisfied by the so-called *stratified media* (see Sections 6 and 7). The condition $b_j(x) \equiv 0$ for $2 \le j \le k$ of course means that there is no small scale velocity in directions other than that in the x_1-direction. The conditions $\beta_j(x_0) = 0$ $(2 \le j \le k)$ and $(\partial\beta_j/\partial x_i)(x_0) = 0$ $(1 \le j \le k,\ 2 \le i \le k)$ are specific requirements on the initial point.

REMARK 2.1.2. It follows from (2.4) that

$$\|P_{0,t} - P_{1,t}\|_{TV} \to 0 \quad \text{as } \frac{t^{3/2}}{a} \to 0, \tag{2.7}$$

that is, $Y(\cdot)$ is a good approximation of $X(\cdot)$ for times

$$t \ll a^{2/3} \quad \text{or for } \frac{t}{a^{2/3}} \text{ small.} \tag{2.8}$$

Under the additional assumptions in part (b) of Theorem 2.1,

$$\|P_{0,t} - P_{1,t}\|_{\mathrm{TV}} \le c_2 \frac{t}{a} + c_3 \frac{t^{3/2}}{a^2} \to 0 \quad \text{as } \frac{t}{a} \to 0, \tag{2.9}$$

that is, $Y(\cdot)$ provides a good approximation to $X(\cdot)$ over a period of time

$$t \ll a. \tag{2.10}$$

Under the hypothesis of part (c),

$$\|P_{0,t} - P_{1,t}\|_{\mathrm{TV}} \le c_3 \frac{t^{3/2}}{a^2} \to 0 \quad \text{as } \frac{t}{a^{4/3}} \to 0, \tag{2.11}$$

that is, the initial phase of asymptotics governed by $Y(\cdot)$ holds over times satisfying

$$t \ll a^{4/3}. \tag{2.12}$$

Examples in Section 6 show that the estimates in Theorem 2.1 are, in general, optimal.

REMARK 2.1.3. The assumption that $\beta(\cdot)$ *is bounded* is only used in part (a) of Theorem 2.1. In the absence of this assumption, only the constants c_1', c_1'' in (2.4), (2.2) need to be changed to $c_1(1 + \|\beta(x_0)\|)$ and $c_1''(1 + \|\beta(x_0)\|)$, respectively.

PROOF OF THEOREM 2.1. By the Cameron–Martin–Girsanov theorem [see, e.g., Ikeda and Watanabe (1981), pages 176–181 or Friedman (1975), pages 164–169],

$$\|P_{0,t} - P_{1,t}\|_{\mathrm{TV}} = E\big|\exp\{Z(t)\} - 1\big|, \tag{2.13}$$

where, with $Y(0) = X(0)$,

$$Z(t) = \int_0^t \sigma^{-1}(Y(s))\{\beta(Y(s)/a) - \beta(Y(0)/a)\}\,dB(s) - \tfrac{1}{2}\int_0^t \left|\sigma^{-1}(Y(s))\{\beta(Y(s)/a) - \beta(Y(0)/a)\}\right|^2 ds. \tag{2.14}$$

By Itô's lemma [Ikeda and Watanabe (1981), pages 66 and 67, Bhattacharya and Waymire (1990), page 585],

$$\begin{aligned} &\beta_j(Y(s)/a) - \beta_j(Y(0)/a) \\ &\quad = \int_0^s \big(L_0\beta_j(\cdot/a)\big)(Y(s'))\,ds' \\ &\qquad + \int_0^s \big(\nabla(\beta_j(\cdot/a))\big)(Y(s'))\sigma(Y(s'))\,dB(s'), \qquad \nabla := \text{grad}, \end{aligned} \tag{2.15}$$

where, writing $D(x) := \sigma(x)\sigma'(x) = ((D_{ij}(x)))$,

$$L_0 = \frac{1}{2}\sum_{i,\,i'} D_{ii'}(\cdot)\frac{\partial^2}{\partial x_i \partial x_{i'}} + \big(b(\cdot) + \beta(x_0)\big)\cdot\nabla. \tag{2.16}$$

Thus,

$$\begin{aligned} \big(L_0(\beta_j(\cdot/a))\big)(Y(s')) = {}& \frac{1}{2a^2}\sum_{i,\,i'} D_{ii'}(Y(s'))\bigg(\frac{\partial^2\beta_j(\cdot)}{\partial x_i \partial x_{i'}}\bigg)(Y(s')/a) \\ & + \frac{1}{a}\big(b(Y(s')) + \beta(x_0)\big)\cdot\big(\nabla\beta_j(\cdot)\big)(Y(s')/a). \end{aligned} \tag{2.17}$$

Denoting the Reimann integral on the right side of (2.15) by $I_{1j}(s)$ and the stochastic integral by $I_{2j}(s)$, we have

$$E\big(\beta_j(Y(s)/a) - \beta_j(Y(0)/a)\big)^2 \le 2EI_{1j}^2(s) + 2EI_{2j}^2(s). \tag{2.18}$$

Letting λ denote the infimum (over $x \in \mathbb{R}^k$) of the smallest eigenvalue of $D(x)$, one has

$$\begin{aligned} E|Z(t)| \le {}& \frac{1}{\sqrt{\lambda}}\bigg(\int_0^t \sum_{j=1}^k E\big(\beta_j(Y(s)/a) - \beta_j(Y(0)/a)\big)^2 ds\bigg)^{1/2} \\ & + \frac{1}{2\lambda}\int_0^t \sum_{j=1}^k E\big(\beta_j(Y(s)/a) - \beta_j(Y(0)/a)\big)^2 ds. \end{aligned} \tag{2.19}$$

Also, denoting by $\|\cdot\|_\infty$ the supremum of the Euclidean norm of a real-, vector-, or matrix-valued function, one has

$$
\begin{aligned}
EI_{1j}^2(s) &\le s^2\left\{\frac{c_1'}{a^4}\,\|D(\cdot)\|_\infty^2\left\|\left(\max_{i,i'}\left|\frac{\partial^2\beta_j(\cdot)}{\partial x_i\partial x_{i'}}\right|\right)\right\|_\infty^2\right.\\
&\qquad\left.+\frac{2}{a^2}\left\|\left(b(\cdot)+\beta(x_0)\right)\cdot(\nabla\beta_j)(\cdot/a)\right\|_\infty^2\right\},\\
EI_{2j}^2(s) &\le \frac{s}{a^2}\,\|D(\cdot)\|_\infty\,\|\nabla\beta_j(\cdot)\|_\infty^2,
\end{aligned}
\tag{2.20}
$$

so that

$$
\begin{aligned}
&\int_0^t\sum_{j=1}^k\left(EI_{ij}^2(s)+EI_{2j}^2(s)\right)ds\\
&\quad\le\frac{t^3}{a^4}\,c_2'\|D(\cdot)\|_\infty^2\left[\sum_{j=1}^k\left\|\left(\max_{i,i'}\left|\frac{\partial^2\beta_j(\cdot)}{\partial x_i\partial x_{i'}}\right|\right)\right\|_\infty^2\right]\\
&\qquad+\frac{2t^3}{3a^2}\sum_{j=1}^k\left\|\left(b(\cdot)+\beta(x_0)\right)\cdot\left(\nabla\beta_j\right)(\cdot/a)\right\|_\infty^2\\
&\qquad+\frac{t^2}{2a^2}\|D(\cdot)\|_\infty\left(\sum_{j=1}^k\|\nabla\beta_j(\cdot)\|_\infty^2\right).
\end{aligned}
\tag{2.21}
$$

Using the last inequality in (2.19), we get

$$
\begin{aligned}
E|Z(t)| &\le\left(c_1''\frac{t^3}{a^2}+c_2''\frac{t^2}{a^2}+c_3''\frac{t^3}{a^4}\right)^{1/2}\frac{1}{\sqrt{\lambda}}\\
&\qquad+\left(c_1''\frac{t^3}{a^2}+c_2''\frac{t^2}{a^2}+c_3''\frac{t^3}{a^4}\right)\frac{1}{2\lambda}.
\end{aligned}
\tag{2.22}
$$

Here the constants c_i'' $(i=1,2,3)$ do not depend on a, t or x_0. Next note that $\exp\{Z(t)\}$, $t\ge 0$, is a martingale and, in particular, $E\exp\{Z(t)\}=1$, or

$$
\begin{aligned}
0=E\left(1-\exp\{Z(t)\}\right)&=E\left(1-\exp\{Z(t)\}\right)^+-E\left(1-\exp\{Z(t)\}\right)^-,\\
E\left|1-\exp\{Z(t)\}\right|&=2E\left(1-\exp\{Z(t)\}\right)^+\le 2\left[E(|Z(t)|\wedge 1)\right].
\end{aligned}
\tag{2.23}
$$

The last inequality follows from the relation $1-e^x\le|x|\wedge 1$ for $x\le 0$. The desired result (2.4) is now a consequence of (2.22) and (2.23).

To prove part (b), note that the second term on the right of (2.21) now vanishes. It remains to prove part (c). Under the additional assumption (2.6), one may express $Y_i(\cdot)$, $2\le i\le k$, as

$$
Y_i(t)=Y_i(0)+\int_0^t\sum_{r=1}^k\sigma_{ir}(Y(s))\,dB_r(s),\qquad 2\le i\le k,
\tag{2.24}
$$

so that the expected square of the stochastic integral in (2.15) may be estimated as

$$
\begin{aligned}
EI_{2j}^2(s) &= \frac{1}{a^2} E\left[\sum_{i=2}^{k} \int_0^s \left(\frac{\partial \beta_j}{\partial x_i}\right)\left(\frac{Y(s')}{a}\right) \sum_{r=1}^{k} \sigma_{ir}(Y(s'))\, dB_r(s')\right]^2 \\
&\le \frac{c_4''}{a^2} \int_0^s \sum_{i=2}^{k} E\left[\left(\frac{\partial \beta_j}{\partial x_i}\right)\left(\frac{Y(s')}{a}\right)\right]^2 ds'.
\end{aligned} \tag{2.25}
$$

In view of (2.6), $(\partial\beta_j/\partial x_i)(Y(0)/a) \equiv (\partial\beta_j/\partial x_i)(x_0) = 0$ for $i \ge 2$. Thus

$$
\begin{aligned}
&E\left[\left(\frac{\partial \beta_j}{\partial x_i}\right)\left(\frac{Y(s')}{a}\right)\right]^2 \\
&\quad = E\left[\left(\frac{\partial \beta_j}{\partial x_i}\right)\left(\frac{Y(s')}{a}\right) - \left(\frac{\partial \beta_j}{\partial x_i}\right)\left(\frac{Y(0)}{a}\right)\right]^2 \\
&\quad = E\left[\left(\frac{Y(s')}{a} - \frac{Y(0)}{a}\right) \cdot \left(\nabla \frac{\partial \beta_j}{\partial x_i}\right)(\tilde{Y}/a)\right]^2 \\
&\qquad\qquad \left[\tilde{Y} \text{ lying in the line segment joining } Y(0) \text{ and } Y(s')\right] \\
&\quad = E\left[\frac{1}{a} \sum_{i'=1}^{k} (Y_{i'}(s') - Y_{i'}(0)) \frac{\partial^2 \beta_j(\cdot)}{\partial x_{i'} \partial x_i} (\tilde{Y}/a)\right]^2 \\
&\quad = \frac{1}{a^2} E\left[\sum_{i'=2}^{k} (Y_{i'}(s') - Y_{i'}(0)) \left(\frac{\partial}{\partial x_i} \frac{\partial \beta_j(\cdot)}{\partial x_{i'}}\right)(\tilde{Y}/a)\right]^2 \\
&\quad \le \frac{c_5''}{a^2} s' \quad \text{by (2.24).}
\end{aligned} \tag{2.26}
$$

Use this and (2.25) to get

$$
\sum_{j=1}^{k} EI_{2j}^2(s) \le \frac{c_6''}{a^4} s^2. \tag{2.27}
$$

Using this estimate in place of the estimate of $EI_{2j}^2(s)$ in (2.20), the last term on the right side involving t^2/a^2 may be replaced by $c_7'' t^3/a^4$. Since the second term on the right of (2.21) (involving t^3/a^2) vanishes, as for part (b), the proof of part (c) is complete. □

REMARK 2.1.4. The significance of Theorem 2.1 is that it identifies the time scale for a change in the behavior of $X(\cdot)$, and shows that, prior to this threshold, $X(\cdot)$ and $Y(\cdot)$ are close in total variation distance. This is especially important in those cases in which $Y(\cdot)$ has interesting analyzable behavior. For example, if $b(\cdot) \equiv 0$ and $\sigma(\cdot)$ is a constant matrix, then $Y(\cdot)$ is a Brownian motion, so that $X(\cdot)$ is approximately a Brownian motion for times $t \ll a^{\frac{2}{3}}$. More important, Theorem 2.2 below identifies a class of coefficients $b(\cdot)$ such that $Y(\cdot)$ is asymptotically a Brownian motion and, for $1 \ll t \ll a^{2/3}$, so is

$X(\cdot)$. One may also consider a class of (nonperiodic) coefficients $b(\cdot)$, $\beta(x_0)$, such that $Y(\cdot)$ is ergodic, that is, $Y(\cdot)$ has a unique invariant probability and is Harris recurrent.

For Theorem 2.2 below, assume $b(\cdot)$, $\sigma(\cdot)$ are periodic having the same period lattice. Since by an appropriate nonsingular linear transformation of $X(\cdot)$, the period lattice of the transformed coefficients becomes the standard lattice $\mathbb{Z}^k$, we will assume without loss of generality that $b(\cdot)$, $\sigma(\cdot)$ are periodic with period one in each coordinate, that is,

$$b(x+r) = b(x), \qquad \sigma(x+r) = \sigma(x) \qquad \forall\, x \in \mathbb{R}^k,\, r \in \mathbb{Z}^k. \tag{2.28}$$

In this case the process $\dot{Y}(\cdot)$ defined by

$$\dot{Y}(t) := Y(t) \bmod 1 \equiv \big(Y_1(t) \bmod 1, \dots, Y_k(t) \bmod 1\big), \tag{2.29}$$

is a Markov process, a diffusion on the *unit torus*

$$\mathscr{T}_1 := \big\{x \bmod 1\colon x \in \mathbb{R}^k\big\} \equiv \big\{(x_1 \bmod 1, \dots, x_k \bmod 1)\colon x = (x_1, \dots, x_k) \in \mathbb{R}^k\big\}$$

[see, e.g., Bhattacharya and Waymire (1990), page 518]. Given that the transition probability density of $Y(\cdot)$ [and, therefore, of $\dot{Y}(\cdot)$] is positive, it is simple to check that $\dot{Y}(\cdot)$ has a unique invariant probability $\pi(x)\,dx$ and that $\dot{Y}(\cdot)$ has an exponentially decaying phi-mixing rate. Also, as shown in Bensoussan, Lions and Papanicolaou [(1978), Chapter 3] and Bhattacharya (1985), $Y(\cdot)$ is asymptotically a Brownian motion, in the sense that the sequence of processes

$$\frac{Y(nt) - Y(0) - nt(\overline{b} + \beta(x_0))}{\sqrt{n}}, \qquad 0 \le t \le 1 \tag{2.30}$$

converges in distribution, as $n \to \infty$, to a Brownian motion with zero drift and dispersion matrix K. Here

$$\begin{aligned} \overline{b} &= (\overline{b}_1, \overline{b}_2, \dots, \overline{b}_k),\\ \overline{b}_j &:= \int_{\mathscr{T}_1} b_j(x)\pi(x)\,dx, \qquad 1 \le j \le k,\\ K &= \int_{\mathscr{T}_1} (\operatorname{grad}\psi(x) - I_k)D(x)(\operatorname{grad}\psi(x) - I_k)'\pi(x)\,dx, \end{aligned} \tag{2.31}$$

I_k being the $k \times k$ identity matrix and $\psi = (\psi_1, \psi_2, \dots, \psi_k)'$ being the unique mean-zero periodic solution of

$$L_0\psi_j(x) = b_j(x) - \overline{b}_j, \qquad 1 \le j \le k. \tag{2.32}$$

Recall that L_0 is the generator of $Y(\cdot)$ [see (2.16)], and therefore of $\dot{Y}(\cdot)$ when restricted to periodic functions. The existence and uniqueness of the solution of (2.32) follows from a general theorem for ergodic Markov processes [see Bhattacharya (1982)]. Indeed, the solution is given by (2.37) below. A proof of the convergence in distribution of (2.30) is sketched in the course of the proof of the theorem below. We will occasionally write $\to_{\mathscr{L}}$ to denote *convergence in law,* or in distribution.

The normal distribution on $\mathbb{R}^k$ having mean vector zero and dispersion matrix K will be denoted by Φ_K or $\mathscr{N}(0, K)$.

THEOREM 2.2. *Assume that* $b(\cdot)$ *is continuously differentiable, and* $\sigma(\cdot)$ *is Lipschitzian, and* $b(\cdot)$ *and* $\sigma(\cdot)$ *are periodic as shown in* (2.28). *Assume also that* $\beta(\cdot)$ *has continuous and bounded derivatives of orders one and two. Let the diffusion* $X(\cdot)$ *be as defined in* (2.1) *with initial value* (2.3). *Then, as* $n \to \infty$, $a \to \infty$, *such that*

$$\frac{n}{a^{2/3}} \to 0, \tag{2.33}$$

the process

$$\frac{X(nt) - X(0) - nt(\overline{b} + \beta(x_0))}{\sqrt{n}}, \qquad 0 \le t \le 1, \tag{2.34}$$

converges in distribution to a Brownian motion with zero drift and dispersion matrix K. *In particular*,

$$\frac{X(t) - X(0) - t(\overline{b} + \beta(x_0))}{\sqrt{t}} \to_{\mathscr{L}} \Phi_K \tag{2.35}$$

as $t \to \infty$, $a \to \infty$ *such that*

$$\frac{t}{a^{2/3}} \to 0. \tag{2.36}$$

PROOF. In view of Theorem 2.1(a) (also see Remark 2.1.3), it is enough to prove that the process (2.30) converges to a Brownian motion, as $n \to \infty$. Although the latter is proved in Bensoussan, Lions and Papanicolaou [(1978), Chapter 3] and Bhattacharya (1985), we will sketch the arguments here for completeness and for later use. First note that

$$\psi_j(\cdot) := -\int_0^\infty T_t(b_j(\cdot) - \overline{b}_j)\, dt \tag{2.37}$$

is well defined as an element of $\mathscr{L}^2(\mathscr{T}_1, \pi)$, where T_t is the transition operator

$$(T_t f)(y) := E\big[f(\dot{Y}(t)) \mid \dot{Y}(0) = y\big], \qquad f \in \mathscr{L}^2(\mathscr{T}_1, \pi). \tag{2.38}$$

Note that $T_t f \to \overline{f}$, exponentially fast as $t \to \infty$, in the $\mathscr{L}^2$-norm. By applying T_h to both sides of (2.37), one obtains $(T_h\psi_j - \psi_j)/h \to b_j - \overline{b}_j$ in $\mathscr{L}^2(\mathscr{T}_1, \pi)$. In other words, (2.32) holds. By Itô's lemma,

$$\begin{aligned}
&\psi_j(\dot{Y}(t)) - \psi_j(\dot{Y}(0))\\
&\quad = \int_0^t L_0\psi_j(\dot{Y}(s))\, ds + \int_0^t \operatorname{grad}\psi_j(\dot{Y}(s)) \cdot \sigma(\dot{Y}(s))\, dB(s)\\
&\quad = \int_0^t \big(b_j(\dot{Y}(s)) - \overline{b}_j\big)\, ds + \int_0^t \operatorname{grad}\psi_j(\dot{Y}(s)) \cdot \sigma(\dot{Y}(s))\, dB(s).
\end{aligned} \tag{2.39}$$

Hence,

$$
\begin{aligned}
& Y(t) - Y(0) - t(\overline{b} + \beta(x_0)) \\
(2.40) \qquad & \equiv \int_0^t \big(b_j(\dot{Y}(s)) - \overline{b}\big)\, ds + \int_0^t \sigma(\dot{Y}(s))\, dB(s) \\
& = \psi(\dot{Y}(t)) - \psi(\dot{Y}(0)) - \int_0^t \big(\operatorname{grad} \psi(\dot{Y}(s)) - I_k\big)\sigma(\dot{Y}(s))\, dB(s).
\end{aligned}
$$

On dividing both sides of (2.40) by $\sqrt{t}$ and letting $t \to \infty$, one shows that the asymptotic distribution of $t^{-1/2}(Y(t) - Y(0) - t\overline{b} - t\beta(x_0))$ is the same as that of

$$
(2.41) \qquad -\frac{1}{\sqrt{t}} \int_0^t \big(\operatorname{grad} \psi(\dot{Y}(s)) - I_k\big)\sigma(\dot{Y}(s))\, dB(s).
$$

But the integrand in (2.41) is stationary and ergodic. Thus if $\dot{Y}(0)$ has distribution π then, by the Billingsley–Ibragimov central limit theorem for martingales [Billingsley (1968), page 206], (2.41) converges in law to $\Phi_K \equiv \mathscr{N}(0, K)$. Since the transition probability density $\dot{p}(t; z, y)$ converges to $\pi(y)$ exponentially fast as $t \to \infty$, uniformly in z and y, the limit law of (2.41) under $\dot{Y}(0) = x_0$ is the same as under the initial distribution π. $\square$

REMARK 2.2.1. Under the hypothesis of Theorem 2.2, a Berry–Esséen type bound may be derived for the process $Y(t)$ defined by (2.2), namely,

$$
(2.42) \qquad \sup_{C \in \mathscr{C}} \left| P\left(\frac{Y(t) - Y(0) - t(\overline{b} + \beta(x_0))}{\sqrt{t}} \in C \right) - \Phi_K(C) \right| \le \frac{c_4}{\sqrt{t}},
$$

where $\mathscr{C}$ is the class of all Borel measurable convex sets in $\mathbb{R}^k$ and c_4 is a positive constant which depends only on $b(\cdot)$, $\beta(\cdot)$ and $D(\cdot)$; in particular, c_4 is independent of a [see, e.g., Nagaev (1961) or Tikhomirov (1980)]. Combining (2.4) and (2.42), we get the following refinement of (2.35):

$$
\begin{aligned}
& \sup_{C \in \mathscr{C}} \left| P\left(\frac{X(t) - X(0) - t(\overline{b} + \beta(x_0))}{\sqrt{t}} \in C \right) - \Phi_K(C) \right| \\
(2.43) \qquad & \quad \le c_1 \frac{t^{3/2}}{a} + c_2 \frac{t}{a} + c_3 \frac{t^{3/2}}{a^2} + \frac{c_4}{\sqrt{t}}.
\end{aligned}
$$

This goes to zero as $t \to \infty$ and $t/a^{2/3} \to 0$. Indeed, one may bound the right side by $c_5 \frac{t^{3/2}}{a} + \frac{c_4}{\sqrt{t}}$, if $t/a^{2/3} < 1$, $a > 1$. Assuming that this is the precise order of the error of normal approximation, the approximation by Φ_K improves as t ($\gg 1$) increases to an order such that $\frac{t^{3/2}}{a} = O(\frac{1}{\sqrt{t}})$, that is, $t = 0(a^{1/2})$ (a large), the minimum error being $O(a^{-1/4})$. After this time, this normal approximation worsens, and it breaks down for t of order $a^{2/3}$ or larger. Under the special assumptions in part (b) of Theorem 2.1, in addition to the assumptions of Theorem 2.2, one may use (2.9), instead of (2.4), to take $c_1 = 0$ in (2.43), so that the error may be bounded by $c_5(t/a) + (c_4/\sqrt{t})$, which has its smallest value $O(a^{-1/3})$ at a time $t = 0(a^{2/3})$. If, in addition, the assumption in part (c) of Theorem 2.1 holds, then one may take $c_1 = 0 = c_2$ in (2.43) to get an error bound $c_3(t^{2/3}/a^2) + c_4/\sqrt{t}$, which becomes minimum for $t = O(a)$, the minimum error being $O(a^{-1/2})$.

REMARK 2.2.2. Central limit theorems for a process such as $Y(\cdot)$ in (2.2) have been studied in the literature under assumptions other than periodicity of $b(\cdot)$, $\sigma(\cdot)$. For example, one may take $b(\cdot)$, $\sigma(\cdot)$ to be (i) almost periodic, or (ii) stationary ergodic random fields [see Papanicolaou and Varadhan (1979), Kozlov (1979, 1980) and Bhattacharya and Ramasubramanian (1988)]. If a is sufficiently large, so that these Gaussian approximations for $Y(t)$ hold for $1 \ll t \ll a^{2/3}$, then they hold for $X(t)$ over the same time scale.

3. Analysis of dispersion in the final phase: the divergence-free case. In this section we first analyze the functional dependence of the asymptotic dispersion of a diffusion with periodic coefficients on a large *velocity parameter* u_0. This is of importance in itself, and has been studied extensively in the hydrology literature [see, e.g., Fried and Combarnous (1971)]. More important for us is the fact (see Proposition 3.1 below) that the asymptotic dispersion matrix is the same function of the spatial parameter "a" in the absence of u_0, as it is of u_0 in the absence of "a." We will use this fact later in the section to analyse the dispersion in the final phase. Consider then the k-dimensional diffusion $\hat{X}(t)$ governed by the Itô equation,

$$\hat{X}(t) = \hat{X}(0) + u_0 \int_0^t \beta(\hat{X}(s))\,ds + \int_0^t \sigma(\hat{X}(s))\,dB(s), \tag{3.1}$$

where $\beta(\cdot)$ is continuously differentiable and periodic with period lattice $\mathbb{Z}^k$, $\sigma(\cdot)$ is a Lipschitzian matrix-valued periodic function of period one whose eigenvalues are bounded away from zero, $\hat{X}(0)$ is independent of the k-dimensional standard Brownian motion $B(\cdot)$, and u_0 is a "large" parameter scaling the velocity magnitude. We have seen in Section 2 that $(\hat{X}(t) - \hat{X}(0) - t\overline{\beta})/\sqrt{t}$ converges in distribution to a Gaussian $\mathcal{N}(0, K)$ with mean zero and dispersion matrix $K = K(u_0)$, say. On the other hand, one may consider the diffusion $\tilde{X}(\cdot)$ governed by

$$\tilde{X}(t) = \tilde{X}(0) + \int_0^t \beta(\tilde{X}(s)/a)\,ds + \int_0^t \sigma(\tilde{X}(s/a)d\tilde{B}(s), \tag{3.2}$$

with the same assumptions on $\beta(\cdot)$, $\sigma(\cdot)$ as above, $\tilde{B}(\cdot)$ a standard Brownian motion independent of $\tilde{X}(0)$, and a a "large" parameter scaling distance. Let $\tilde{K}(a)$ denote its asymptotic dispersion matrix as computed in Section 2.

PROPOSITION 3.1. *Under the above assumptions*, $K(\cdot) \equiv \tilde{K}(\cdot)$.

PROOF. Define the process

$$\hat{Y}(t) := u_0 \hat{X}(t/u_0^2), \qquad t \geq 0. \tag{3.3}$$

Then

$$d\hat{Y}(t) = \beta(\hat{Y}(t)/u_0)\,dt + \sigma(\hat{Y}(t)/u_0)\,d\overline{B}(t), \tag{3.4}$$

where $\overline{B}(t) := u_0 B(t/u_0^2)$ is again a k-dimensional standard Brownian motion. Thus, with $a = u_0$, $\hat{Y}(\cdot)$ and $\tilde{X}(\cdot)$ have the same law, if their initial states are the same. However, irrespective of initial states, the scaled processes converge weakly to the same Gaussian law. Finally, the asymptotic dispersion matrix of $\hat{Y}(\cdot)$ is the same as that of $\hat{X}(\cdot)$. For $\lim_{t\to\infty} \operatorname{var} \hat{Y}(t)/t = \lim_{t\to\infty} \operatorname{var} \hat{X}(t/u_0^2)/(t/u_0^2) = \lim_{t\to\infty} \operatorname{var} \hat{X}(t)/t = K(u_0)$. Therefore, $K(\cdot) \equiv \tilde{K}(\cdot)$. □

The analysis of the asymptotic dispersion of $\hat{X}(\cdot)$ [governed by (3.1)] will be carried out under the additional assumptions,

$$\operatorname{div}\beta(x) \equiv \sum_{j=1}^{k} \frac{\partial \beta_j(x)}{\partial x_j} = 0 \qquad \forall\, x \tag{3.5}$$

and

$$\sigma(x) \equiv \sigma, \tag{3.6}$$

where σ is a constant nonsingular $k \times k$ matrix. An extension to nonconstant σ is indicated later (see Remark 4.5.2). We will write $D = ((D_{jj'}))$ for $\sigma\sigma'$. The *divergence-free* condition (3.5) means that the medium through which the transport (of a solute, e.g.) is taking place is *incompressible*. The spectral method of this section does not extend to velocity fields which are not divergence free. The latter are treated in Section 5 by direct calculations for the case of dimension one.

Under the assumptions that $\beta(\cdot)$ is periodic with period lattice $\mathbb{Z}^k$, and (3.5), (3.6) hold, the diffusion $\dot{\hat{X}}(t) := \hat{X}(t) \bmod 1$ on the torus $\mathscr{T}_1 := \{x \bmod 1 \colon x \in \mathbb{R}^k\}$ has the normalized Lebesgue measure dx as the invariant distribution. To see this check that $L^*1 = 0$, where L^* is the *formal adjoint* of the generator L of $\dot{\hat{X}}$,

$$\begin{aligned}
Lf(x) &= \frac{1}{2}\sum_{j,\,j'=1}^{k} D_{jj'} \frac{\partial^2 f(x)}{\partial x_j \partial x_{j'}} + u_0 \sum_{j=1}^{k} \beta_j(x) \frac{\partial f}{\partial x_j},\\
L^* f &= \frac{1}{2}\sum_{j,\,j'=1}^{k} \frac{\partial^2}{\partial x_j \partial x_{j'}} (D_{jj'} f(x)) - u_0 \sum_{j=1}^{k} \frac{\partial}{\partial x_j} (\beta_j(x) f(x))\\
&= \frac{1}{2}\sum_{j,\,j'=1}^{k} D_{jj'} \frac{\partial^2 f(x)}{\partial x_j \partial x_{j'}} - u_0 \sum_{j=1}^{k} \beta_j(x) \frac{\partial f(x)}{\partial x_j}.
\end{aligned} \tag{3.7}$$

The last equality follows from (3.5).

It follows from Section 2 [see (2.30)–(2.32)] that

$$\frac{\hat{X}(t) - \hat{X}(0) - u_0 t\overline{\beta}}{\sqrt{t}} \to_{\mathscr{L}} \Phi_K \quad \text{as } t \to \infty, \tag{3.8}$$

where $\overline{\beta} = (\overline{\beta}_1, \dots, \overline{\beta}_k)$, with $\overline{\beta}_j$ given by

$$\overline{\beta}_j = \int_{[0,1]^k} \beta_j(x)\, dx \tag{3.9}$$

and

$$\begin{aligned} K &= \int_{[0,1]^k} \big(\operatorname{grad}\psi(x) - I_k\big) D \big(\operatorname{grad}\psi(x) - I_k\big)'\, dx \\ &= ((K_{jj'})). \end{aligned} \tag{3.10}$$

Here $\psi(\cdot) = (\psi_1(\cdot), \dots, \psi_k(\cdot))$ is the unique mean-zero solution in the domain of L [in $\mathscr{L}^2(\mathscr{T}_1, dx)$],

$$L\psi_j(\cdot) = u_0\big(\beta_j(\cdot) - \overline{\beta}_j\big), \qquad 1 \le j \le k. \tag{3.11}$$

Further, one has

$$\begin{aligned} K_{jj'} &= E_{jj'} + D_{jj'}, \\ E_{jj'} &:= \int_{[0,1]^k} \operatorname{grad}\psi_j(x) \cdot D \operatorname{grad}\psi_{j'}(x)\, dx. \end{aligned} \tag{3.12}$$

This follows from (3.10) using periodic boundary conditions, namely,

$$\int_{[0,1]^k} \frac{\partial \psi_j(x)}{\partial x_r}\, dx = 0, \qquad 1 \le j, r \le k. \tag{3.13}$$

To analyze $E_{jj'}$ let us introduce the complex Hilbert space,

$$H^0 = \mathscr{L}^2(\mathscr{T}_1, dx) \cap 1^\perp = 1^\perp, \tag{3.14}$$

where dx denotes Lebesgue measure, or the uniform distribution on the unit torus $\mathscr{T}_1$, and $\mathscr{L}^2(\mathscr{T}_1, dx) \equiv \mathscr{L}^2$ is the space of complex-valued square integrable (w.r.t. dx) functions on $\mathscr{T}_1$. Here $1^\perp$ is the subspace of $\mathscr{L}^2$ orthogonal to constants, that is, the set of all mean-zero elements of $\mathscr{L}^2$. This identifies H^0 and its inner product as

$$\begin{aligned} H^0 &= \left\{ h \text{ periodic: } \int_{[0,1]^k} |h(x)|^2\, dx < \infty, \int_{[0,1]^k} h(x)\, dx = 0 \right\}, \\ \langle f, g\rangle_0 &:= \int_{[0,1]^k} f(x) g^-(x)\, dx. \end{aligned} \tag{3.15}$$

Here g^- is the *complex conjugate* of g. The spectral expansion of $E_{jj'}$ is carried out on the Hilbert space H^1 defined by

$$(3.16)\qquad \begin{aligned} H^1 &= \left\{h \in H^0\colon \int_{[0,1]^k} |\nabla h|^2(x)\,dx < \infty\right\}, \qquad \nabla := \text{grad}, \\ \langle f, g\rangle_1 &:= \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'} \int_{[0,1]^k} \frac{\partial f}{\partial x_j}\frac{\partial g^-}{\partial x_{j'}}\,dx. \end{aligned}$$

Note that for all twice continuously differentiable periodic f and once continuously differentiable periodic g, integration by parts yields

$$(3.17)\qquad \langle f, g\rangle_1 = -\langle \mathscr{D} f, g\rangle_0,$$

where

$$(3.18)\qquad \mathscr{D} := \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'} \frac{\partial^2}{\partial x_j \partial x_{j'}}.$$

The *Sobolev space* H^1, given in (3.16), is the closure in the norm $\|f\|_1 := (\langle f, f\rangle_1)^{1/2}$ of the space of all twice continuously differentiable periodic functions in H^0, and the elements of H^1 are those elements of H^0 which have square integrable derivatives on $[0,1]^k$. Finally let H^2 be the subspace of H^1 having square integrable derivatives of order two. The operator $\mathscr{D}$ maps H^2 onto H^0 and is in fact invertible. Indeed, if $g \in H^2$ is such that $\mathscr{D} g = f$, then the *Fourier transforms* $\hat{g}$, $\hat{f}$ of g and f are related, on integration by parts, by

$$(3.19)\qquad \begin{aligned} \hat{f}(r) &= \int_{[0,1]^k} f(x)\exp(-2\pi i r\cdot x)\,dx \\ &= \int_{[0,1]^k} (\mathscr{D} g)(x)\exp(-2\pi i r\cdot x)\,dx \\ &= \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'} \int_{[0,1]^k} \frac{\partial^2 g(x)}{\partial x_j \partial x_{j'}}\exp(-2\pi i r\cdot x)\,dx \\ &= \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'}(2\pi i r_j)(2\pi i r_{j'}) \int_{[0,1]^k} g(x)\exp(-2\pi i r\cdot x)\,dx \\ &= -2\pi^2\left(\sum_{j,j'=1}^{k} D_{jj'} r_j r_{j'}\right)\hat{g}(r), \qquad r \in \mathbb{Z}^k. \end{aligned}$$

Thus $\hat{g}$ is given by

$$(3.20)\qquad \hat{g}(0) = 0, \qquad \hat{g}(r) = -\frac{1}{2\pi^2}\left(\sum_{j,j'=1}^{k} D_{jj'} r_j r_{j'}\right)^{-1}\hat{f}(r), \qquad r \in \mathbb{Z}^k\backslash\{0\}.$$

Note that

$$(3.21)\qquad |\hat{g}(r)| \geq \frac{1}{2\alpha_2\pi^2|r|^2}|\hat{f}(r)|, \qquad r \in \mathbb{Z}^k\backslash\{0\},$$

where α_1 is the smallest eigenvalue of D and α_2 is the largest. We will show that the operator $\mathscr{D}^{-1}$: $H^0 \to H^1$ is *compact*. For this, let f_k $(k = 1, 2, \ldots)$ be a bounded sequence in H^0, say $\|f_k\|_0 \le 1 \,\forall\, k$. Then there exists a subsequence $f_{k'}$ $(k = 1, 2, \ldots)$ which converges weakly to some element f_0 of the unit ball of H^0. In particular,

$$\begin{aligned} &\hat{f}_{k'}(r) \to \hat{f}_0(r) \quad \text{as } k' \to \infty \quad (r \in \mathbb{Z}^k \backslash \{0\}),\\ &\hat{f}_{k'}(0) = 0 = \hat{f}_0(0) \qquad \forall\, k'. \end{aligned} \tag{3.22}$$

Let

$$g_k := \mathscr{D}^{-1} f_k, \qquad g_0 := \mathscr{D}^{-1} f_0. \tag{3.23}$$

We now show that $\|g_{k'} - g_0\|_1 \to 0$ as $k' \to \infty$. For this write [see (3.17)–(3.21)]

$$\begin{aligned} \|g_{k'} - g_0\|_1^2 &= -\langle \mathscr{D}(g_{k'} - g_0), g'_k - g_0\rangle_0 = -\langle f_{k'} - f_0, \mathscr{D}^{-1}(f_{k'} - f_0)\rangle_0\\ &= \frac{1}{2\pi^2} \sum_{r \ne 0} \big|\hat{f}_{k'}(r) - \hat{f}_0(r)\big|^2 \frac{1}{\Sigma D_{jj'} r_j r_{j'}}\\ &\le \frac{1}{2\alpha_1 \pi^2} \sum_{r \ne 0} \big|\hat{f}_{k'}(r) - \hat{f}_0(r)\big|^2 \frac{1}{|r|^2}\\ &\le \frac{1}{2\alpha_1 \pi^2} \Bigg\{ \sum_{|r| \le R} \big|\hat{f}_{k'}(r) - \hat{f}_0(r)\big|^2 + \frac{1}{R^2} \sum_{|r| > R} \big|\hat{f}_{k'}(r) - \hat{f}_0(r)\big|^2 \Bigg\}\\ &\le \frac{1}{2\alpha_1 \pi^2} \Bigg\{ \sum_{|r| \le R} \big|\hat{f}_{k'}(r) - \hat{f}_0(r)\big|^2 + \frac{4}{R^2} \Bigg\}, \end{aligned} \tag{3.24}$$

since $\|f_{k'}\|_0 \le 1$, $\|f_0\|_0 \le 1$. Given $\varepsilon > 0$, choose R_ε such that $(1/2\alpha_1\pi^2)((4/R^2)) < \varepsilon/2$ for $R \ge R_\varepsilon$. Now choose k'_ε large so that $\sum_{|r| \le R_\varepsilon} |\hat{f}_{k'}(r) - \hat{f}_0(r)|^2 < \varepsilon/2 \;\forall\, k' \ge k'_\varepsilon$. Then

$$\|g_{k'} - g_0\|_1^2 < \varepsilon \qquad \forall\; k' \ge k'_\varepsilon.$$

Thus $g_{k'} \to g_0$ in H^1, proving the compactness of $\mathscr{D}^{-1}$.

We may now express (3.11) as

$$\big(\mathscr{D} + u_0 \beta(\cdot) \cdot \nabla\big)\psi_j(\cdot) = u_0\big(\beta_j(\cdot) - \overline{\beta}_j\big), \qquad 1 \le j \le k. \tag{3.25}$$

Rewrite (3.25) as

$$(\mathscr{I} + u_0 S)\psi_j = u_0 \mathscr{D}^{-1}(\beta_j(\cdot) - \overline{\beta}_j), \qquad 1 \le j \le k, \tag{3.26}$$

where $\mathscr{I}$ is the *identity operator on* H^1, and S is the linear operator

$$S = \mathscr{D}^{-1}\beta(\cdot) \cdot \nabla \tag{3.27}$$

acting on H^1. Note that, for $f, g \in H^1$,

$$
\begin{aligned}
\langle \mathscr{D}^{-1}\beta(\cdot)\cdot\nabla f, g\rangle_1 &= -\langle \beta(\cdot)\cdot\nabla f, g\rangle_0 \\
&= -\sum_{j=1}^{k}\int_{[0,1]^k}\beta_j(x)\frac{\partial f(x)}{\partial x_j}\,g^-(x)\,dx \\
&= \int_{[0,1]^k} f(x)\sum_{j=1}^{k}\frac{\partial}{\partial x_j}\big(\beta_j(x)g^-(x)\big)\,dx \\
&= \int_{[0,1]^k} f(x)\left\{g^-(x)(\operatorname{div}\beta)(x)+\sum_{j=1}^{k}\beta_j(x)\frac{\partial g^-(x)}{\partial x_j}\right\}dx \\
&= \int_{[0,1]^k} f(x)(\beta(\cdot)\cdot\nabla g)(x)\,dx = \langle f, \beta(\cdot)\cdot\nabla g\rangle_0 \\
&= -\langle \mathscr{D}^{-1}f, \beta(\cdot)\cdot\nabla g\rangle_1.
\end{aligned}
\tag{3.28}
$$

Thus S is skew symmetric. Noting that $\beta(\cdot)\cdot\nabla\colon H^1 \to H^0$ is bounded, while $\mathscr{D}^{-1}\colon H^0 \to H^1$ is compact, we have the following result [see Reed and Simon (1980), page 200].

PROPOSITION 3.2. *Let $\beta(\cdot)$ be continuously differentiable and periodic and* (3.5), (3.6) *hold. Then $S := \mathscr{D}^{-1}\beta(\cdot)\cdot\nabla$ is a skew symmetric compact operator on H^1 and, therefore, may be expressed as $S = iG$ where G is a compact and self–adjoint operator on H^1.*

Applying the spectral theorem for compact self-adjoint operators [see Reed and Simon (1980), page 203], it now follows that G has a sequence of nonzero eigenvalues $\lambda_n \to 0$ with corresponding eigenfunctions φ_n $(n \ge 1)$ such that $\{\varphi_n\colon n \ge 1\}$ form a complete orthonormal sequence for $N^\perp$, the subspace of H^1 orthogonal to the *null space* N of G or S. Hence one has the eigenfunction expansion,

$$
f = f_N + \sum_{n=1}^{\infty}\langle f, \varphi_n\rangle_1\varphi_n, \qquad f \in H^1, \tag{3.29}
$$

where f_N is the *orthogonal projection* of f onto N. Also,

$$
Sf = \sum_{n=1}^{\infty} i\lambda_n\langle f, \varphi_n\rangle_1\varphi_n, \qquad f \in H^1. \tag{3.30}
$$

Taking $f = \psi_j$ in (3.29), (3.30), one may now express the equation (3.26) in spectral form

$$
\begin{aligned}
(\psi_j)_N &= u_0\big(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\big)_N, \\
(1 + iu_0\lambda_n)\langle\psi_j, \varphi_n\rangle_1 &= u_0\langle\mathscr{D}^{-1}\big(\beta_j - \overline{\beta}_j\big), \varphi_n\rangle_1, \qquad n \ge 1.
\end{aligned}
\tag{3.31}
$$

Hence,

$$
\langle\psi_j, \varphi_n\rangle_1 = \frac{u_0\beta_{jn}}{1 + iu_0\lambda_n}, \qquad \beta_{jn} := \langle\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j), \varphi_n\rangle_1, \qquad n \ge 1. \tag{3.32}
$$

Thus the components $E_{jj'}$ of the dispersion matrix $((K_{jj'}))$ arising from the heterogeneity of the medium of transport may be expressed as [see (3.12), (3.16)]

$$
\begin{aligned}
E_{jj'} &= 2\langle \psi_j, \psi_{j'} \rangle_1 \\
&= 2\big\langle (\psi_j)_N, (\psi_{j'})_N \big\rangle_1 + 2\sum_{n=1}^{\infty} \langle \psi_j, \varphi_n \rangle_1 \overline{\langle \psi_{j'}, \varphi_n \rangle_1} \qquad (3.33) \\
&= 2u_0^2 \big\langle (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_N, (\mathscr{D}^{-1}(\beta_{j'} - \overline{\beta}_{j'}))_N \big\rangle_1 + 2\sum_{n=1}^{\infty} \frac{u_0^2 \beta_{jn} \overline{\beta_{j'n}}}{1 + u_0^2 \lambda_n^2}.
\end{aligned}
$$

In particular,

$$
E_{jj} = 2u_0^2 \left\{ \big\| (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_N \big\|_1^2 + \sum_{n=1}^{\infty} \frac{|\beta_{jn}|^2}{1 + u_0^2 \lambda_n^2} \right\}, \qquad 1 \le j \le k. \tag{3.34}
$$

THEOREM 3.3. *Suppose the assumptions in Proposition* 3.2 *hold.*

(a) *If* $(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_N \neq 0$, *then*

$$
\lim_{u_0 \to \infty} \frac{K_{jj}}{u_0^2} = \lim_{u_0 \to \infty} \frac{E_{jj}}{u_0^2} = 2\big\| (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_N \big\|_1^2 > 0. \tag{3.35}
$$

(b) *If* $\beta_j - \overline{\beta}_j$ *belongs to the range of* $\beta(\cdot) \cdot \nabla$, *say*, $\beta(\cdot) \cdot \nabla h = \beta_j - \overline{\beta}_j$ *for some* $h \in H^1$, *then*

$$
\lim_{u_0 \to \infty} K_{jj} = \lim_{u_0 \to \infty} E_{jj} + D_{jj} = 2\|h_0\|_1^2 + D_{jj}, \tag{3.36}
$$

where h_0 *is the projection of* h *on* $N^\perp$ *or, equivalently*, h_0 *is the unique element in* $N^\perp$ *such that* $\beta(\cdot) \cdot \nabla h_0 = \beta_j - \overline{\beta}_j$.

PROOF. (a) If $(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_N \neq 0$, then (3.35) is an immediate consequence of (3.12) and (3.34).

(b) In this case, $\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j) = \mathscr{D}^{-1}\beta(\cdot) \cdot \nabla h \equiv Sh$ belongs to the range of S and is, therefore, orthogonal to N. Writing

$$
h = \sum_{n=1}^{\infty} \langle h, \varphi_n \rangle_1 \varphi_n + h_N, \qquad Sh = \sum_{n=1}^{\infty} i\lambda_n \langle h, \varphi_n \rangle_1 \varphi_n,
$$

one has [see (3.32), (3.34)]

$$
\begin{aligned}
\beta_{jn} &= \big\langle \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j), \varphi_n \big\rangle_1 = \langle Sh, \varphi_n \rangle_1 = i\lambda_n \langle h, \varphi_n \rangle_1, \\
E_{jj} &= 2u_0^2 \sum_{n=1}^{\infty} \frac{\lambda_n^2 \langle h, \varphi_n \rangle_1^2}{1 + u_0^2 \lambda_n^2} \\
&\to 2\sum_{n=1}^{\infty} \langle h, \varphi_n \rangle_1^2 = 2\|h_0\|_1^2.
\end{aligned}
$$

□

REMARK 3.3.1. Under the hypothesis of part (a) of Theorem 3.3, the dispersion coefficient $K_{jj} = E_{jj} + D_{jj}$ [see (3.12)] grows quadratically with u_0. Experimental studies have shown a similar growth pattern for solute dispersion in saturated porous media. See Fried and Combarnous (1971), and Figures 1 and 2 in Section 7.

Consider now the diffusion $\tilde{X}(t)$ governed by the Itô equation (3.2) involving a spatial scale parameter a. Then $\tilde{X}(t) \bmod a$ is a diffusion on the torus $\mathscr{T}_a := \{x \bmod a\colon x \in \mathbb{R}^k\}$ and therefore $\tilde{X}(t)$ is asymptotically Gaussian. Indeed, in view of Proposition 3.1, the matrix $((\tilde{K}_{jj'}))$ of dispersion coefficients of this asymptotic distribution is the same as that of $\hat{X}(t)$, namely, $((K_{jj'}))$ for $u_0 = a$. The following is then an immediate consequence of Theorem 3.3.

COROLLARY 3.4. *Suppose the hypothesis of Proposition* 3.2 *holds for the coefficients of the diffusion* $\tilde{X}(t)$.

(a) *If* $(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_N \neq 0$, *then*

$$\lim_{a\to\infty} \frac{\tilde{K}_{jj}}{a^2} = 2\left\|(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_N\right\|_1^2 > 0. \tag{3.37}$$

(b) *If* $\beta_j - \overline{\beta}_j$ *belongs to the range of* $\beta(\cdot)\cdot\nabla$, $\beta_j - \overline{\beta}_j = \beta\cdot\nabla h$, *then, with* h_0 *as the projection of* h *on* $N^{\perp}$,

$$\lim_{a\to\infty} \tilde{K}_{jj} = 2\|h_0\|_1^2 + D_{jj}. \tag{3.38}$$

We now turn to the multiscale process of interest, namely [see (2.1)],

$$X(t) = X(0) + \int_0^t \{b(X(s)) + \beta(X(s)/a)\}\, ds + \sigma B(t), \tag{3.39}$$

where it is assumed that

(3.40)(A1) $b(\cdot)$, $\beta(\cdot)$ are continuously differentiable, divergence free and periodic with period lattice $\mathbb{Z}^k$;
(3.40)(A2) σ is a constant $k\times k$ nonsingular matrix;
(3.40)(A3) a is a positive integer.

Then

$$\dot{X}(t) := X(t) \bmod a \equiv (X_1(t) \bmod a, \dots, X_k(t) \bmod a) \tag{3.41}$$

is a diffusion on the torus $\mathscr{T}_a := \{x \bmod a\colon x \in \mathbb{R}^k\}$, whose unique invariant probability is the uniform distribution on $\mathscr{T}_a$. It is convenient to scale this process to bring it to the unit torus $\mathscr{T}_1 = \{x \bmod 1, x \in \mathbb{R}^k\}$. For this, define

$$Y(t) := \frac{X(a^2t)}{a}, \qquad \dot{Y}(t) := Y(t) \bmod 1 \equiv \frac{\dot{X}(a^2t)}{a}. \tag{3.42}$$

Note that apart from the scaling of distance, in which one unit of length in the Y-scale equals "a" units of length in the original X-scale, one unit of time for Y equals a^2 units of time for X. The process $Y(t)$ is governed by the Itô equation

$$Y(t) = Y(0) + \int_0^t a\{b(aY(s)) + \beta(Y(s))\}\,ds + \sigma\overline{B}(t), \tag{3.43}$$

where $\overline{B}(t) := B(a^2 t)/a$ is a standard Brownian motion on $\mathbb{R}^k$. The infinitesimal generator of $\dot{Y}(t)$ is given by $A_a = \mathscr{D} + a(b(a\cdot) + \beta)\cdot\nabla$, that is,

$$A_a f(x) = \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'}\frac{\partial^2 f(x)}{\partial x_j \partial x_{j'}} + a(b(ax) + \beta(x))\cdot\nabla f(x) \tag{3.44}$$

for smooth functions which are periodic: $f(y+r) = f(y)$ for all $r \in \mathbb{Z}^k$, all $y \in \mathbb{R}^k$; that is, A_a acts on (a dense subspace of) $\mathscr{L}^2(\mathscr{T}_1, dx)$. Let g_j $(1 \le j \le k)$ be the unique solution in $1^\perp$ of

$$A_a g_j(x) = b_j(ax) + \beta_j(x) - \overline{b}_j - \overline{\beta}_j. \tag{3.45}$$

By Itô's lemma [see (2.40)], writing $g = (g_1, \dots, g_k)'$,

$$\begin{aligned} &Y(t) - Y(0) - at(\overline{b} + \overline{\beta}) \\ &\quad = g(\dot{Y}(t)) - g(\dot{Y}(0)) - \int_0^t a\{\operatorname{grad} g(\dot{Y}(s)) - I\}\sigma\,d\overline{B}(s) \\ &\qquad\qquad \left(\operatorname{grad} g(x) := \left(\left(\frac{\partial g_j(x)}{\partial x_{j'}}\right)\right)\right). \end{aligned} \tag{3.46}$$

Hence, for a fixed "a,"

$$\frac{Y(t) - Y(0) - at(\overline{b} + \overline{\beta})}{\sqrt{t}} \to_{\mathscr{L}} \Phi_K \quad \text{as } t \to \infty, \tag{3.47}$$

where

$$\begin{aligned} K_{jj'} &= E_{jj'} + D_{jj'}, \\ E_{jj'} &:= a^2(\langle g_j, g_{j'}\rangle_1 + \langle g_{j'}, g_j\rangle_1), \qquad 1 \le j,\ j' \le k. \end{aligned} \tag{3.48}$$

Since the function $x \to b(ax)$ is *rapidly oscillating* for large a, one may think of approximating A_a by $\overline{A} = \mathscr{D} + a(\overline{b} + \beta)\cdot\nabla$,

$$\overline{A} f(x) = \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'}\frac{\partial^2 f(x)}{\partial x_j \partial x_{j'}} + a(\overline{b} + \beta)\cdot\nabla f(x). \tag{3.49}$$

Correspondingly, define on H^1 the skew symmetric compact operators

$$S_a = \mathscr{D}^{-1}(b(a\cdot) + \beta)\cdot\nabla, \qquad \overline{S} = \mathscr{D}^{-1}(\overline{b} + \beta)\cdot\nabla. \tag{3.50}$$

Let $\underline{N}$ denote the *null space* of $\overline{S}$. We will denote by $f_{\underline{N}}$ the orthogonal projection of an element f of H^1 on $\underline{N}$. The following result provides preliminary estimates of the norms of the solution g_j of the equation (3.45) in H^0 and H^1.

LEMMA 3.5. *Under assumptions* (A1)–(A3) *in* (3.40) *one has*

$$\sup_a \|g_j\|_1^2 < \infty, \qquad \|g_j\|_0^2 \le \frac{1}{2\pi^2\alpha_1}\|g_j\|_1^2, \tag{3.51}$$

where α_1 is the smallest eigenvalue of the diffusion matrix $D = \sigma\sigma'$.

PROOF. The operators S_a, $\overline{S}$ are skew symmetric so that

$$\langle S_a f, f\rangle_1 = 0, \qquad \langle \overline{S} f, f\rangle_1 = 0 \qquad \forall f \in H^1. \tag{3.52}$$

Therefore,

$$\big\|(\mathscr{I} + aS_a)f\big\|_1^2 = \|f\|_1^2 + a^2\|S_a f\|_1^2 \ge \|f\|_1^2. \tag{3.53}$$

Rewrite the defining equation (3.45) for g_j as

$$(\mathscr{I} + aS_a)g_j = \mathscr{D}^{-1}\big[b_j(a\cdot) - \overline{b}_j + \beta_j(\cdot) - \overline{\beta}_j\big]. \tag{3.54}$$

It now follows from (3.53) [also see (3.17)] that

$$\begin{aligned} \|g_j\|_1^2 &\le \big\|\mathscr{D}^{-1}\big[b_j(a\cdot) - \overline{b}_j + \beta_j(\cdot) - \overline{\beta}_j\big]\big\|_1^2 \\ &= \big\langle b_j(a\cdot) - \overline{b}_j + \beta_j(\cdot) - \overline{\beta}_j, -\mathscr{D}^{-1}\big[b_j(a\cdot) - \overline{b}_j + \beta_j(\cdot) - \overline{\beta}_j\big]\big\rangle_0. \end{aligned} \tag{3.55}$$

Writing $r = (r_1, \dots, r_k) \in \mathbb{Z}^k$, and using Parseval's relation and (3.20), one has for all $f \in H^0$,

$$\begin{aligned} \big\langle -\mathscr{D}^{-1} f, f\big\rangle_0 &= \sum_{r\in\mathbb{Z}^k\setminus\{\mathbf{0}\}} \Big(2\pi^2 \sum_{j,j'} D_{jj'} r_j r_{j'}\Big)^{-1} \big|\hat{f}(r)\big|^2 \\ &\le \sum_{r\in\mathbb{Z}^k\setminus\{\mathbf{0}\}} \big(2\pi^2\alpha_1|r|^2\big)^{-1}|\hat{f}(r)|^2 \\ &\le \frac{1}{2\pi^2\alpha_1}\|f\|_0^2. \end{aligned} \tag{3.56}$$

Therefore, (3.55) leads to the first inequality in (3.51). The second inequality in (3.51) follows from

$$\begin{aligned} \|f\|_1^2 = \langle -\mathscr{D} f, f\rangle_0 &= \textstyle\sum_{r\neq 0} 2\pi^2\Big(\sum_{j,j'} D_{jj'} r_j r_{j'}\Big)|\hat{f}(r)|^2 \\ &\ge \sum_{r\neq 0} 2\pi^2\alpha_1|r|^2|\hat{f}(r)|^2 \ge 2\pi^2\alpha_1\sum_{r\neq 0}|\hat{f}(r)|^2 \\ &= 2\pi^2\alpha_1\|f\|_0^2 \qquad \forall f \in H^1. \end{aligned} \tag{3.57}$$

□

The next lemma enables one to estimate the error in replacing $b(a\cdot)$ by $\overline{b}$ in variance calculations.

LEMMA 3.6. *Suppose $f \in H^0$ and "a" is a positive integer.*

(a) *If $\mathscr{G}$ is a relatively compact subset of H^1 then*

$$\sup_{g\in\mathscr{G}}\left|\langle f(a\cdot), g\rangle_0\right| = o\left(\frac{1}{a}\right) \quad \textit{as } a\to\infty, \tag{3.58}$$

(b) *If $\mathscr{G}$ is a relatively compact subset of H^0, then*

$$\sup_{g\in\mathscr{G}}\left|\langle f(a\cdot), g\rangle_0\right| = o(1) \quad \textit{as } a\to\infty. \tag{3.59}$$

PROOF. Assume first that f is continuously differentiable of all orders up to at least $[k/2]+1 = k_0$, say. Then

$$\langle f(a\cdot), g\rangle_0 = \sum_{r\in\mathbb{Z}^k\setminus\{0\}} f(a\cdot)\hat{\ }(r)\hat{g}^-(r), \qquad r\in\mathbb{Z}^k\setminus\{0\}. \tag{3.60}$$

Now

$$\begin{aligned}\sum_{r\neq 0}|\hat{f}(r)| &= \sum_{r\neq 0}\frac{|r|^{k_0}|\hat{f}(r)|}{|r|^{k_0}}\\ &\leq \left(\sum_{r\neq 0}|r|^{2k_0}|\hat{f}(r)|^2\right)^{1/2}\left(\sum_{r\neq 0}\frac{1}{|r|^{2k_0}}\right)^{1/2} < \infty.\end{aligned} \tag{3.61}$$

It follows that the Fourier series for f, namely $\sum_{r\neq 0}\hat{f}(r)\exp\{2\pi i r\cdot x\}$, converges uniformly to $f(x)$ so that

$$f(ax) = \sum_{r\neq 0}\hat{f}(r)\exp(2\pi i r\cdot ax) = \sum_{r\neq 0}\hat{f}(r)\exp(2\pi i a r\cdot x). \tag{3.62}$$

In particular,

$$f(a\cdot)^\wedge(r) = \begin{cases} 0, & \text{if } r\notin a\mathbb{Z}^k\setminus\{0\},\\ \hat{f}(r/a), & \text{if } r\in a\mathbb{Z}^k\setminus\{0\}.\end{cases} \tag{3.63}$$

Using this in (3.60) we get

$$\begin{aligned}\left|\langle f(a\cdot), g\rangle_0\right| &= \left|\sum_{r\in a\mathbb{Z}^k\setminus\{0\}}\hat{f}(r/a)\hat{g}^-(r)\right|\\ &\leq \left(\sum_{r\in a\mathbb{Z}^k\setminus\{0\}}\left|\hat{f}(r/a)\right|^2\right)^{1/2}\left(\sum_{|r|\geq a}\left|\hat{g}(r)\right|^2\right)^{1/2}\\ &\leq \|f_0\|\left(\sum_{|r|\geq a}\frac{1}{a^2}|r|^2|\hat{g}(r)|^2\right)^{1/2}\\ &\leq \frac{\|f\|_0}{a}\left(\sum_{|r|\geq a}|r|^2|\hat{g}(r)|^2\right)^{1/2} \leq c\|f\|_0\|g\|_1/a.\end{aligned} \tag{3.64}$$

To prove (3.58), note that if $\mathscr{G}$ is a relatively compact subset of H^1 then

$$\sup_{g\in\mathscr{G}}\biggl(\sum_{|r|\geq a}|r|^2|\hat{g}(r)|^2\biggr)^{1/2}\to 0\quad\text{as } a\to\infty. \tag{3.65}$$

To prove part (b), use the first inequality in (3.64) to get

$$\bigl|\langle f(a\cdot), g\rangle_0\bigr|\leq \|f\|_0\biggl(\sum_{|r|\geq a}|\hat{g}(r)|^2\biggr)^{1/2}, \tag{3.66}$$

and note that the right side goes to zero as $a\to\infty$, uniformly for g belonging to a relatively compact subset of H^0.

Since the final estimates in (3.64), (3.66) involve only the H^0-norm of f, and the set of all infinitely differentiable functions in H^0 is dense in H^0, the proof of the lemma is complete. □

We are now ready to prove two of the main results of this section. The following technical condition will be made use of in the proof.

Consider the "approximation" of g_j provided by the solution h_j in H^1 to the equation

$$\begin{gathered}\overline{A}h_j=\beta_j-\overline{\beta}_j\quad\text{or}\\ (\mathscr{I}+a\overline{S})h_j=\mathscr{D}^{-1}(\beta_j-\overline{\beta}_j),\end{gathered} \tag{3.67}$$

and let $i\lambda_n$ be the eigenvalues of $\overline{S}$ corresponding to normalized eigenfunctions φ_n $(n\geq 1)$. We assume

(3.68)(A4)j (i) $\{g_j\colon a\geq 1\}$, $\{g_j^{-}\partial/\partial x_s(\mathscr{D}^{-1}(\beta_j-\overline{\beta}_j))_{\underline{N}}\colon a\geq 1\}$ $(1\leq s\leq k)$ are relatively compact subsets of H^1, and
(ii) $\{\varphi_n\partial g_j^{-}/\partial x_s\colon a\geq 1\}$ $(1\leq s\leq k,\ n\geq 1)$ are relatively compact in H^0.

See Remark 3.7.1 for some simpler conditions which guarantee (A4)j.

For the statement of the theorem below recall that $K_{jj'}=E_{jj'}+D_{jj'}$ are the elements of the dispersion matrix of the limiting Gaussian distribution of the scaled $Y(t)$ process (3.47) [see (3.48)].

THEOREM 3.7. *Assume* (A1)–(A3) *in* (3.40), *and* (A4)j *in* (3.67). *Then*

$$\lim_{a\to\infty}\frac{K_{jj}}{a^2}=2\bigl\|(\mathscr{D}^{-1}(\beta_j-\overline{\beta}_j))_{\underline{N}}\bigr\|_1^2. \tag{3.69}$$

PROOF. Since [see (3.48)]

$$K_{jj}=a^2E_{jj}+D_{jj}=2a^2\|g_j\|_1^2+D_{jj}, \tag{3.70}$$

it is enough to show that

$$\lim_{a\to\infty}\|g_j\|_1^2=\bigl\|(\mathscr{D}^{-1}(\beta_j-\overline{\beta}_j))_{\underline{N}}\bigr\|_1^2. \tag{3.71}$$

Now g_j solves (3.45) or (3.54). Hence [see (3.52) and Lemma 3.6],

$$
\begin{aligned}
\|g_j\|_1^2 &= \langle g_j, (\mathscr{I}+aS_a)g_j\rangle_1 = \langle g_j, \mathscr{D}^{-1}[b_j(a\cdot)-\overline{b}_j+\beta_j-\overline{\beta}_j]\rangle_1 \\
&= -\langle g_j, b_j(a\cdot)-\overline{b}_j+\beta_j-\overline{\beta}_j\rangle_0 \simeq -\langle g_j, \beta_j-\overline{\beta}_j\rangle_0 \\
&= \langle g_j, \mathscr{D}^{-1}(\beta_j-\overline{\beta}_j)\rangle_1.
\end{aligned}
\tag{3.72}
$$

Here $\simeq$ indicates that the difference between its two sides goes to zero as $a \to \infty$. As in the proof of Theorem 3.3(a), letting $i\lambda_n$ be the eigenvalues of $\overline{S}$ corresponding to eigenfunctions φ_n ($n \geq 1$), one may express the second equation in (3.67) as

$$
\begin{aligned}
&(h_j)_{\underline{N}} = \big(\mathscr{D}^{-1}(\beta_j-\overline{\beta}_j)\big)_{\underline{N}},\\
&\langle h_j, \varphi_n\rangle_1 = \frac{\beta_{jn}}{1+ia\lambda_n}, \qquad \beta_{jn} := \langle \mathscr{D}^{-1}(\beta_j-\overline{\beta}_j), \varphi_n\rangle_1, \qquad n \geq 1.
\end{aligned}
\tag{3.73}
$$

Hence

$$
\big\|h_j - (\mathscr{D}^{-1}(\beta_j-\overline{\beta}_j))_{\underline{N}}\big\|_1^2 = \sum_{n=1}^{\infty} \frac{\beta_{jn}^2}{1+a^2\lambda_n^2} \to 0 \quad \text{as } a \to \infty,
\tag{3.74}
$$

since the sum on the right is bounded above by $\sum_{n=1}^{\infty} \beta_{jn}^2 \leq \|\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\|_1^2$ for all a. Hence,

$$
h_j \to \big(\mathscr{D}^{-1}(\beta_j-\overline{\beta}_j)\big)_{\underline{N}} \quad \text{in } H^1\text{-norm, as } a \to \infty.
\tag{3.75}
$$

Now, using $(\mathscr{I} + aS_a)g_j = \mathscr{D}^{-1}[b_j(a\cdot) - \overline{b}_j + \beta_j - \overline{\beta}_j]$ and $(\mathscr{I} + a\overline{S})h_j = \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)$, one gets

$$
(\mathscr{I}+a\overline{S})(g_j-h_j) = \mathscr{D}^{-1}(b_j(a\cdot)-\overline{b}_j) - a\mathscr{D}^{-1}(b(a\cdot)-\overline{b})\cdot\nabla g_j.
\tag{3.76}
$$

Therefore,

$$
(g_j-h_j)_{\underline{N}} = \big(\mathscr{D}^{-1}(b_j(a\cdot)-\overline{b}_j)\big)_{\underline{N}} - a\big(\mathscr{D}^{-1}[(b(a\cdot)-\overline{b})\cdot\nabla g_j]\big)_{\underline{N}}
\tag{3.77}
$$

and

$$
\begin{aligned}
\langle g_j - h_j, \varphi_n\rangle_1 &= \frac{1}{1+ia\lambda_n}\langle \mathscr{D}^{-1}(b_j(a\cdot)-\overline{b}_j), \varphi_n\rangle_1\\
&\quad - \frac{a}{1+ia\lambda_n}\langle \mathscr{D}^{-1}(b(a\cdot)-\overline{b})\cdot\nabla g_j, \varphi_n\rangle_1, \qquad n \geq 1.
\end{aligned}
\tag{3.78}
$$

The first term on the right in (3.78) goes to zero by (3.58) or (3.59). To evaluate the second term, express the inner product as

$$
\begin{aligned}
\langle \mathscr{D}^{-1}(b(a\cdot)-\overline{b})\cdot\nabla g_j, \varphi_n\rangle_1 &= \langle -(b(a\cdot)-\overline{b})\cdot\nabla g_j, \varphi_j\rangle_0\\
&= -\sum_{s=1}^{k} \langle (b_s(a)-\overline{b}_s)\frac{\partial g_j}{\partial x_s}, \varphi_n\rangle_0\\
&= -\sum_{s=1}^{k} \left\langle (b_s(a\cdot)-\overline{b}_s), \varphi_n \frac{\partial \overline{g_j}}{\partial x_s}\right\rangle_0 \to 0 \quad \text{as } a \to \infty,
\end{aligned}
\tag{3.79}
$$

using the assumption that $\{\varphi_n \partial g_j^- / \partial x_s\colon a \geq 1\}$, $1 \leq s \leq k$, are relatively compact subsets of H^0 [see (A4)j and Lemma 3.6(b)]. Thus

$$\langle g_j - h_j, \varphi_n \rangle_1 \to 0 \quad \text{as } a \to \infty \ (n = 1, 2, \ldots). \tag{3.80}$$

Since $\{g_j - h_j\colon a = 1, 2, \ldots\}$ is relatively compact in H^1, (3.80) implies $(g_j - h_j)_{\underline{N}^\perp} \to 0$ weakly in H^1, that is, $(g_j)_{\underline{N}^\perp} \to 0$ weakly in H^1. Now use (3.72) to write

$$\begin{aligned}
\|g_j\|_1^2 &\simeq \langle g_j, \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\rangle_1 \\
&= \langle g_j, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\rangle_1 + \langle g_j, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}^\perp}\rangle_1 \\
&\simeq \langle g_j, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\rangle_1 \\
&= \langle h_j, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\rangle_1 + \langle g_j - h_j, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\rangle_1 \\
&\simeq \left\|(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\right\|_1^2 + \langle (g_j - h_j)_{\underline{N}}, \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\rangle_1.
\end{aligned} \tag{3.81}$$

Now, by (3.77) and Lemma 3.6,

$$\begin{aligned}
&\langle (g_j - h_j)_{\underline{N}}, \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\rangle_1 \\
&\quad \simeq -a\langle (\mathscr{D}^{-1}[(b(a\cdot) - \overline{b}) \cdot \nabla g_j])_{\underline{N}}, \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\rangle_1 \\
&\quad = -a\langle \mathscr{D}^{-1}[(b(a\cdot) - \overline{b}) \cdot \nabla g_j], (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\rangle_1 \\
&\quad = a\langle (b(a\cdot) - \overline{b}) \cdot \nabla g_j, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\rangle_0 \\
&\quad = a\sum_{s=1}^{k} \left\langle \frac{\partial}{\partial x_s}\{(b_s(a\cdot) - \overline{b}_s)g_j\}, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\right\rangle_0 \\
&\quad = -a\sum_{s=1}^{k} \left\langle (b_s(a\cdot) - \overline{b}_s), g_j^- \frac{\partial}{\partial x_s}(\mathscr{D}^{-1}(\beta_j - \overline{\beta})_j))_{\underline{N}}\right\rangle_0 \to 0,
\end{aligned} \tag{3.82}$$

since $\{g_j^-(\partial/\partial x_s)(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\colon a \geq 1\}$, $1 \leq s \leq k$, are relatively compact subsets of H^1. Relations (3.81) and (3.82) imply (3.71). $\square$

REMARK 3.7.1. Assumption (A4)j is probably redundant, in the presence of assumptions (A1)–(A3), for the proof of the theorem above. But we are unable to dispense with it.

A set of sufficient conditions for (A4)j to hold are

$$\sup_{a \geq 1} \sup_x |\nabla g_j(x)| < \infty, \quad \lim_{a \to \infty} \nabla g_j(x) \text{ exists a.e.} \tag{3.83}$$

and

$$(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}} \text{ have bounded first, and second-order derivatives.} \tag{3.84}$$

Indeed, (3.83) guarantees that g_j converges in H^1 to some q, say. One may then use the inequality

$$\|uv\|_1^2 \le c'\left(\|u\|_1^2\|v\|_\infty^2 + \|u\|_0^2\left(\sum_{s'=1}^{k}\left\|\frac{\partial v}{\partial x_{s'}}\right\|_\infty^2\right)\right), \tag{3.85}$$

with $u = g_j - q$ and $v = \partial/\partial x_s(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}$. One also has the more symmetric inequality,

$$\|uv\|_1^2 \le c'\big(\|u\|_1^2\|v\|_\infty^2 + \|u_\infty\|^2\|v\|_1^2\big). \tag{3.86}$$

It may be noted in this connection that the inequality (3.86) corrects a careless error in Bhattacharya and Götze [(1995), relation (4.86)]. Finally, it may be noted that (3.83), (3.84) hold in the examples in Section 6.

The next result deals with the case

$$\big(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\big)_{\underline{N}} = 0. \tag{3.87}$$

We will make use of the following assumption, in addition to (A1)–(A3):

(3.88)(A5)j There exists a twice continuously differentiable solution $p \in H^1$ of the equation

$$(\overline{b} + \beta)\cdot\nabla p = \beta_j - \overline{\beta}_j.$$

Note that (A5)j says that $\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)$ belongs to the *range of* $\overline{S}$, so that (3.87) holds.

THEOREM 3.8. *Assume* (A1)–(A3), (A5)j. *Then*

$$D_{jj} \le \liminf_{a\to\infty} K_{jj} \le \limsup_{a\to\infty} K_{jj} < \infty. \tag{3.89}$$

PROOF. In view of (3.48) [see (3.70)], it is enough to show that

$$\limsup_{a\to\infty} a^2\|g_j\|_1^2 < \infty. \tag{3.90}$$

Letting p be as in (3.88) one has, by the last inequality in (3.64),

$$\begin{aligned}\|g_j\|_1^2 &= \langle g_j, (\mathscr{I} + aS_a)g_j\rangle_1\\ &= \langle g_j, \mathscr{D}^{-1}(b_j(a\cdot) - \overline{b_j})\rangle_1 + \langle g_j, \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)\rangle_1\\ &\le \frac{c_1\|g_j\|_1\|b_j\|_0}{a} + \langle g_j, \overline{S}p\rangle_1.\end{aligned} \tag{3.91}$$

Now

$$
(3.92)\qquad
\begin{aligned}
\langle g_j, \overline{S}p\rangle_1 &= \langle g_j, S_a p\rangle_1 - \langle g_j, (S_a - \overline{S})p\rangle_1 \\
&= -\langle S_a g_j, p\rangle_1 - \langle g_j, \mathscr{D}^{-1}(b(a\cdot) - \overline{b})\cdot \nabla p\rangle_1 \\
&= -\frac{1}{a}\langle -g_j + \mathscr{D}^{-1}(b_j(a\cdot) - \overline{b}_j) + \mathscr{D}^{-1}(\beta_j - \overline{\beta}_j), p\rangle_1 \\
&\quad - \langle g_j, \mathscr{D}^{-1}(b(a\cdot) - \overline{b})\cdot \nabla p\rangle_1 \\
&= \frac{1}{a}\langle g_j, p\rangle_1 + \frac{1}{a}\langle b_j(a\cdot) - \overline{b}_j, p\rangle_0 \\
&\quad - \frac{1}{a}\langle \overline{S}p, p\rangle_1 - \langle g_j, \mathscr{D}^{-1}(b(a\cdot) - \overline{b})\cdot \nabla p\rangle_1 .
\end{aligned}
$$

Since $\langle \overline{S}p, p\rangle_1 = 0$, and $p \in H^1$, (3.91) and (3.92) lead to

$$
(3.93)\qquad \|g_j\|_1^2 \le \frac{c_2\|g_j\|_1}{a} + \frac{c_3}{a^2} + \left|\langle g_j, \mathscr{D}^{-1}(b(a\cdot) - \overline{b})\cdot \nabla p\rangle_1\right|.
$$

By Lemma 3.6(a),

$$
(3.94)\qquad
\begin{aligned}
&\left|\langle g_j, \mathscr{D}^{-1}(b(a\cdot) - \overline{b})\cdot \nabla p\rangle_1\right| \\
&\quad = \left|\left\langle g_j, \sum_{s=1}^{k}(b_s(a\cdot) - \overline{b}_s)\frac{\partial p}{\partial x_s}\right\rangle_0\right| \\
&\quad = \left|\sum_{s=1}^{k}\left\langle g_j\frac{\partial p^-}{\partial x_s}, b_s(a\cdot) - \overline{b}_s\right\rangle_0\right| \le c_4\|g_j\|_1/a.
\end{aligned}
$$

The last inequality follows from (3.58) using the fact that [see (3.85)]

$$
(3.95)\qquad
\begin{aligned}
\left\| g_j \tfrac{\partial p^-}{\partial x_s}\right\|_1^2 &\le c'\left\{\|g_j\|_1^2\left\|\frac{\partial p^-}{\partial x_s}\right\|_\infty^2 + \|g_j\|_0^2\sum_{s'=1}^{k}\left\|\frac{\partial^2 p^-}{\partial x_{s'}\partial x_s}\right\|_\infty^2\right\} \\
&\le c_5\|g_j\|_1^2.
\end{aligned}
$$

From (3.93), (3.94), one derives the relation

$$
(3.96)\qquad a^2\|g_j\|_1^2 \le c_6 a\|g_j\|_1 + c_7,
$$

where c_6 and c_7 do not depend on "a." It is clear from (3.96) that $\{a\|g_j\|_1 : a = 1, 2, \ldots\}$ is a bounded sequence, that is, (3.90) holds. □

REMARK 3.8.1. The assumption of boundedness of derivatives of p in (A5)j is probably redundant. In any case, it is satisfied in Example 2 of Section 6.

4. Speed of convergence to equilibrium of diffusions on a big torus. A crucial element in the analysis of the asymptotic behavior of multiscale diffusions with periodic coefficients, such as $X(t)$ in (3.39), is the estimation of the total variation distance between the distribution at large times t of the corresponding diffusion $\dot{X}(\cdot) := X(\cdot) \bmod a$ on the big torus $\mathcal{T}_a = \{x \mod a \colon x \in \mathbb{R}^k\}$ and its equilibrium distribution uniformly with respect to all initial states of $\dot{X}(\cdot)$. To derive such an estimate we first obtain an analogue of a result of Fill (1991) [also see Diaconis and Stroock (1991) for the time-reversible case], which holds for general Markov processes in continuous time.

Let $U(t)$, $t \geq 0$, be a Markov process on a measurable state space $(M, \mathcal{M})$, having a transition probability density $r(t; x, y)$ with respect to a sigma-finite measure ν. Suppose r admits a unique invariant probability $\pi(dx) = \pi(x)\nu(dx)$. Let $\mathcal{L}^2 = \mathcal{L}^2(M, \pi)$ be the real Hilbert space of square integrable (w.r.t. π) functions on M and T_t $(t > 0)$ the semigroup of transition operators on $\mathcal{L}^2$,

$$(T_t f)(x) = \int f(y) r(t; x, y)\nu(dy), \qquad f \in \mathcal{L}^2. \tag{4.1}$$

Define also the transition probability density $q(t; x, y)$ of the *time-reversed Markov process,* by

$$q(t; x, y) = r(t; y, x)\pi(y)/\pi(x) \tag{4.2}$$

if $\pi(x) > 0$ [and arbitrarily, measurably, if $\pi(x) = 0$]. Let $\tilde{T}_t$, $t > 0$, denote the corresponding transition semigroup,

$$(\tilde{T}_t g)(x) = \int g(y) q(t; x, y)\nu(dy), \qquad g \in \mathcal{L}^2. \tag{4.3}$$

It is simple to check that $\tilde{T}_t$ is the *adjoint* of T_t, that is,

$$\langle T_t f, g\rangle = \langle f, \tilde{T}_t g\rangle, \qquad f, g \in \mathcal{L}^2, \tag{4.4}$$

where $\langle\, ,\, \rangle$ is the *inner product* on $\mathcal{L}^2$. Let B and $\tilde{B}$ denote the infinitesimal generators of the semigroups T_t $(t > 0)$ and $\tilde{T}_t$ $(t > 0)$, respectively, and $\mathbf{D}_B$, $\mathbf{D}_{\tilde{B}}$ their domains. Let $1^{\perp}$ denote the subspace of $\mathcal{L}^2$ orthogonal to constants and write $\|\cdot\|$ for the *norm* in $\mathcal{L}^2$.

PROPOSITION 4.1. *Assume that $\mathbf{D}_{\tilde{B}}$ is dense in $\mathcal{L}^2$, and define*

$$\lambda = \inf\bigl\{\langle -\tilde{B} f, f\rangle \colon f \in 1^{\perp} \cap \mathbf{D}_{\tilde{B}}, \|f\| = 1\bigr\}. \tag{4.5}$$

Then if $U(0)$ has a probability density η w.r.t. ν, the density η_t of $U(t)$ satisfies

$$\int \bigl|\eta_t(y) - \pi(y)\bigr|\nu(dy) \leq e^{-\lambda t}\|\psi_0\|, \tag{4.6}$$

where ψ_0 is given by

$$\psi_0(y) = \frac{\eta(y) - \pi(y)}{\pi(y)} \quad a.e., w.r.t.\ \pi(dy). \tag{4.7}$$

PROOF. Without loss of generality, assume $\|\psi_0\| < \infty$, that is, $\psi_0 \in 1^\perp$. Now if $g \in 1^\perp \cap \mathbf{D}_{\tilde{B}}$, then $\tilde{T}_t g \in 1^\perp \cap \mathbf{D}_{\tilde{B}}$ $\forall\, t > 0$ so that, by (4.5),

$$\begin{aligned} \frac{d}{dt}\|\tilde{T}_t g\|^2 &= \frac{d}{dt}\langle \tilde{T}_t g, \tilde{T}_t g\rangle = 2\langle \tilde{T}_t g, \tilde{B}\tilde{T}_t g\rangle \\ &\leq -2\lambda\|\tilde{T}_t g\|^2, \end{aligned} \tag{4.8}$$

leading to

$$\|\tilde{T}_t g\|^2 \leq e^{-2\lambda t}\|g\|^2, \qquad g \in 1^\perp \cap \mathbf{D}_{\tilde{B}}. \tag{4.9}$$

Note that $1^\perp \cap \mathbf{D}_{\tilde{B}}$ is dense in $1^\perp$, since $\mathbf{D}_{\tilde{B}}$ is dense in $\mathscr{L}^2$. Hence (4.9) holds for all $g \in 1^\perp$. Now one may write

$$\begin{aligned} \tilde{T}_t\left(\frac{\eta}{\pi}\right)(y) &= \int \frac{\eta(x)}{\pi(x)}\, q(t; y, x)\nu(dx) \\ &= \int \frac{\eta(x)}{\pi(x)}\, r(t; x, y)\frac{\pi(x)}{\pi(y)}\,\nu(dx) = \frac{\eta_t(y)}{\pi(y)}, \end{aligned} \tag{4.10}$$

which implies $\tilde{T}_t\psi_0 = \eta_t/\pi - 1$. Therefore, by the Cauchy–Schwarz inequality and (4.9),

$$\begin{aligned} \int |\eta_t(y) - \pi(y)|\nu(dy) &= \int \left|\frac{\eta_t(y)}{\pi(y)} - 1\right|\pi(y)\nu(dy) \\ &= \int \big|\tilde{T}_t\psi_0(y)\big|\pi(y)\nu(dy) \\ &\leq \|\tilde{T}_t\psi_0\| \leq e^{-\lambda t}\|\psi_0\|. \end{aligned} \tag{4.11}$$

□

REMARK 4.1.1. If B is *self–adjoint*, that is, if $\tilde{B} = B$, λ defined in (4.5) is the *spectral gap* of B. Note that in this case the spectrum of B lies on the negative half of the real-axis (in the complex plane), with 0 as the simple eigenvalue corresponding to the eigenspace of constants in $\mathscr{L}^2(M, \pi)$. The point of the rest of the spectrum closest to 0 is $-\lambda$, if $\lambda > 0$. If B is not self-adjoint then, assuming that the symmetric operator $B + \tilde{B}$ is closed with a domain dense in $\mathscr{L}^2(M, \pi)$, the quantity λ in (4.5) is the spectral gap of $\frac{1}{2}(B+\tilde{B})$. For notational purposes, we will often write λ_B for λ in (4.5).

The following simple lemma shows the change in λ that occurs under a change in the time scale.

LEMMA 4.2. *Assume the hypothesis of Proposition* 4.1 *and consider the Markov process* $V(t) := U(ct)$, $t \geq 0$.

(a) *Then* $V(\cdot)$ *has invariant probability* π *and for its infinitesimal generator* B_c, *say, one has*

$$\lambda_{B_c} := \inf\big\{\langle -\tilde{B}_c f, f\rangle \colon \|f\| = 1, f \in 1^\perp \cap \mathbf{D}_{\tilde{B}_c}\big\} = c\lambda_B. \tag{4.12}$$

(b) *Also, if* $V(0)$ *has a probability density* η *w.r.t.* ν, *and* $V(t)$ *has the corresponding density* η_t, *then one has*

$$\int_M \left|\eta_t(y) - \pi(y)\right| \nu(dy) \le \exp(-c\lambda_B t)\|\psi_0\|, \tag{4.13}$$

where $\psi_0(y) = (\eta_0(y) - \pi(y))/\pi(y)$.

PROOF. Clearly, $V(\cdot)$ has the same invariant probability as $U(\cdot)$. Also, the infinitesimal generator of $V(\cdot)$ is $B_c = cB$ (with domain $\mathbf{D}_{B_c} = \mathbf{D}_B$), so that $\tilde{B}_c = c\tilde{B}$ and λ_{B_c} is given by

$$\inf\left\{\left\langle -c\tilde{B}f, f\right\rangle : \|f\| = 1, f \in 1^{\perp} \cap \mathbf{D}_{\tilde{B}}\right\} = c\lambda_B. \tag{4.14}$$

This proves part (a). Part (b) follows from (4.7). □

We now apply this lemma to the scaled diffusion $\dot{Y}$ on the unit torus and its generator A_a on $\mathscr{L}^2(\mathscr{T}_1, dx)$ [see (3.42), (3.44)]. Recall that, under the divergence-free assumption in (3.40), dx is the unique invariant probability of $\dot{Y}$. The adjoint operator $\tilde{A}_a$ is then easily seen to be

$$\begin{aligned}\tilde{A}_a &= \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'} \frac{\partial^2}{\partial x_j \partial x_{j'}} - a(b(a\cdot) + \beta)\cdot\nabla \\ &= \mathscr{D} - a(b(a\cdot) + \beta)\cdot\nabla.\end{aligned} \tag{4.15}$$

Denote by $\overline{\lambda}$ the infimum in (4.5) for the case $B = A_a$, $\tilde{B} = \tilde{A}_a$. That is, $\overline{\lambda}$ is the spectral gap of $\frac{1}{2}(\mathscr{D} + a(b(a\cdot) + \beta)\cdot\nabla) + \frac{1}{2}(\mathscr{D} - a(b(\cdot) + \beta)\cdot\nabla) = \mathscr{D}$ on $\mathscr{L}^2(\mathscr{T}_1, dx)$.

PROPOSITION 4.3. *Under assumptions* (A1)–(A3) *in* (3.40), *writing* α_1 *for the smallest eigenvalue of the matrix* $((D_{jj'}))$, *and* $\lambda_1 = \min\{D_{jj}: 1 \le j \le k\}$, *one has*

$$2\pi^2\alpha_1 \le \overline{\lambda} \le 2\pi^2\lambda_1. \tag{4.16}$$

PROOF. Denote by $\mathbf{D}$ the domain of $\mathscr{D}$ on $\mathscr{L}^2(\mathscr{T}_1, dx)$. As before, let $\hat{f}$ denote the Fourier transform of f on $\mathscr{L}^2(\mathscr{T}_1, dx)$. Then one has

$$\begin{aligned}\overline{\lambda} &= \inf\left\{-\langle f, \mathscr{D}f\rangle : \|f\| = 1, f \in 1^{\perp} \cap \mathbf{D}\right\} \\ &= \inf\left\{2\pi^2 \sum_{r\in\mathbb{Z}^k\setminus\{\mathbf{0}\}} \left|\hat{f}(r)\right|^2 \left(\sum_{j,j'} D_{jj'} r_j r_{j'}\right)\right\} \\ &\ge 2\pi^2 \inf\left\{\sum_{r\in\mathbb{Z}^k\setminus\{0\}} \left|\hat{f}(r)\right|^2 \alpha_1 |r|^2\right\} \ge 2\pi^2\alpha_1.\end{aligned} \tag{4.17}$$

On the other hand, letting $f(x) = \sqrt{2}\cos 2\pi x_j$, one gets $-\langle f, \mathscr{D}f\rangle = 2\pi^2 D_{jj}$. Hence $\overline{\lambda} \le 2\pi^2\lambda_1$. □

By using Lemma 4.2, one arrives at the following corollary of Proposition 4.3. To state it, assume (A1)–(A3). Let L denote the generator of the diffusion $\dot{X}(t)$ on the big torus $\mathscr{T}_a = \{x \bmod a\colon x \in \mathbb{R}^k\}$, and let m denote the normalized Lebesgue measure or the *uniform distribution* on $\mathscr{T}_a$. Note that m is the unique invariant probability of $\dot{X}$ and that $\frac{1}{2}(L+\tilde{L}) = \mathscr{D}$ on $\mathscr{L}^2(\mathscr{T}_a, m)$. Let $\mathbf{D}$ denote the domain of $\mathscr{D}$ in $\mathscr{L}^2(\mathscr{T}_a, m)$.

COROLLARY 4.4. *Assume* (A1)–(A3) *in* (3.40). *Then the quantity* $\lambda_L := \inf\{-\langle f, \tilde{L} f\rangle\colon \|f\| = 1,\ f \in 1^{\perp} \cap \mathbf{D}_{\tilde{L}}\}$ *satisfies*

$$2\pi^2 \frac{\alpha_1}{a^2} \leq \lambda_L \leq \frac{2\pi^2}{a^2}\lambda_1, \tag{4.18}$$

where α_1, λ_1 *are as in Proposition* 4.3.

PROOF. First note that $V(t) := \dot{X}(a^2 t)$, $t \geq 0$, has the generator $a^2 L$ on $\mathscr{L}^2(\mathscr{T}_a, m)$. The generator A_a of $\dot{Y}(t) = V(t)/a$ has the same spectrum on $\mathscr{L}^2(\mathscr{T}_1, dx)$ as that of $V(t)$ on $\mathscr{L}^2(\mathscr{T}_a, m)$. Therefore, $\overline{\lambda} = \lambda_{a^2 L} \in [c_1, c_2]$, where c_1, c_2 are as in (4.16). On the other hand, by Lemma 4.2, $\lambda_{a^2L} = a^2 \lambda_L$. Hence $\lambda_L = 1/a^2 \lambda_{a^2 L} \in [c_1/a^2, c_2/a^2]$. $\square$

One of the main results of this section may now be stated and proved.

THEOREM 4.5. *Assume* (A1)–(A3) *in* (3.40), *and let* $p_a(t; x, y)$ *denote the transition probability density of* $\dot{X}(t)$ *with respect to Lebesgue measure on* $[0, a)^k$. *Then there exists a positive constant* c_5 *independent of* a *such that*

$$\sup_x \int_{[0,a)^k} \left| p_a(t; x, y) - \frac{1}{a^k} \right| dy \leq c_5 a^{k/2} \exp\{-2\pi^2 \alpha_1 t/a^2\}, \tag{4.19}$$

α_1 *being the smallest eigenvalue of the matrix* $((D_{jj'}))$.

PROOF. By Corollary 4.4 and Proposition 4.1 one has, for every initial density η of $\dot{X}$,

$$\int_{[0,a)^k} \left| \eta_t(y) - \frac{1}{a^k} \right| dy \leq \exp\left(\frac{-2\pi^2 \alpha_1 t}{a^2}\right) \|\psi_0\|, \tag{4.20}$$

where $\psi_0(y) = (\eta(y) - a^{-k})/a^{-k}$, and η_t is the density of $\dot{X}(t)$. Now

$$\begin{aligned} \|\psi_0\|^2 &= a^{2k} \int_{[0,a)^k} \big(\eta^2(y) + a^{-2k} - 2a^{-k}\eta(y)\big) a^{-k}\, dy \\ &= a^k \int_{[0,a)^k} \eta^2(y) dy - 1 \\ &\leq a^k \sup\{\eta(y)\colon y \in [0,a)^k\}. \end{aligned} \tag{4.21}$$

Since $p_a(t; x, y)$ is the density of $\dot{X}(t)$, when $\dot{X}(0)$ has the degenerate distribution δ_x, we will apply (4.20), (4.21) to $\eta(y) = p_a(1; x, y)$ and with t replaced by $t-1$ to get

$$\int_{[0,a)^k}\left|p_a(t;x,y)-\frac{1}{a^k}\right|dy \\ \le \left(a^k \sup\left\{p_a(1;x,y)\colon y\in[0,a)^k\right\}\right)^{1/2}\exp\left\{-2\pi^2\alpha_1(t-1)/a^2\right\}. \tag{4.22}$$

To estimate the supremum on the right side we apply a result of Aronson (1967), which implies that the transition probability density $f(1;x,y)$ of the process $X(t)$ satisfies

$$f(1;x,y)\le c'\exp\{-c|x-y|^2\},\qquad x,y\in\mathbb{R}^k, \tag{4.23}$$

where c and c' are positive constants not depending on a. Now

$$p_a(1;x,y)=\sum_{r\in\mathbb{Z}^k} f(1;x,y+ar),\qquad x,y\in[0,a)^k \\ \le c'\sum_{r\in\mathbb{Z}^k}\exp\left\{-c|x-y-ar|^2\right\}\le c'', \tag{4.24}$$

where c'' does not depend on a. Therefore, (4.22) and (4.24) lead to (4.19). □

REMARK 4.5.1. It follows from the above proof that the transition density $q_a(t;x,y)$, say, of $\dot{Y}(t)$ satisfies the inequality

$$\sup_x \int_{[0,1)^k}\left|q_a(t;x,y)-1\right|dy\le c_5 a^{k/2}\exp\{-2\pi^2\alpha_1 t\}. \tag{4.19)$'$}$$

REMARK 4.5.2. One may extend Theorem 4.5 by relaxing the assumptions (A1)–(A3) to the case of diffusions $X(t)$ with generators of the form

$$L=\frac{1}{2}\sum_{j,j'=1}^{k}\frac{\partial}{\partial x_j}(D_{jj'}(x))\frac{\partial}{\partial x_{j'}}+\sum_{j=1}^{k}\{b_j(x)+\beta_j(x/a)\}\frac{\partial}{\partial x_j}, \tag{4.25}$$

where the assumptions (A1), (A3) hold for b_j, β_j and a, but (A2) is replaced by

(A2)′ $((D_{jj'}(x)))$ is a (positive definite)-matrix valued continuously differentiable periodic function with period lattice $\mathbb{Z}^k$.

In this case the diffusion $\dot{X}(t)=X(t)\bmod a$ on the big torus $\mathscr{T}_a$ has again as its unique invariant probability the normalized Lebesgue measure $m=a^{-k}\,dx$, whose generator on $\mathscr{L}^2(\mathscr{T}_a,m)$ is L-restricted to periodic functions. Also, $\frac{1}{2}(L+\tilde{L})=\mathscr{D}_1:=\frac{1}{2}\sum_{j,j'}(\partial/\partial x_j)\,(D_{jj'}(x))\,(\partial/\partial x_{j'})$ is self-adjoint on $\mathscr{L}^2(\mathscr{T}_a,m)$ and has a spectral gap $O(1/a^2)$. This last statement is a consequence of the fact that for the generator A_a of $\dot{Y}(t):=\dot{X}(a^2t)/a$ one has $\frac{1}{2}(A_a+\tilde{A}_a)=\mathscr{D}_1$ on $\mathscr{L}^2(\mathscr{T}_1,dx)$, and the latter has a spectral gap independent of a. Thus under the hypotheses (A1), (A2)′, (A3) the transition probability density $p_a(t;x,y)$ of $\dot{X}(t)$ satisfies (4.19).

We next turn to a special class of diffusions with periodic diffusion coefficients whose drift terms are *not divergence free.* This is the class of one-dimensional multiscale diffusions with periodic coefficients. But, first, some general facts concerning diffusions on the *unit circle* $S^1 = \{x \bmod 1 \colon x \in \mathbb{R}^1\}$ are needed. For detailed derivations see Bhattacharya, Denker and Goswami (1999). Consider the one-dimensional Itô equation

$$Z(t) = Z(0) + \int_0^t \mu(Z(s))\,ds + \int_0^t \sigma(Z(s))\,dB(s), \tag{4.26}$$

where $\mu(\cdot)$, $\sigma(\cdot)$ are continuously differentiable periodic functions with period one, $\sigma^2(x) > 0\ \forall\, x$, $B(t)$ is a standard one-dimensional Brownian motion independent of $Z(0)$. The diffusion $\dot{Z}(t) := Z(t) \bmod 1$ on S^1 has a unique invariant probability with density π given by

$$\pi(x) = d\exp(I(0,x))/\sigma^2(x), \qquad I(0,x) := \int_0^x \frac{2\mu(y)}{\sigma^2(y)}\,dy, \tag{4.27}$$

provided one has

$$\int_0^1 \frac{\mu(y)}{\sigma^2(y)}\,dy = 0. \tag{4.28}$$

If (4.28) does not hold, then

$$\begin{aligned}\pi(x) = \frac{d'\exp(I(0,x))}{\sigma^2(x)}\biggl\{&\frac{\exp(I(0,1))}{\exp(I(0,1)) - 1}\int_0^1 \exp(-I(0,y))\,dy \\ &- \int_0^x \exp(-I(0,y))\,dy\biggr\}.\end{aligned} \tag{4.29}$$

The constant d in (4.27) is the normalizing constant, as is the constant d' in (4.29). The infinitesimal generator A of $\dot{Z}(t)$ on $\mathscr{L}^2(S^1, \pi)$ is $\frac{1}{2}\sigma^2(x)(d^2/dx^2) + \mu(x)(d/dx)$ acting on periodic functions. One can show that A is *self-adjoint*, that is, $A = \tilde{A}$, if and only if (4.28) holds. Write $\tilde{A}$ for the adjoint of A. Then, irrespective of whether A is self-adjoint or not, one can show on direct integration, using integration by parts and periodic boundary conditions [see Bhattacharya, Denker and Goswami (1999)] that

$$\begin{aligned}\langle -f, \tilde{A}f\rangle &= \frac{1}{2}\|\sigma(\cdot)f'\|^2 \qquad \forall\, f \in \mathbf{D}_{\tilde{A}}, \\ \lambda_A &:= \inf\Bigl\{-\langle f, \tilde{A}f\rangle \colon \|f\| = 1,\ f \in 1^{\perp} \cap \mathbf{D}_{\tilde{A}}\Bigr\} \\ &\geq \frac{1}{2M},\end{aligned} \tag{4.30}$$

where

$$\begin{aligned}M &:= \sup\biggl\{\bigl(\sigma^2(y)\pi(y)\bigr)^{-1}\int_y^1 x\pi(x)\,dx \colon 0 \le y < 1\biggr\} \\ &\le \Bigl(\min_y \sigma^2(y)\pi(y)\Bigr)^{-1}\Bigl(\max_y \pi(y)\Bigr)\Big/2.\end{aligned} \tag{4.31}$$

Now consider a general multiscale one-dimensional diffusion with periodic coefficients,

$$
(4.32)\qquad
\begin{aligned}
X(t) &= X(0) + \int_0^t \{b(X(s)) + \beta(X(s)/a)\}\,ds \\
&\qquad + \int_0^t \sigma(X(s))\,dB(s), \\
X(0) &= ax_0.
\end{aligned}
$$

Assume

(4.33)(B1) $b(\cdot)$, $\beta(\cdot)$, $\sigma(\cdot)$ are continuously differentiable and periodic with period one,
(4.33)(B2) $\sigma(x)$ does not vanish for any x,
(4.33)(B3) "a" *is a positive integer.*

Also, without any essential loss of generality, assume

(4.34)(B4) $\int_0^1 (b(x)/\sigma^2(x))\,dx = 0,$

by adding a constant to $\beta(\cdot)$ if necessary. As before, $\dot{X}(t) := X(t) \bmod a$ is a diffusion on the *big circle* $S_a^1 := \{x \bmod a\colon x \in \mathbb{R}^1\}$, which we identify with $[0, a)$ for purposes of integration. Let $\tilde{\pi}_a$ denote the unique invariant probability density of $\dot{X}(t)$. The infinitesimal generator of $\dot{X}(t)$ on $\mathscr{L}^2(S_a^1, \tilde{\pi}_a)$ is

$$
(4.35)\qquad L = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + \{b(x) + \beta(x/a)\}\frac{d}{dx}
$$

acting on periodic functions. The diffusion $Y(t) := X(a^2t)/a$ is governed by the Itô equation

$$
(4.36)\qquad
\begin{aligned}
Y(t) &= Y(0) + \int_0^t a\{b(aY(s)) + \beta(Y(s))\}\,ds + \int_0^t \sigma(aY(s))\,d\overline{B}(s), \\
Y(0) &= x_0,
\end{aligned}
$$

where $\overline{B}(t) := B(a^2t)/a$ is a standard Brownian motion. Also, $\dot{Y}(t) := Y(t) \bmod 1$ is a diffusion on the unit circle S^1 having the invariant probability density π_a related to the invariant density of $\dot{X}(t)$ by $\pi_a(y) = a\tilde{\pi}_a(ay)$. Note that the generator of $\dot{Y}(t)$ is given by

$$
(4.37)\qquad A_a = \frac{1}{2}\sigma^2(ax)\frac{d^2}{dx^2} + a\{b(ax) + \beta(x)\}\frac{d}{dx}.
$$

Assume now that $\beta(\cdot)$ is bounded away from zero. Then, in the presence of (B4) in (4.34), the relation (4.28) does not hold for the drift of $\dot{Y}(t)$. Hence in this case the invariant density π_a is given by (4.29) with

$$
(4.38)\qquad I(0, x) = a\int_0^x \frac{b(ay)}{\sigma^2(ay)}\,dy + a\int_0^x \frac{\beta(y)}{\sigma^2(ay)}\,dy,
$$

and the generator A_a is not self-adjoint. In order to estimate λ_{A_a} using (4.30), (4.31), assume $\beta(x) > 0\ \forall x$. [The case $\beta(x) < 0\ \forall\ x$ is entirely analogous.] Write

$$
\begin{aligned}
&d_1 = \min_y \sigma^2(y), \qquad d_2 = \max_y \sigma^2(y), \qquad \delta = \int_0^1 |b(y)|\,dy,\\
&\beta_* = \min_y \beta(y), \qquad \beta^* = \max_y \beta(y).
\end{aligned}
\tag{4.39}
$$

By direct calculation one may now show [Bhattacharya, Denker and Goswami (1999)]

$$
\begin{aligned}
&\left(\frac{d_1}{2a\beta^*}\right)\exp(-4\delta/d_1)\left(1-\exp\left(\frac{-2a\beta^*}{d_1}\right)\right)\\
&\quad\le \frac{\sigma^2(ax)\pi_a(x)}{d'}\\
&\quad\le \left(\frac{d_2}{2a\beta_*}\right)\left[\exp\left((2\delta/d_1)+\left(1+\frac{d_2}{2a\beta_*}\right)\exp(4\delta/d_1)\right].
\end{aligned}
\tag{4.40}
$$

From (4.40), (4.30), (4.31), one arrives at

$$\lambda_{A_a} \ge d_1 \frac{\min_y \pi_a(y)}{\max_y \pi_a(y)} \ge c_6 > 0,$$

where c_6 is independent of "a." To get an upper bound, let $f(x) = \sin 2\pi x - \int \sin 2\pi(y)\pi_a(y)\,dy$ to get [see (4.30)] $\langle -f, \tilde{A}f\rangle = \langle -f, Af\rangle = \frac{1}{2}\sigma^2\|f'\|^2 \le c_6'\|f\|^2$ for some c_6' independent of a. Thus one obtains

$$c_6 \le \lambda_{A_a} \le c_6'. \tag{4.41}$$

For the generator L of $\dot{X}(t)$ one then has, by the same argument as given in the proof of Corollary 4.4,

$$\frac{c_6}{a^2} \le \lambda_L \le \frac{c_6'}{a^2}. \tag{4.42}$$

Using this together with the Aronson estimate (4.23), and the relation (4.24), as in the proof of Theorem 4.5, one arrives at the following result.

THEOREM 4.6. *Assume* (B1)–(B4) *in* (4.33), (4.34). *In addition, assume* $\beta(x) > 0$ *for all* x. *Then*:

(a) (4.42) *holds, and*

(b) *The* L^1*-distance between the transition probability density* $p_a(t; x, y)$ *of* $\dot{X}(t)$ *and its invariant density* $\tilde{\pi}_a(y)$ *is estimated by*

$$\sup_x \int_{[0,a)} \left|p_a(t; x, y) - \tilde{\pi}_a(y)\right| dy \le c_7 a^{1/2}\exp\{-c_7' t/a^2\}, \tag{4.43}$$

where c_7 *and* c_7' *are positive constants independent of* a.

The next result concerns the self-adjoint case. For this we assume that the diffusion coefficient $\sigma^2(\cdot)$ is a (positive) *constant* $\sigma^2 > 0$, so that (B4) in (4.34) becomes

$$\text{(4.44)(B4)}' \quad \int_0^1 b(x)\,dx = 0.$$

The generator A_a of $\dot{Y}(t)$ is then self-adjoint if and only if

$$\text{(4.45)(B5)} \quad \int_0^1 \beta(x)\,dx = 0.$$

As stated earlier, the invariant probability density π_a of $\dot{Y}(t)$ in this case is given by (4.27), with $I(0,x) = 2a/\sigma^2\{\int_0^x b(ay)\,dy + \int_0^x \beta(y)\,dy\}$. A fairly straightforward calculation [see Bhattacharya, Denker and Goswami (1999)] yields

$$\begin{aligned} &\frac{\exp(-2\delta/\sigma^2)}{d}\pi_a(x) \le \exp\left(\frac{2a\theta^*}{\sigma^2}\right), \\ &\exp\left(\frac{2\delta/\sigma^2)}{d}\right)\pi_a(x) \ge \exp\left(\frac{2a\theta_*}{\sigma^2}\right), \end{aligned} \tag{4.46}$$

where $\delta = \int_0^1 |b(x)|\,dx$ and

$$\theta_* = \min_x \int_0^x \beta(y)\,dy, \qquad \theta^* = \max_x \int_0^x \beta(y)\,dy. \tag{4.47}$$

From (4.46) one gets

$$\frac{\max_x \pi_a(x)}{\min_x \pi_a(x)} \le \exp\left(\frac{4\delta}{\sigma^2}\right)\exp\left(2a(\theta^* - \theta_*)/\sigma^2\right). \tag{4.48}$$

Using this in (4.30) one obtains

$$\lambda_{A_a} \ge \left(\sigma^2 \exp\left(\frac{-4\delta}{\sigma^2}\right)\right)\exp\left(\frac{-2a(\theta^* - \theta_*)}{\sigma^2}\right), \tag{4.49}$$

so that the spectral gap λ_L of the generator L of the diffusion $\dot{X}(t)$ on the big circle $S_a^1 = \{x \bmod a\colon x \in \mathbb{R}^1\}$ is estimated by

$$\lambda_L = \frac{1}{a^2}\lambda_{A_a} \ge \left(\sigma^2\exp\left(\frac{-4\delta}{\sigma^2}\right)\right)\frac{1}{a^2}\exp(\frac{-2a(\theta^* - \theta_*)}{\sigma^2}\Bigg). \tag{4.50}$$

Proceeding as in the proof of Theorem 4.5, or Theorem 4.6, one arrives at the following estimate of the speed of convergence to equilibrium in this case.

THEOREM 4.7. *In the self-adjoint case* (*B*4)′, (*B*5) *with constant* $\sigma^2 > 0$, *the* L^1-*distance between the transition probability density of* $\dot{X}(t)$ *and its equilibrium density* $\tilde{\pi}_a$ *is estimated by*

$$\begin{aligned} &\sup_x \int_{[0,a)} \left|p_a(t;x,y) - \tilde{\pi}_a(y)\right| dy \\ &\qquad \le c_8 a^{1/2}\exp\{c_9 a/2\}\exp\{-(c_8'/a^2)\exp(-c_9 a)t\} \end{aligned} \tag{4.51}$$

where c_8, c_8' $c_9 \equiv 2(\theta^* - \theta_*)/\sigma^2$ *are positive constants independent of* "*a*."

REMARK 4.7.1. The speed of convergence to equilibrium as estimated in Theorem 4.7 is exceedingly slow and, going by it, the process may take times $t \gg (a^2 \log a) \exp\{c_9 a\}$ to be close to equilibrium. This is in contrast to the nonself-adjoint case considered in Theorem 4.6, where for times $t \gg a^2 \log a$, the process is near equilibrium. The estimate (4.51) concerns the "worst case" scenario such as holds under the hypothesis of part (b) of Theorem 4.9 below. On the other hand, under the hypothesis of part (a) of Theorem 4.9, the speed of convergence is shown to be as fast as in the case of Theorem 4.6.

An important difference in the asymptotic behavior between the two classes of diffusions considered in Theorems 4.6 and 4.7 is provided by the following result.

PROPOSITION 4.8. (a) *Under the hypothesis of Theorem* 4.6, *the invariant probability density* π_a *of the diffusion* $\dot{Y}(t) = \dot{X}(a^2 t)/a$ *on* S^1 *is bounded away from zero uniformly in* a.

(*b*) *Assume the hypothesis of Theorem* 4.7. *If the "potential function"* $\psi(x) := \int_0^x \beta(y)\,dy$ *has its maximum attained at a single point* x^*, *then the invariant probability* $\pi_a(x)\,dx$ *converges weakly to the point mass* δ_{x^*} *as* $a \to \infty$. *More generally*, *if the maximum of* ψ *is attained at a finite number of points*, *then all weak limit points of* $\pi_a(x)\,dx$ *have support contained in this finite set*.

PROOF. Part (a) follows from the estimate $\min_y \pi_a(y)/\max_y \pi_a(y) \geq c_6/d_1$ [see (4.41)]. To prove part (b), let $x_1, x_2, \ldots, x_m$ be the distinct points in $[0, 1)$ where the maximum of ψ is attained. Since $|a \int_0^x b(ay)\,dy| = |\int_0^{ax} b(y)\,dy| = |\int_{[ax]}^{ax} b(y)\,dy| \leq \delta \equiv \int_0^1 |b(y)|\,dy$ [in view of (4.44)], it is simple to check that $\pi_a(x)/\max_y \pi_a(y) \to 0$ if $x \notin \{x_1, x_2, \ldots, x_m\}$. It follows that for any $\varepsilon > 0$, however small, the π_a-probability of the ε–neighborhood of the finite set $\{x_1, x_2, \ldots, x_m\}$ goes to one as $a \to \infty$. $\square$

The next result provides a dichotomy of the class of time-reversible diffusions on the big circle into (1) those for which the speed of convergence to equilibrium is the same as in the nonself-adjoint case considered in Theorem 4.6 and (2) those for which the convergence is exceedingly slow, requiring times $t \gg e^{ca}$. For this we need a result of Holley, Kusuoka and Stroock (1989) specialized to the circle. Consider the self-adjoint case with $b(\cdot) \equiv 0$, that is,

$$A_a = \frac{1}{2}\sigma^2 \frac{d^2}{dy^2} + a\beta(y)\frac{d}{dy}, \qquad \int_0^1 \beta(y)\,dy = 0. \tag{4.52}$$

In this case, according to Theorem 1.14 in Holley, Kusuoka and Stroock (1989), there exist constants $c^{(1)} > 0$, $c^{(2)} \geq 0$, independent of a, such that

$$c^{(1)} a^{-2} \exp\{-c^{(2)} a\} \leq \lambda_{A_a} \leq c^{(1)} a^6 \exp\{-c^{(2)} a\}. \tag{4.53}$$

The constant $c^{(2)}$ is computed as follows. Let $U(x) = (\theta^* - \psi(x))/\sigma^2$, where θ^* is given by (4.47) and $\psi(x) = \int_0^x \beta(y)\,dy$. For any given pair of points x, y in S^1 and a continuous curve γ joining x and y, let $H_\gamma(x, y)$ denote the maximum value of U on (the image of) γ. Define $H(x, y)$ to be the infimum over all such γ. Then

$$c^{(2)} = \sup_{x,y}\{H(x, y) - U(x) - U(y)\}. \tag{4.54}$$

THEOREM 4.9. *Consider the self-adjoint case* (B4)′, (B5) *with* $\sigma^2 > 0$, *and assume that the number of zeros of* β *on* $[0, 1)$ *is finite.*

(a) *If* $\psi(x) \equiv \int_0^x \beta(y)\,dy$ *has a unique maximum, then there exists a positive constant* $c^{(3)}$ *independent of* a *such that*

$$\lambda_L \geq \frac{c^{(3)}}{a^2}. \tag{4.55}$$

(b) *If* ψ *has more than one maximum then there exist positive constants* $c^{(2)}$, $c^{(4)}$, $c^{(5)}$ *independent of* a, *with* $c^{(2)}$ *as in* (4.54), *such that*

$$c^{(4)}a^{-5}\exp\{-c^{(2)}a\} \leq \lambda_L \leq c^{(5)}a^4\exp\{-c^{(2)}a\}. \tag{4.56}$$

PROOF. (a) Consider the generator A_a of $\dot{Y}(t)$ on $\mathscr{L}^2(S^1, \pi_a)$. Let π denote the probability density on $[0, 1]$ given by

$$\pi(x) = d'\exp\left\{\frac{2a}{\sigma^2}\psi(x)\right\}, \qquad 0 \leq x \leq 1, \tag{4.57}$$

d' being the normalizing constant. Since $|\int_0^x ab(ay)\,dy| \leq \delta \equiv \int_0^1 |b(y)|\,dy$ [see (4.44)] one has

$$\pi_a(x) \leq \exp(4\delta/\sigma^2)\pi(x), \qquad x \in [0, 1]. \tag{4.58}$$

Now let $f \in \mathbf{D}_{\tilde{A}_a} \cap 1^{\perp}$. Then, writing $c_{11} = \exp(8\delta/\sigma^2)$,

$$\begin{aligned}
\|f\|^2 &= \tfrac{1}{2}\int_0^1\int_0^1 (f(x) - f(y))^2\pi_a(x)\pi_a(y)\,dx\,dy \\
&\leq \tfrac{1}{2}c_{11}\int_0^1\int_0^1 (f(x) - f(y))^2\pi(x)\pi(y)\,dx\,dy \\
&= \tfrac{1}{2}c_{11}\left(\iint_{\{x<y\}} + \iint_{\{y<x\}}\right)\bigl(f(x) - f(y)\bigr)^2\pi(x)\pi(y)\,dx\,dy \\
&= c_{11}\iint_{\{x<y\}}(f(x) - f(y))^2\pi(x)\pi(y)\,dx\,dy \qquad (4.59)\\
&= c_{11}\int_0^1\int_0^y (f(x) - f(y))^2\pi(x)\pi(y)\,dx\,dy \\
&= c_{11}\int_0^1\int_0^y\left(\int_x^y f'(z)\,dz\right)^2\pi(x)\pi(y)\,dx\,dy
\end{aligned}$$

$$\le c_{11}\int_0^1\int_0^y (y-x)\left\{\int_x^y (f'(z))^2\,dz\right\}\pi(x)\pi(y)\,dx\,dy$$
$$= c_{11}\int_0^1 (f'(z))^2\left[\int_z^1\left\{\int_0^z (y-x)\pi(x)\,dx\right\}\pi(y)\,dy\right]dz.$$

By translation, if necessary, we may assume that the minimum of ψ is at 0, and the maximum is at x^*. Then $\psi(x)$ increases from $x = 0$ to $x = x^*$ and decreases from $x = x^*$ to $x = 1$. One thus has, for $z \le x^*$,

$$\begin{aligned}
&\int_z^1\left\{\int_0^z (y-x)\pi(x)\,dx\right\}\pi(y)\,dy\\
&\qquad\le \int_z^1\left\{\int_0^z (y-x)\pi(z)\,dx\right\}\pi(y)\,dy\\
&\qquad= \pi(z)\int_z^1\left\{\int_0^z (y-x)\,dx\right\}\pi(y)\,dy\\
&\qquad= \pi(z)\int_z^1\left(yz-\frac{z^2}{2}\right)\pi(y)\,dy\\
&\qquad\le \pi(z)z.
\end{aligned}\tag{4.60}$$

For $z > x^*$,

$$\begin{aligned}
&\int_z^1\left\{\int_0^z (y-x)\pi(x)\,dx\right\}\pi(y)\,dy\\
&\quad= \int_0^z\left\{\int_z^1 (y-x)\pi(y)\,dy\right\}\pi(x)\,dx\\
&\quad\le \int_0^z\left\{\int_z^1 (y-x)\pi(z)\,dy\right\}\pi(x)\,dx\\
&\quad= \pi(z)\int_0^z\left\{\frac{1-z^2}{2}-x(1-z)\right\}\pi(x)\,dx \le \pi(z)\left(\frac{1-z^2}{2}\right).
\end{aligned}\tag{4.61}$$

Using (4.60), (4.61) in (4.59) we get [see (4.30)]

$$\|f\|^2 \le c_{11}\int_0^1 (f'(z))^2\pi(z)\,dz = \frac{c_{11}}{\sigma^2}\langle -f, \tilde{A}f\rangle, \tag{4.62}$$

so that $\lambda_{A_a} \ge \sigma^2/c_{11}$ and, as a consequence, the spectral gap λ_L of the generator L of $\dot{X}(t)$ satisfies

$$L = \frac{1}{a^2}\lambda_{A_a} \ge \frac{c_{12}}{a^2}, \tag{4.63}$$

where $c_{12} \equiv \sigma^2/c_{11}$ does not depend on a.

For part (b), first make the additional assumption $b(\cdot) \equiv 0$. Let x be a point where ψ attains its absolute maximum value θ^* and let y be another maximum of ψ. Then $U(x) = 0$ and y is a minimum of U, $U(y) \geq 0$. Every continuous curve γ joining x and y contains (in its range) a maximum of U, that is, $H_\gamma(x, y) \geq U(z) - U(x) - U(y) \equiv U(z) - U(y) > 0$, where $U(z) = \min\{U(z_1), U(z_2)\}$, and z_1, z_2 are the two points on the two arcs joining x and y at which U attains its maximum values. Hence $c^{(2)}$ defined by (4.54) is positive, and (4.56) is just (4.53) in this case. For the general case under (b), one shows, as in part (a) above, that the ratio of the invariant density to that with $b(\cdot) \equiv 0$ is bounded away from zero and infinity. Hence, using (4.30), one derives (4.53) with $c^{(1)}$ replaced by a smaller constant $c^{(4)}$ on the left and by a larger constant $c^{(5)}$ on the right. Since $\lambda_L = \lambda_{A_a}/a^2$, (4.56) follows. □

Using Lemma 4.2 and an estimate of Aronson (1967), exactly as in the proof of Theorem 4.6, we derive the following theorem.

THEOREM 4.10. *In addition to the hypothesis of Theorem* 4.7, *assume that the potential function* $\psi(x) = \int_0^x \beta(y)\, dy$ *on* $[0, 1)$ *has a unique maximum and a unique minimum. Then*

$$\sup_x \int_{[0,a)} \left| p_a(t; x, y) - \tilde{\pi}_a(y) \right| dy \leq c_{10} a^{1/2} \exp(-c'_{10} t/a^2), \tag{4.64}$$

where c_{10} *and* c'_{10} *are independent of* a.

EXAMPLE 4.10.1. Let $b(\cdot)$ be arbitrary (periodic and differentiable) satisfying $\int_0^1 b(y)\, dy = 0$. Let $\beta(x) = \pi \cos \pi x$, so that $\psi(x) = \sin \pi x$. Then, on the unit circle, the *flow* $dx(t)/dt = \beta(x(t))$ has one *stable equilibrium* $x = \frac{1}{2}$, where ψ is maximum, and one *unstable equilibrium* $x = 0$, where ψ is minimum. Thus Theorem 4.9 applies. One may expect a *relatively fast convergence* to equilibrium here for $\dot{Y}(t)$, since from every initial point $x \neq 0$ the flow approaches the stable equilibrium fast.

EXAMPLE 4.10.2. Let $b(\cdot)$ be arbitrary, as above, and $\beta(x) = 4\pi \cos 4\pi x$. Then $\psi(x) = \sin 4\pi x$ attains its maximum value at $x = \frac{1}{8}$ and $x = \frac{5}{8}$; these are the stable equilibria of the flow $dx(t)/dt = \beta(x(t))$. The minimum value of $\psi(x)$ is attained at $x = \frac{3}{8}$ and $x = \frac{7}{8}$; these are the unstable equilibria of the flow. In this case one would expect a relatively slow convergence to equilibrium of $\dot{Y}(t)$ starting from any point x, and Theorem 4.9(b) applies. The spectral gap in this case is exponentially small, namely, $O(e^{-\alpha a})$, for some $\alpha > 0$ which does not depend on a, and a slow convergence to equilibrium such as provided for by Theorem 4.7 results.

5. Final phase of asymptotics.

5.1. *The divergence-free case.* Consider again the multiscale diffusion on $\mathbb{R}^k$ with periodic coefficients as given in (3.39), namely,

$$X(t) = X(0) + \int_0^t \{b(X(s)) + \beta(X(s)/a)\}\, ds + \sigma B(t), \tag{5.1}$$

and its scaled version $Y(t) = X(a^2t)/a$ satisfying the Itô equation (4.36). Recall the diffusion $\dot{X}(t) = X(t) \bmod a$ on the big torus $\mathscr{T}_a$ and the diffusion $\dot{Y}(t) = Y(t) \bmod 1$ on the unit torus $\mathscr{T}_1$. We first derive a simple consequence of Theorem 4.5. To state it, write E_x for expectation under $\dot{X}(0) = x$ or $\dot{Y}(0) = x$, as the case may be, and E as the expectation under equilibrium, that is, the invariant distribution. Also, "cov_x" denotes covariance under $\dot{X}(0) = x$ [or $\dot{Y}(0) = x$], while "cov" denotes covariance under equilibrium. As before, $\|f\|_\infty$ denotes the supremum of $|f(x)|$ over all x for some measurable real-valued function f. The constants c_i c_i' below are positive and independent of a.

PROPOSITION 5.1. *Assume* (A1)–(A3) *in* (3.40). *There exist positive constants* c_i, c_i' $(i = 13, 14)$ *not depending on "a" such that for all bounded measurable* f, g *on* $\mathscr{T}_a$, *one has*

$$\begin{aligned} &\big|E_x f(\dot{X}(t)) - Ef(\dot{X}(t))\big| \\ &\qquad \le c_{13} a^{k/2} \|f\|_\infty \exp\{-c_{13}' t/a^2\}, \qquad t \ge 0, \\ &\big|\mathrm{cov}_x\{f(\dot{X}(s)), g(\dot{X}(t))\}\big| \\ &\qquad \le c_{14} a^{k/2} \|f\|_\infty \|g\|_\infty \exp\{-c_{13}'(t-s)/a^2\}, \qquad 0 \le s \le t. \end{aligned} \tag{5.2}$$

Similarly, for all bounded measurable f, g *on* $\mathscr{T}_1$, *one has*

$$\begin{aligned} &\big|E_y f(\dot{Y}(t)) - Ef(\dot{Y}(t))\big| \\ &\qquad \le c_{13} a^{k/2} \|f\|_\infty \exp\{-c_{13}' t\}, \qquad t \ge 0, \\ &\big|\mathrm{cov}_y\{f(\dot{Y}(s)), g(\dot{Y}(t))\}\big| \\ &\qquad \le c_{14} a^{k/2} \|f\|_\infty \|g\|_\infty \exp\{-c_{13}'(t-s)\}, \qquad 0 \le s \le t. \end{aligned} \tag{5.3}$$

PROOF. The first relation in (5.2) is an immediate consequence of Theorem 4.5 with $c_{13} = c_5$ and $c_{13}' = 2\pi^2\alpha_1$. For the second relation, use conditioning given $\sigma\{\dot{X}(u)\colon 0 \le u \le s\}$ to write

$$\begin{aligned} \mathrm{cov}_x\{f(\dot{X}(s)), g(\dot{X}(t))\} &= E_x\big[\{f(\dot{X}(s)) - E_x f(\dot{X}(s))\} \\ &\qquad \times \big\{E_z g(\dot{X}(t-s))_{z=\dot{X}(s)} - E_x g(\dot{X}(t))\big\}\big]. \end{aligned} \tag{5.4}$$

Applying the first inequality in (5.2) to the second factor in (5.4), one gets the second relation in (5.2). Relations (5.3) follow from those in (5.2), noting that, for functions f, g on $\mathscr{T}_1$, $f(\dot{Y}(t)) = f(\dot{X}(a^2t)/a)$, $g(\dot{Y}(t)) = g(\dot{X}(a^2t)/a)$ so that (5.2) may be applied to functions $x \to f(x/a)$, $g(x/a)$ with times a^2t, a^2s in place of t, s. $\square$

An immediate consequence of (5.2) and (5.3) is

$$
\begin{aligned}
&\big|\mathrm{cov}\{f(\dot X(s)), g(\dot X(t))\}\big| \\
&\qquad \le c'_{14} a^{k/2} \|f\|_\infty \|g\|_\infty \exp\{-c'_{13}(t-s)/a^2\}, \\
&\big|\mathrm{cov}\{f(\dot Y(s)), g(\dot Y(t))\}\big| \\
&\qquad \le c'_{14} a^{k/2} \|f\|_\infty \|g\|_\infty \exp\{-c'_{13}(t-s)\}, \qquad 0 \le s \le t.
\end{aligned}
\tag{5.5}
$$

For this, simply replace cov_x, E_x in (5.4) by cov and E, respectively.

We are now ready to prove one of the main results of this article. Below, $\to_{\mathscr{L}}$ denotes convergence in law or distribution.

THEOREM 5.2. *Assume* (A1)–(A3) *in* (3.40). *Also assume that* (A4)j *in* (3.68) *holds for $1 \le j \le k_1$ for some $k_1 \le k$. If, in addition, the assumption*

(5.6)(A6) $(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}$, $1 \le j \le k_1$, *are linearly independent elements of H^1, holds, then for $t \gg a^2(\log a)^2$, that is, as*

$$
a \to \infty, \qquad \frac{t}{a^2(\log a)^2} \to \infty, \tag{5.7}
$$

one has

$$
\left\{\frac{1}{a\sqrt{t}}(X_j(t) - X_j(0) - t(\overline{b}_j + \overline{\beta}_j))\colon 1 \le j \le k_1\right\} \to_{\mathscr{L}} \mathscr{N}(0, \Sigma_1), \tag{5.8}
$$

no matter what the initial state $X(0)$ may be. Here $\Sigma_1 = ((\overline{\sigma}_{ij}))$ is given by

$$
\overline{\sigma}_{ij} = 2\langle(\mathscr{D}^{-1}(\beta_i - \overline{\beta}_i))_{\underline{N}}, (\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}}\rangle_1, \qquad 1 \le i, j \le k_1. \tag{5.9}
$$

PROOF. One needs to prove that an arbitrary non-zero linear combination of the random variables in (5.8), with coefficients ξ_j, say, converges in distribution to a normal law $\mathscr{N}(0, \gamma)$ where $\gamma = \sum_{i,j=1}^{k_1} \overline{\sigma}_{ij}\xi_i\xi_j$. To avoid a somewhat messy notation, we will prove the result for the case $\xi_j = 1$, $\xi_i = 0$ for $i \ne j$. The proof in the general case is entirely analogous. We will prove that for times t satisfying (5.7),

$$
\frac{1}{a\sqrt{t}}\big(X_j(t) - X_j(0) - t(\overline{b}_j + \overline{\beta}_j)\big) \to_{\mathscr{L}} \mathscr{N}(0, \overline{\sigma}_{jj}), \tag{5.10}
$$

under the assumptions (A1)–(A4) and $(\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j))_{\underline{N}} \ne 0$. This assertion is equivalent to

$$
\frac{1}{a\sqrt{t}}\big(Y_j(t) - Y_j(0) - at(\overline{b}_j + \overline{\beta}_j)\big) \to_{\mathscr{L}} \mathscr{N}(0, \overline{\sigma}_{jj}), \tag{5.11}
$$

under the same assumptions, but for times $t \gg (\log a)^2$, that is,

$$
a \to \infty, \qquad \frac{t}{(\log a)^2} \to \infty. \tag{5.12}
$$

Recalling that the left side of (5.11) equals [see (3.43)]

$$\frac{1}{\sqrt{t}}\int_0^t \{b_j(a\dot{Y}(s)) + \beta_j(\dot{Y}(s)) - \overline{b}_j - \overline{\beta}_j\}\,ds + \frac{(\sigma\overline{B}(t))_j}{a\sqrt{t}}, \tag{5.13}$$

where $(\sigma\overline{B}(t))_j$ is the jth component of the vector $\sigma\overline{B}(t)$, (5.11) is equivalent to

$$\frac{1}{\sqrt{t}}\int_0^t \{b_j(a\dot{Y}(s)) + \beta_j(\dot{Y}(s)) - \overline{b}_j - \overline{\beta}_j\}ds \to_{\mathscr{L}} \mathscr{N}(0, \overline{\sigma}_{jj}). \tag{5.14}$$

By Itô's lemma, the left side of (5.14) equals [see (3.45), (3.46)]

$$\frac{1}{\sqrt{t}}\left\{g_j(\dot{Y}(t)) - g_j(\dot{Y}(0))\right\} - \frac{1}{\sqrt{t}}\int_0^t \operatorname{grad} g_j(\dot{Y}(s))\sigma\, d\overline{B}(s). \tag{5.15}$$

Therefore, one has

$$\begin{aligned} &\frac{1}{\sqrt{t}}\int_0^t \{b_j(a\dot{Y}(s)) + \beta_j(\dot{Y}(s)) - \overline{b}_j - \overline{\beta}_j\}\,ds \\ &\qquad - \frac{1}{\sqrt{t}}\{g_j(\dot{Y}(t)) - g_j(\dot{Y}(0))\} \\ &\quad = -\frac{1}{\sqrt{t}}\int_0^t \operatorname{grad} g_j(\dot{Y}(s))\sigma\, d\overline{B}(s). \end{aligned} \tag{5.16}$$

Assume first that $\dot{Y}(0)$ has the *uniform* (equilibrium) *distribution*. Then, by Lemma 3.5,

$$E\left(\frac{1}{\sqrt{t}}\{g_j(\dot{Y}(t)) - g_j(\dot{Y}(0))\}\right)^2 \le \frac{2}{t}\|g_j\|_0^2 \to 0 \quad \text{as } t \to \infty. \tag{5.17}$$

Letting $t = \varphi(a) \gg (\log a)^2$, $\varphi(a)$ integral, one may express the left side of (5.14) as

$$\begin{aligned} &\sum_{r=1}^{\varphi(a)} V_r, \\ &V_r := \frac{1}{\sqrt{\varphi(a)}}\int_{r-1}^r \{b_j(a\dot{Y}(s)) + \beta_j(\dot{Y}(s)) - \overline{b}_j - \overline{\beta}_j\}\,ds \ (1 \le r \le \varphi(a)). \end{aligned} \tag{5.18}$$

In view of (5.16), (5.17) and Theorem 3.7, one has

$$EV_r = 0, \quad E\left(\sum_{r=1}^{\varphi(a)} V_r\right)^2 \to \overline{\sigma}_{jj} \qquad \text{as } a \to \infty. \tag{5.19}$$

We will prove the asymptotic normality of $\sum_{r=1}^{\varphi(a)} V_r$ by representing it approximately as the sum of a number of nearly independent block sums. For this purpose, define

$$\begin{aligned} &\delta \equiv \delta(a) := \varphi(a)/(\log a)^2, \qquad \eta \equiv \eta(a) := [\delta^{1/8}\log a], \\ &\psi \equiv \psi(a) := [\delta^{3/8}\log a], \qquad m \equiv m(a) := \left[\frac{\varphi(a)}{\eta + \psi}\right], \end{aligned} \tag{5.20}$$

where $[z]$ denotes the integer part of z. Consider the "big" block sums

$$Z_1 = \sum_{r=1}^{\psi(a)} V_r, \qquad Z_2 = \sum_{r=1}^{\psi(a)} V_{r+\psi+\eta}, \dots, \qquad Z_m = \sum_{r=1}^{\psi(a)} V_{r+(m-1)(\psi+\eta)} \tag{5.21}$$

and the "little" block sums

$$\begin{aligned} \xi_1 &= \sum_{r=1}^{\eta(a)} V_{r+\psi}, \\ \xi_2 &= \sum_{r=1}^{\eta(a)} V_{r+2\psi+\eta}, \dots, \xi_m = \sum_{r=1}^{\eta(a)} V_{r+m(\psi+\eta)-\eta}. \end{aligned} \tag{5.22}$$

Then

$$\sum_{r=1}^{\varphi(a)} V_r \simeq \sum_{r=1}^{m(a)} Z_r + \sum_{r=1}^{m(a)} \xi_r. \tag{5.23}$$

To verify this, note that the right side of (5.23) is missing at most $\psi + \eta$ terms V_r from the left. By applying the convergence in (5.19), but with $\psi + \eta$ in place of φ, it follows that the expected value of the squared sum of the missing terms is no more than $O((\psi + \eta)/\varphi(a)) \to 0$. Next, by a similar argument,

$$E\xi_r^2 \le c_{15}\eta/\varphi(a), \qquad \sum_{r=1}^{m} E\xi_r^2 \le c'_{15} m\eta/\varphi(a) \to 0. \tag{5.24}$$

Also, for $r' \ge 1$,

$$\begin{aligned} E\xi_r\xi_{r+r'} &= \sum_{i=1}^{\eta} \sum_{i'=1}^{\eta} E\bigl(V_{i+\psi} V_{i'+(r'+1)(\psi+\eta)-\eta}\bigr) \\ &= \frac{1}{\varphi(a)} \sum_{i,\, i'=1}^{\eta} \int_0^1 \bigl\langle h, T_{i'-i-1+r'(\psi+\eta)+s} f\bigr\rangle\, ds, \end{aligned} \tag{5.25}$$

where $h(y) := b_j(ay) + \beta_j(y) - \overline{b}_j - \overline{\beta}_j$, $f(y) = E_y \int_0^1 h(\dot{Y}(s))\, ds$, and T_u is the transition operator of $\dot{Y}$ $(u \ge 0)$. By Proposition 5.1, the integral on the right in (5.25) is bounded in magnitude by $c_{16}\|h\|_\infty^2 a^{k/2} \exp\{-c'_{13} r'\psi\}$, so that

$$\begin{aligned} \bigl|E\xi_r\xi_{r+r'}\bigr| &\le c'_{16} \frac{\eta^2 a^{k/2}}{\varphi(a)} \exp\{-c'_{13} r'\psi\}, \\ \sum_{r=1}^{m} \sum_{r'=1}^{m-r} \bigl|E\xi_r\xi_{r+r'}\bigr| &\le c_{17} \frac{m\eta^2 a^{k/2}}{\varphi(a)} \exp\{-c'_{13}\psi\} \to 0 \quad \text{as } a \to \infty. \end{aligned} \tag{5.26}$$

Thus $E(\sum_{r=1}^{m} \xi_r)^2 \to 0$ as $a \to \infty$, and we get from (5.23) the relation

$$\sum_{r=1}^{\varphi(a)} V_r \simeq \sum_{r=1}^{m} Z_r. \tag{5.27}$$

We next show that the characteristic function of the right side of (5.27) is asymptotically the same as that of the sum of m i.i.d. random variables, each having the same distribution as Z_1. For this write, for any fixed $\xi \in \mathbb{R}^1$, $f(y) := E[\exp\{i\xi Z_1\} \mid \dot{Y}(0) = y]$, to derive the following approximation using Proposition 5.1:

$$
\begin{aligned}
&\left| E\left(\exp\left\{ i\xi \sum_{r=1}^{m} Z_r \right\}\right)\right. \\
&\qquad \left. - E\left(\exp\left\{ i\xi \sum_{r=1}^{m-1} Z_r \right\}\right) E\left(\exp\{i\xi Z_m\}\right) \right| \\
&\quad = \left| E\left[\exp\left\{ i\xi \sum_{r=1}^{m-1} Z_r \right\}\left(T_\eta f(\dot{Y}(r')) - \overline{f}\right)\right]\right| \\
&\qquad \times \left(r' := (m-1)(\psi+\eta) - \eta\right) \\
&\quad \le c_{18} a^{k/2} \exp\{-c'_{13}\eta\}.
\end{aligned} \tag{5.28}
$$

Telescoping this process one arrives at

$$
\begin{aligned}
&\left| E\left(\exp\left\{ i\xi \sum_{r=1}^{m} Z_r \right\}\right) - \prod_{r=1}^{m} E\left(\exp\{i\xi Z_r\}\right) \right| \\
&\qquad \le c_{18} m a^{k/2} \exp\{-c'_{13}\eta\} \to 0 \quad \text{as } a \to \infty.
\end{aligned} \tag{5.29}
$$

We will now verify Lindeberg's condition for the sum $m = m(a)$ i.i.d. random variables Z_r. Note that, for each $\varepsilon > 0$,

$$
\sum_{r=1}^{m} E\left(Z_r^2 \, \mathbb{1}_{\{|Z_r| > \varepsilon\}}\right) = m E Z_1^2 \, \mathbb{1}_{\{|Z_1| > \varepsilon\}} = 0 \tag{5.30}
$$

for all sufficiently large a, since $|Z_1| \le c\psi/\sqrt{\varphi(a)} \to 0$ as $a \to \infty$. This proves (5.11) under the invariant initial distribution [of $\dot{Y}(0)$].

It remains to consider the case of an arbitrary initial distribution [of $\dot{Y}(0)$]. Let $t = \varphi(a) \gg (\log a)^2$, $s = \psi(a) = \delta^{3/8} \log a$ as in (5.20). Write

$$
\begin{aligned}
\frac{Y_j(t) - t(\overline{b}_j + \overline{\beta}_j)}{a\sqrt{t}} &= \frac{Y_j(s) - s(\overline{b}_j + \overline{\beta}_j)}{a\sqrt{t}} \\
&\quad + \frac{Y_j(t) - Y_j(s) - (t-s)(\overline{b}_j + \overline{\beta}_j)}{a\sqrt{t}}.
\end{aligned} \tag{5.31}
$$

Using the integral representation of $Y(\cdot)$ [see (3.43), (5.13)], it follows that

$$
E\left(\frac{Y_j(s) - s(\overline{b}_j + \overline{\beta}_j)}{a\sqrt{t}}\right)^2 \le \frac{c_{19} s^2}{t} \to 0 \quad \text{as } a \to \infty. \tag{5.32}
$$

Now the conditional distribution of $Y(t) - Y(s)$, given $\{Y(u)\colon 0 \le u \le s\}$, depends only on $\dot{Y}(s)$ and is, in fact, the same as the distribution of $Y(t-s) - z$ with an initial state $z = \dot{Y}(s)$. Therefore, by Theorem 4.5 and Proposition 4.3

(see Remark 4.5.1), the total variation distance between the distribution of $Y(t) - Y(s)$ under an arbitrary $\dot{Y}(0)$ and that under a uniformly distributed $\dot{Y}(0)$ goes to zero as $a \to \infty$. Using this fact and (5.32) in (5.31), it follows that the left side of (5.31) converges in law to $\mathscr{N}(0, \overline{\sigma}_{jj})$ as $a \to \infty$, no matter what the initial distribution may be. $\square$

REMARK 5.2.1. Theorem 5.2 may be strengthened to the following functional form under the given hypothesis: *for any given sequence of integers* $\varphi(a)$ *such that*

$$\frac{\varphi(a)}{a^2(\log a)^2} \to \infty \quad \textit{as } a \to \infty,$$

one has

$$\frac{1}{a\sqrt{\varphi(a)}}\left\{X_j(\varphi(a)t) - X_j(0) - \varphi(a)t(\overline{b}_j + \overline{\beta}_j)\colon 1 \le j \le k_1\right\}_{t\ge 0}$$

$$\to_{\mathscr{L}} \{W(t)\}_{t\ge 0} \quad \textit{as } a \to \infty,$$

where $\{W(t)\}_{t\ge 0}$ *is a Brownian motion on* $\mathscr{C}([0,\infty) \to \mathbb{R}^k)$ *having the dispersion coefficients* (5.9). To prove this, one first uses the negligibility of ξ_r's to reduce the problem to that of the asymptotic distribution of the polygonal process corresponding to the partial sums of Z_r $(r \ge 1)$. We then show that the total variation distance between the distribution of $(Z_1, Z_2, \dots, Z_m)$ under equilibrium and the product measure G_a^m, where G_a is the distribution of Z_1, goes to zero as $m \to \infty$. To establish the latter, consider a real-valued bounded measurable function f on $\mathbb{R}^m$ and show, by using the Markov property, Proposition 5.1, and telescoping [as in (5.28), (5.29)], that

$$\left|Ef(Z_1, Z_2, \dots, Z_m) - \int f\,dG_a^m\right| \le c'_{18} m \|f\|_\infty a^{k/2} \exp\{-c'_{13}\eta\}.$$

Hence the proof of the functional limit theorem stated above boils down to that for triangular arrays of i.i.d. summands, making use of Lindeberg's condition (5.30) [Billingsley (1968), page 77]. The argument when $\dot{X}(0)$ or $\dot{Y}(0)$ is not in equilibrium remains the same as given at the end of the proof of Theorem 5.2.

The next result complements Theorem 5.2 by analyzing the case where $\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)$ belongs to the range of $\overline{S} = \mathscr{D}^{-1}(\overline{b} + \beta)\cdot\nabla$ for certain j's. Dramatic differences in the growth of dispersion in the two cases (see Theorems 3.7, 3.8) lead to significantly different scalings in Theorems 5.2 and 5.3.

For the statement of the following theorem, recall that $K = ((K_{jj'}))$ where $K_{jj'} = 2a^2\langle g_j, g_{j'}\rangle_1 + D_{jj'}$ [see (3.48)]. For a set of k_2 coordinates, $1 \le j \le k_2$, let K_2 denote the $k_2 \times k_2$ submatrix of K comprising elements belonging to the first k_2 rows and to the first k_2 columns of K. Also write I_{k_2} for the $k_2 \times k_2$ identity matrix.

THEOREM 5.3. *In addition to* (A1)–(A3) *in* (3.40), *assume that* (A5)j *in* (3.88) *holds for* $1 \le j \le k_2$ *and that the functions* p_j *in* H^1 *satisfying*

$(\overline{b} + \beta) \cdot \nabla p_j = \beta_j - \overline{\beta}_j$, $1 \le j \le k_2$, *are linearly independent. Then for* $t \gg a^4(\log a)^2$ *one has*

$$\frac{1}{\sqrt{t}} K_2^{-1/2}\big(\{X_j(t) - X_j(0) - t(\overline{b}_j + \overline{\beta}_j)\}_{1 \le j \le k_2}\big) \to_{\mathscr{L}} \mathscr{N}(0, I_{k_2}) \tag{5.33}$$

as $a \to \infty$, *whatever be the initial distribution.*

PROOF. As in the proof of Theorem 5.2, we will prove that for $t \gg a^4(\log a)^2$ one has

$$\frac{1}{\sqrt{t K_{jj}}}\big(X_j(t) - X_j(0) - t(\overline{b}_j + \overline{\beta}_j)\big) \to_{\mathscr{L}} \mathscr{N}(0, 1), \qquad 1 \le j \le k_2, \tag{5.34}$$

as $a \to \infty$. The proof for an arbitrary linear combination of X'_js $(1 \le j \le k_2)$ is analogous.

First assume $\dot{X}(0)$ has the uniform (equilibrium) distribution. Let w_j be the solution in H^1 of the equation

$$Lw_j = b_j + \beta_j(\cdot/a) - \overline{b}_j - \overline{\beta}_j, \tag{5.35}$$

where L is the generator of $\dot{X}(t)$ on $\mathscr{L}^2(\mathscr{T}_a, a^{-k}dx)$. Then $w_j(x) = a^2 g_j(x/a)$, g_j being as in (3.45). By Itô's lemma, with $t = \varphi(a) \gg a^4(\log a)^2$, $\varphi(a)$ integral,

$$\begin{aligned}
&\frac{1}{\sqrt{tK_{jj}}}\big(X_j(t) - X_j(0) - t(\overline{b}_j + \overline{\beta}_j)\big)\\
&\quad= \frac{1}{\sqrt{tK_{jj}}}\bigg[w_j(\dot{X}(t)) - w_j(\dot{X}(0))\\
&\qquad\qquad - \int_0^t \operatorname{grad} w_j(\dot{X}(s))\sigma\, dB(s) + (\sigma B(t))_j\bigg]\\
&\quad= \sum_{r=1}^{\varphi(a)} V_r,\\
&V_r := \frac{1}{\sqrt{\varphi(a)K_{jj}}}\bigg[\int_{r-1}^{r}\Big\{b_j(\dot{X}(s)) + \beta_j(\dot{X}(s)/a) - \overline{b}_j - \overline{\beta}_j\Big\}\, ds\\
&\qquad\qquad + \big(\sigma B(r) - \sigma B(r-1)\big)_j\bigg].
\end{aligned} \tag{5.36}$$

Since $Ew_j^2(\dot{X}(t)) = Ew_j^2(\dot{X}(0)) = a^4 E g_j^2(\dot{Y}(0)) \le c_{20}a^2$ by Theorem 3.8 (also see Lemma 3.5), one has

$$E\bigg(\frac{1}{\sqrt{tK_{jj}}}\big(w_j(\dot{X}(t)) - w_j(\dot{X}(0))\big)\bigg)^2 \to 0 \quad \text{as } a \to \infty. \tag{5.37}$$

Therefore,

$$
\begin{aligned}
&\operatorname{var}\biggl(\sum_{r=1}^{\varphi(a)} V_r\biggr) \\
&\equiv E\biggl(\sum_{r=1}^{\varphi(a)} V_r\biggr)^2 \qquad (5.38)\\
&\to \lim_{a\to\infty} \operatorname{var}\biggl(\frac{1}{\sqrt{tK_{jj}}}\biggl\{\int_0^t \operatorname{grad} w_j(\dot{X}(s))\sigma B(s) + (\sigma B(t))_j\biggr\}\biggr) = 1.
\end{aligned}
$$

Indeed, the variance on the right is exactly 1. We will now prove the asymptotic normality of $\sum_{r=1}^{\varphi(a)} V_r$ by representing it approximately as the sum of a number of nearly independent block sums. For this purpose, define

$$
\begin{aligned}
&\delta = \varphi(a)/(a^4(\log a)^2), \qquad \eta = [\delta^{1/8} a^2 \log a], \\
&\psi = [\delta^{3/8} a^2 \log a], \qquad m = \biggl[\frac{\varphi(a)}{\eta + \psi}\biggr].
\end{aligned} \tag{5.39}
$$

Define the "big" and "little" block sums as in (5.21), (5.22), respectively, but with V_r as in (5.36). The rest of the proof that $\sum_{r=1}^{\varphi(a)} V_r \to_{\mathscr{L}} \mathscr{N}(0,1)$ is entirely analogous to the corresponding proof for Theorem 5.2. The only changes are in replacing ψ by ψ/a^2 and η by η/a^2 in the exponents in (5.26), and (5.28), (5.29) respectively. The reason for this adjustment is that we are directly considering the $X(\cdot)$ process, and not its scaled version $Y(\cdot)$. To check that the Lindeberg condition holds, as $a \to \infty$, for the sum of m i.i.d. random variables each having the same distribution as Z_1, write

$$
\begin{aligned}
&U_1 = \int_0^{\psi} \Bigl\{ b_j(\dot{X}(s)/a) + \beta_j(\dot{X}(s)/a) - \overline{b}_j - \overline{\beta}_j \Bigr\}\, ds, \\
&U_1' = \biggl(\int_0^{\psi} \sigma\, dB(s)\biggr)_j.
\end{aligned} \tag{5.40}
$$

Then $Z_1^2 \le (2/\varphi(a))(U_1^2 + U_1'^2)$, so that

$$
\begin{aligned}
mE\bigl(Z_1^2\, \mathbb{1}_{\{|Z_1|>\varepsilon\}}\bigr) &\le \frac{2m}{\varphi(a)} E\bigl(U_1^2\, \mathbb{1}_{\{|U_1|>(\varepsilon/2)\sqrt{\varphi(a)}\}}\bigr) \\
&\quad + \frac{2m}{\varphi(a)} E\Bigl(U_1'^2\, \mathbb{1}_{\{|U_1'|>(\varepsilon/2)\sqrt{\varphi(a)}\}}\Bigr) \\
&\quad + \frac{2m}{\varphi(a)} E\Bigl(U_1^2\, \mathbb{1}_{\{|U_1'|>(\varepsilon/2)\sqrt{\varphi(a)}\}}\Bigr) \\
&\quad + \frac{2m}{\varphi(a)} E\Bigl(U_1'^2\, \mathbb{1}_{\{|U_1|>\varepsilon/2\sqrt{\varphi(a)}\}}\Bigr).
\end{aligned} \tag{5.41}
$$

Since $|U_1| \le c\psi \le (\varepsilon/2)\sqrt{\varphi(a)}$ for all sufficiently large a, the first and last terms on the right side of (5.41) vanish for large a. Also, the second term is estimated by

$$
\begin{aligned}
&\frac{2m}{\varphi(a)} E\Big(U_1'^2 \, \mathbb{1}_{\{|U_1'|>(\varepsilon/2)\sqrt{\varphi(a)}\}}\Big) \\
(5.42) \qquad &\leq \frac{2m}{\varphi(a)} (EU_1'^4)^{1/2} \bigg(P\bigg(|U_1'| > \frac{\varepsilon}{2}\sqrt{\varphi(a)}\bigg)\bigg)^{1/2} \\
&\leq \frac{2m}{\varphi(a)} (c_{21}\psi)\big(EU_1'^2/\varepsilon^2\varphi(a)\big)^{1/2} \leq c_{21}' \frac{m\psi^{3/2}}{\varepsilon\varphi^{3/2}(a)} \to 0
\end{aligned}
$$

as $a \to \infty$. Finally, the third term on the right side of (5.41) is estimated by

$$
(5.43) \qquad c_{22} \frac{m}{\varphi(a)} \psi^2 P\bigg(|U_1'| > \frac{\varepsilon}{2}\sqrt{\varphi(a)}\bigg)^{1/2} \leq c_{22}' \frac{m\psi^2}{\varphi(a)} \exp\bigg\{-c_{23}\varepsilon^2 \frac{\varphi}{\psi}\bigg\} \to 0
$$

as $a \to \infty$. We have used an exponential bound for the tail probability of a Gaussian random variable for the last inequality. This completes the proof when $\dot{X}(0)$ has the uniform distribution on $\mathscr{T}_a$.

The proof of (5.34) under an arbitrary (initial) distribution of $\dot{X}(0)$ is analogous to that given for Theorem 5.2. Once again one takes $t = \varphi(a)$, $s = \psi(a)$ as in (5.39) and makes use of Theorem 4.5 and the fact that $s^2/t \to 0$ as $a \to \infty$. □

REMARK 5.3.1. An example in the next section shows that the time scale for large scale asymptotics in Theorem 5.2 cannot be smaller than $t \gg a^2$ in general. The time scale $t \gg a^4(\log a)^2$ in Theorem 5.3, however, seems too large. To understand the nature of technical difficulty encountered in trying to bring down the scale, one may attempt a "more straightforward" martingale CLT using the first equality in (5.36). Leaving aside the term $R := (tK_{jj})^{-1/2}[w_j(\dot{X}(t)) - w_j(\dot{X}(0))]$, one needs to show that the CLT applies to the term M, say, involving the stochastic integral [including $(\sigma B(t))_j$]. The proof of the conditional Lindeberg condition [see, e.g., Bhattacharya and Waymire (1990), page 508] requires an estimate of the growth of the stochastic integrand grad w_j beyond its second moment. Even under equilibrium, we are unable to obtain a precise estimate of this growth. Note that grad $w_j(x) = a(\text{grad}\, g_j)(x/a)$. Thus under equilibrium $\|w^2\|_1^2 \equiv a^2\|g_j\|_1^2$ is bounded by Theorem 3.8. If one could show that grad w_j is bounded in *sup norm* (not just in L^2) then, at least under equilibrium, the martingale term M is asymptotically normal for $t \gg a^2$. Similarly, under equilibrium, the L^2-norm of $w_j(x) \equiv a^2 g_j(x/a)$ is of the order $O(a^2)$, so that $R \to 0$ in probability for $t \gg a^2$. However, a direct estimate of the sup norm of w_j using the identity $w_j(x) = -\int_0^\infty T_s(b_j(\cdot) + \beta_j(\cdot/a) - \overline{b}_j - \overline{\beta}_j)(x)\,ds$ (with T_s as the transition operator of $\dot{X}$), yields a value of order larger than $a^2 \log a$, if one applies the rate of decay of the integrand given by (5.2). Thus, if $\dot{X}(0)$ is an arbitrary state, then to show $R \to 0$ in probability using this last estimate, we need $\sqrt{t} \gg a^2 \log a$. If one could show that ag_j and a grad g_j are *bounded in sup norm*, then the above arguments would lead to an improvement of the time scale in Theorem 5.3 to $t \gg a^2 \log a$. It is worthwhile to write this out as a theorem.

THEOREM 5.4. *If, in addition to the hypothesis in Theorem* 5.3, *one assumes that the functions* ag_j *and* a grad g_j $(1 \le j \le k_2)$ *are bounded in* sup norm, *then* (5.33) *holds for* $t \gg a^2 \log a$.

REMARK 5.4.1. Functional versions of Theorems 5.3 and 5.4 may be derived by arguments analogous to those given under Remark 5.2.1.

5.2. *Final phase of asymptotics for vector fields which are not divergence free—the one-dimensional case.* Since the general case of multiscale diffusions with periodic nondivergence-free vector fields is intractable, we will consider only one-dimensional diffusions. This will provide some insight into the nature and diversity of phenomena in the general case. Let $X(\cdot)$ be a one-dimensional diffusion governed by the Itô equation (2.1) whose coefficients satisfy the assumptions (B1)–(B3) in (4.33). Following the treatment of these processes given in the last part of Section 4, we will consider the nonself-adjoint and the self-adjoint cases separately. Once again, without any essential loss of generality, we will assume that (B4) in (4.34) also holds, that is, $\int_0^1 (b(x)/\sigma^2(x))\,dx = 0$. Using the notation in Section 4, let $\tilde{\pi}_a$ and π_a denote the invariant probability densities of $\dot{X}(t) \equiv X(t) \operatorname{mod} a$, and $\dot{Y}(t) \equiv Y(t) \operatorname{mod} 1$ $(Y(t) := X(a^2 t)/a)$, respectively. Write

$$\overline{b} = \int_0^a b(x)\tilde{\pi}_a(x)\,dx, \qquad \overline{\beta} = \int_0^a \beta(x/a)\tilde{\pi}_a(x)\,dx. \tag{5.44}$$

Note that unlike the case where $\tilde{\pi}_a$ and π_a are uniform densities, in general $\overline{\beta} \neq \int_0^a \beta(x)\tilde{\pi}_a(x)\,dx$. Let L be the generator of $\dot{X}(t)$, as given by (4.35), and let h be the unique mean-zero solution in $L^2(S_a^1, \tilde{\pi}_a)$ of

$$Lh(x) = b(x) + \beta(x/a) - \overline{b} - \overline{\beta}. \tag{5.45}$$

Define

$$\theta^2 = \int_0^a \sigma^2(x)(h'(x) - 1)^2 \tilde{\pi}_a(x)\,dx. \tag{5.46}$$

Note that, by Itô's lemma,

$$\begin{aligned} X(t) - X(0) - t(\overline{b} + \overline{\beta}) &= h(\dot{X}(t)) - h(\dot{X}(0)) \\ &\quad + \int_0^t \sigma(\dot{X}(s))\{1 - h'(\dot{X}(s))\}\,dB(s), \end{aligned} \tag{5.47}$$

so that, for a fixed a, θ^2 is the variance of the asymptotic normal distribution of $t^{-1/2}(X(t) - X(0) - t(\overline{b} + \overline{\beta}))$. The proof of the following theorem is based on Theorem 4.6 and a direct computation of h and is given in detail in Bhattacharya, Denker and Goswami (1999).

THEOREM 5.5. *In addition to* (B1)–(B4) *in* (4.33), (4.34), *assume that* $\beta(\cdot)$ *is bounded away from zero. Then for* $t \gg a^2 \log a$ *one has, for all* $X(0)$,

$$\frac{X(t) - X(0) - t(\overline{b} + \overline{\beta})}{\theta\sqrt{t}} \to_{\mathscr{L}} \mathscr{N}(0,1) \quad \text{as } a \to \infty. \tag{5.48}$$

Here $\theta = \theta(a)$ *is bounded away from zero and infinity.*

Note that the time scale as well as the growth in dispersion here are comparable to those in Theorem 5.3. The next theorem is dramatically different in these respects. For the case $\int_0^1 \beta(y)\,dy = 0$, write

$$\theta^* = \max_x \int_0^x \beta(y)\,dy, \qquad \theta_* = \min_x \int_0^x \beta(y)\,dy. \tag{5.49}$$

THEOREM 5.6. *In addition to* (B1)–(B4) *in* (4.33), (4.34), *assume that* $\sigma(\cdot)$ *is a constant*, $\beta(\cdot)$ *is nonconstant and* $\int_0^1 \beta(y)dy = 0$.

(a) *Then* $\theta = \theta(a)$ *defined by* (5.46) *goes to zero exponentially fast as* $a \to \infty$.

(b) *If* $t \gg a^2 \exp\{(18a/\sigma^2(\theta^* - \theta_*)\}$, *one has, for arbitrary initial states* $X(0) = ax_0$,

$$\frac{X(t) - ax_0}{\sigma a \theta \sqrt{t}} \to_{\mathscr{L}} \mathscr{N}(0,1) \quad \text{as } a \to \infty. \tag{5.50}$$

(c) *If* $t \ll a^{-4} \exp\{(2a/\sigma^2)(\theta^* - \theta_*)\}$ *then* (5.50) *does not hold, unless* $X(t) - ax_0 \to 0$ *in probability.*

Part (a) follows from a direct computation of θ in this case, while part (b) uses this computation of θ and Theorem 4.7 [see Bhattacharya, Denker and Goswami (1999) for details]. For part (c), one shows that, for the given range of t, $a\theta\sqrt{t} \to 0$ [Bhattacharya, Denkar and Goswami (1999)]. Therefore, if (5.50) is to hold, $(X(t) - ax)/ax$ must go to zero in probability.

REMARK 5.6.1. With regard to the centering in (5.50), it may be shown that $\overline{b} + \overline{\beta} = 0$ for all a.

REMARK 5.6.2. Part (a) of Theorem 5.6 shows that the asymptotic variance parameter or dispersion per unit time goes to zero exponentially fast, in dramatic contrast to the divergence-free case (Theorems 5.2, 5.3) and the one-dimensional nonself-adjoint case (Theorem 5.5). A heuristic explanation is that the invariant probability π_a either converges to a point mass or at least gets confined to a small set in the limit, as $a \to \infty$.

REMARK 5.6.3. The exponentially large time needed for the final Gaussian phase to take hold, as indicated in parts (b), (c) of Theorem 5.6, is not really due to the slow convergence to equilibrium as estimated in Theorem 4.7 for the "worst case scenario." Note that Theorem 5.6 holds under the hypothesis of Theorem 4.9 where relatively speedy convergence to equilibrium takes place.

REMARK 5.6.4. Consider the possibility of $X(t) - ax_0$ converging to 0 in probability as indicated in part (c) of Theorem 5.6. For the scaled version of $X(t)$, namely, the $Y(t)$ process, this means $Y(t) \to x_0$ as $a \to \infty$, $t \to \infty$, but $t \ll a^{-6} \exp\{(2a/\sigma^2)(\theta^* - \theta_*)\}$. Under the hypothesis of Theorem 5.6, this is impossible unless $\psi(x) = \int_0^x \beta(y)\, dy$ has a maximum at x_0. To show this, consider x_0 where ψ does not have a maximum. By using standard formulas [see, e.g., Bhattacharya and Waymire (1990), page 422, equation (10.12)], it is not difficult to check that if $x_1 < x_0 < x_2$ are such that ψ does not have a maximum in $[x_1, x_2]$, then the exit time τ of $Y(\cdot)$ from (x_1, x_2) has an expected value $E_y\tau$ which satisfies $\sup\{E_y\tau \colon y \in [x_1, x_2], a \geq 1\} < \infty$.

REMARK 5.6.5. Although Theorems 5.5, 5.6 address the case of one-dimensional multiscale diffusions with periodic coefficients, they point to a range of diverse behavior in the final phase for the general multidimensional nondivergence-free case. Theorems 4.7, 4.10 similarly indicate widely different time scales for approach to equilibrium on the big torus for the latter case. For example, if the invariant density π_a of the scaled diffusion $\dot{Y}(t) = \dot{X}(a^2 t)/a$ on $\mathscr{T}_1$ converges to a point mass as $a \to \infty$, one would expect the dispersion (per unit time) in the final phase to decay as $a \to \infty$ and the time scale for the final Gaussian approximation to be very large. This ought to be true, for example, in the case that the diffusion matrix is $\sigma^2 I_k$ (σ^2 a positive constant) and $b(x) = \operatorname{grad} \psi_1(x)$, $\beta(x) = \operatorname{grad} \psi_2(x)$, where (1) the "potential" functions ψ_1 and ψ_2 are periodic with period lattice $\mathbb{Z}^k$, (2) $\psi_1(\mathbf{n}) = 0 = \psi_2(\mathbf{n})$ for all $\mathbf{n} \in \mathbb{Z}^k$ and (3) on $[0, 1)^k$, ψ_2 has a unique maximum at x^*. In this case, the invariant probability $d(a) \exp\{(2/\sigma^2)(\psi_1(ay) + a\psi_2(y))\}\, dy$ of $\dot{Y}(t)$ converges to the point mass $\delta_{x^*}(dy)$ as $a \to \infty$; one would expect for this case an analog of Theorem 5.6 to hold.

6. Examples. In this section we provide two examples to illustrate the theory presented in Sections 2–5. Example 6.1 satisfies the hypotheses of Theorems 4.5 and 5.2, while Example 6.2 satisfies the hypotheses of Theorems 4.5 and 5.4.

EXAMPLE 6.1. Consider the diffusion on $\mathbb{R}^2$ defined by

$$\begin{aligned} dX_1(t) &= \{c_0 + c_1 \sin(2\pi(X_2(t)) + c_2 \cos(2\pi X_2(t)/a)\}\, dt + dB_1(t), \\ dX_2(t) &= dB_2(t), \qquad X(0) = ax = (ax_1, ax_2). \end{aligned} \tag{6.1}$$

Assume c_1, c_2 are nonzero and

$$\sin(2\pi x_2) = 0. \tag{6.2}$$

Table 1 shows the phase changes that occur along with their time scales. Here $\mathscr{L}(U)$ denotes the law, or distribution of a random variable U. The sign $\pm$ in (ii) is $+$ or $-$ according as $\cos 2\pi x_2 = -1$ or $+1$.

Table 1 is a modification of one derived in Bhattacharya and Götze (1995) under the initial condition $X(0) = x = (x_1, x_2)$. The latter initial condition implies $X(0)/a \to 0$ as $a \to \infty$, thereby essentially requiring that the process start at the origin. This issue becomes more important in the case (6.2) fails, as we show in a modification of Table 1 in Remark 6.1.1 below. The first row in the table is a consequence of Theorem 2.1(c), and Theorem 2.2. An alternative derivation may be given along the lines of case (i) of Example 6.2 below. To derive the second row, write

$$\begin{aligned}
&\frac{X_1(t) - X_1(0) - t(c_0 + c_2\cos 2\pi x_2)}{t^2/a^2}\\
&\quad = \frac{c_1}{t^2/a^2}\int_0^t \sin(2\pi(ax_2 + B_2(s)))\,ds \\
&\qquad + \frac{c_2}{t^2/a^2}\int_0^t \left\{\cos\left(2\pi\left(x_2 + \frac{B_2(s)}{a}\right)\right) - \cos 2\pi x_2\right\} ds + \frac{B_1(t)}{t^2/a^2}.
\end{aligned} \tag{6.3}$$

Note that $a^{4/3} \ll t \Leftrightarrow t^2/a^2 \gg t^{1/2}$. Therefore, $B_1(t)/(t^2/a^2) \to 0$ in probability. Now use Itô's lemma to get

$$\begin{aligned}
&\int_0^t \sin\bigl(2\pi(ax_2 + B_2(s))\bigr)\,ds \\
&\quad = -\frac{1}{2\pi^2}\bigl\{\sin\bigl(2\pi(ax_2 + B_2(t))\bigr) - \sin 2\pi a x_2\bigr\} \\
&\qquad + \frac{1}{\pi}\int_0^t \cos\bigl(2\pi(ax_2 + B_2(s))\bigr)\,dB_2(s).
\end{aligned} \tag{6.4}$$

From this it is clear that the first term on the right in (6.3) goes to zero. It remains to show that the middle term on the right in (6.3) has the asymptotic distribution $\mathscr{L}(\pm 2c_2\pi^2\int_0^1 B_2^2(s)\,ds)$. By a Taylor expansion,

$$\begin{aligned}
&\cos\left(2\pi\left(x_2 + \frac{B_2(s)}{a}\right)\right) - \cos 2\pi x_2 \\
&\quad = -\frac{2\pi B_2(s)}{a}\sin 2\pi x_2 - \frac{4\pi^2}{2a^2}B_2^2(s)\cos 2\pi x_2 + \frac{8\pi^3}{6a^3}B_2^3(s)\theta,
\end{aligned} \tag{6.5}$$

TABLE 1
Phase changes in Example 6.1

Time scale	**Asymptotic law**
(1) $1 \ll t \ll a^{4/3}$	$\dfrac{X_1(t) - X_1(0) - t(c_0 + c_2\cos 2\pi x_2)}{\sqrt{t}} \to_{\mathscr{L}} \mathscr{N}(0, 1 + c_1^2/2\pi^2)$
(2) $a^{4/3} \ll t \ll a^2$	$\dfrac{X_1(t) - X_1(0) - t(c_0 + c_2\cos 2\pi x_2)}{t^2/a^2} \to_{\mathscr{L}} \mathscr{L}\left(\pm 2c_2\pi^2\int_0^1 B_2^2(s)\,ds\right)$
(3) $t/a^2 \to r > 0$	$\dfrac{X_1(t) - X_1(0) - tc_0}{t} \to_{\mathscr{L}} \mathscr{L}\left(\dfrac{c_2}{r}\int_0^r \cos(2\pi(x_2 + B_2(s)))\,ds\right)$
(4) $t \gg a^2$	$\dfrac{X_1(t) - X_1(0) - tc_0}{a\sqrt{t}} \to_{\mathscr{L}} \mathscr{N}(0, c_2^2/2\pi^2)$

where θ is a random variable, $|\theta| \le 1$. The first term on the right is zero by assumption (6.2). Also, $E|B_2^3(s)| \le cs^{3/2}$, so that

$$\frac{c_2}{t^2/a^2}\left|\int_0^t \frac{8\pi^3}{6a^3}\, B_2^3(s)\theta\, ds\right| \le \frac{c'}{t^2 a} t^{5/2} = \frac{c' t^{1/2}}{a} \to 0, \tag{6.6}$$

since $t \ll a^2$. Thus the middle term on the right in (6.3) has the same asymptotic distribution as

$$\begin{aligned}&\frac{c_2}{t^2/a^2}\int_0^t -\frac{4\pi^2}{2a^2}\, B_2^2(s)\cos 2\pi x_2\, ds\\ &\quad = -\frac{2\pi^2 c_2 \cos 2\pi x_2}{t^2}\int_0^t B_2^2(s)\, ds =_{\mathscr{L}} -2\pi^2 c_2 \cos 2\pi x_2 \int_0^1 B_2^2(s)\, ds.\end{aligned} \tag{6.7}$$

For the last equality in law we use the fact that, for every $t > 0$, the distributions of the processes $\{\sqrt{t}\, B_2(s/t)\colon s \ge 0\}$ and $\{B_2(s)\colon s \ge 0\}$ are the same. To derive the third row in the table use the representation with a denominator t, instead of t^2/a^2, and omit the centering term $c_2 \cos 2\pi x_2$ from both sides, to get the desired asymptotic distribution the same as that of

$$\begin{aligned}&\frac{c_2}{t}\int_0^t \cos\left(2\pi\left(x_2 + \frac{B_2(s)}{a}\right)\right) ds\\ &\quad =_{\mathscr{L}} \frac{c_2}{t}\int_0^t \cos\bigl(2\pi(x_2 + B_2(s/a^2))\bigr)\, ds\\ &\quad = \frac{c_2}{t/a^2}\int_0^{t/a^2} \cos\bigl(2\pi(x_2 + B_2(s)\bigr)\, ds\\ &\quad \to \frac{c_2}{r}\int_0^r \cos\bigl(2\pi(x_2 + B_2(s))\bigr)\, ds.\end{aligned} \tag{6.8}$$

The final phase (iv) in Table 1 follows from Theorem 5.2 for time scales $t \gg a^2(\log a)^2$. By explicit computation we now show that it holds for times $t \gg a^2$. As above, since $a\sqrt{t} \gg \sqrt{t}$, one only needs to evaluate the asymptotic distribution of

$$\frac{c_2}{a\sqrt{t}}\int_0^t \cos\left(2\pi\left(x_2 + \frac{B_2(s)}{a}\right)\right) ds. \tag{6.9}$$

Since the function $f(y) = -(c_2 a^2/2\pi^2)\cos(2\pi(x_2 + \frac{y}{a}))$ satisfies $\frac{1}{2}f''(y) = c_2\cos(2\pi(x_2 + y/a))$, Itô's lemma shows that (6.9) equals

$$\begin{aligned}&\frac{1}{a\sqrt{t}}\left(-\frac{c_2 a^2}{2\pi^2}\right)\left\{\cos\left(2\pi\left(x_2 + \frac{B_2(t)}{a}\right)\right) - \cos 2\pi x_2\right\}\\ &\quad - \frac{1}{a\sqrt{t}}\int_0^t \frac{c_2 a}{\pi}\sin\left(2\pi\left(x_2 + \frac{B_2(s)}{a}\right)\right) dB_2(s)\\ &\quad \simeq -\frac{c_2}{\pi\sqrt{t}}\int_0^t \sin\left(2\pi\left(x_2 + \frac{B_2(s)}{a}\right)\right) dB_2(s) = I(t),\end{aligned} \tag{6.10}$$

say. To show that the last expression is asymptotically normal, note that its quadratic variation is

$$
\begin{aligned}
Q(t) &:= \left(\frac{c_2^2}{\pi^2}\right)\frac{1}{t}\int_0^t \sin^2\left(2\pi\left(x_2+\frac{B_2(s)}{a}\right)\right)ds\\
&=_{\mathscr{L}} \left(\frac{c_2^2}{\pi^2}\right)\frac{1}{t}\int_0^t \sin^2\left(2\pi\left(x_2+B_2\left(\frac{s}{a^2}\right)\right)\right)ds\\
&= \frac{c_2^2}{\pi^2}\frac{1}{t/a^2}\int_0^{t/a^2} \sin^2\big(2\pi(x_2+B_2(u))\big)\,du\\
&\to \frac{c_2^2}{\pi^2}\int_0^1 \sin^2(2\pi y)\,dy = \frac{c_2^2}{2\pi^2} \quad \text{a.s.,}
\end{aligned}
\tag{6.11}
$$

since $t/a^2 \to \infty$, and the process $U(t) := (x_2 + B_2(t)) \bmod 1$ is a positive recurrent Markov process on $S^1 = \{x \bmod 1\colon x \in \mathbb{R}\}$ having the uniform distribution as its invariant probability. One may now check that the martingale central limit theorem [see, e.g., Bhattacharya and Waymire (1990), page 508] holds for the last expression $I(t)$ in (6.10), with the asymptotic variance $c_2^2/2\pi^2$. An alternative derivation may be given by noting that $E\exp\{i\xi I(t)+\xi^2/2Q(t)\} = 1\ \forall\,\xi$ and $\forall\,t$. By (6.11), $Q(t) \to c_2^2/2\pi^2$ a.s., as $a \to \infty$, $t \gg a^2$. Since $|Q(t)| \le c_2^2/\pi^2$ for all t and a, one may now easily show that $E\exp\{i\xi I(t)\} \to \exp\{-\xi^2/2\sigma^2\}$ with $\sigma^2 = c_2^2/2\pi^2$.

REMARK 6.1.1. The above example shows that the time scale for the first phase of asymptotics derived in Theorem 2.1(c), Theorem 2.2, is exact, namely, $1 \ll t \ll a^{4/3}$. Indeed, with an additional calculation one may show that if $a \to \infty$, $t/a^{4/3} \to r > 0$, then

$$
\begin{aligned}
&\frac{X_1(t)-X_1(0)-t(c_0+c_2\cos 2\pi x_2)}{\sqrt{t}}\\
&\qquad - \frac{c_2\int_0^t\{\cos(2\pi(x_2+(B_2(s)/a)))-\cos 2\pi x_2\}\,ds}{\sqrt{t}}\\
&\to_{\mathscr{L}} \mathscr{N}\left(0, 1+\frac{c_1^2}{2\pi^2}\right).
\end{aligned}
\tag{6.12}
$$

Now, by (6.5), and (6.6), (6.7) (with t^2/a^2 replaced by $\sqrt{t}$), one shows that

$$
\begin{aligned}
&-\frac{c_2}{\sqrt{t}}\int_0^t\left\{\cos\left(2\pi\left(x_2+\frac{B_2(s)}{a}\right)\right)-\cos 2\pi x_2\right\}ds\\
&\qquad \to_{\mathscr{L}} \mathscr{L}\left(2\pi^2 c_2 r^{3/2}(\cos 2\pi x_2)\int_0^1 B_2^2(u)\,du\right).
\end{aligned}
\tag{6.13}
$$

The limiting law in (6.13) is that of a strictly positive or a strictly negative random variable (depending on whether $c_2\cos 2\pi x_2$ is positive or negative). From this it follows that for $t/a^{4/3} \to r > 0$, the asymptotic law in Table 1(1) does not hold.

REMARK 6.1.2. If in Example 6.1 we drop the assumption (6.2), and instead assume

$$\sin 2\pi x_2 \neq 0, \tag{6.14}$$

then the hypothesis of part (b) of Theorem 2.1 is satisfied, but not that of part (c). Therefore, the time scale for case (1) is $1 \ll t \ll a$. The arguments for cases (3) and (4) remain unchanged. Case (2), however, changes drastically. For $a \ll t \ll a^2$ one has, using (6.3) with t^2/a^2 replaced by $t^{3/2}/a$, and noting that $t^{3/2}/a \gg t^{1/2}$,

$$\begin{aligned}
&\frac{X_1(t) - X_1(0) - t(c_0 + c_2 \cos 2\pi x_2)}{t^{3/2}/a} \\
&\quad \simeq \frac{c_2}{t^{3/2}/a} \int_0^t \left\{ \cos\left(2\pi\left(x_2 + \frac{B_2(s)}{a}\right)\right) - \cos 2\pi x_2 \right\} ds \\
&\quad = \frac{c_2}{t^{3/2}/a} \int_0^t (-\sin 2\pi x_2) \frac{2\pi B_2(s)}{a}\, ds \\
&\qquad + O\left(\frac{a}{t^{3/2}} \int_0^t \frac{B_2^2(s)}{a^2}\, ds\right) \\
&\quad \simeq -\frac{2\pi c_2 \sin 2\pi x_2}{t^{3/2}} \int_0^t B_2(s)\, ds,
\end{aligned} \tag{6.15}$$

since the expected value of the magnitude of the O-term is $O(1/at^{3/2}t^2) = O(t^{1/2}/a) \to 0$ for $t \ll a^2$. Now the last expression in (6.15) has the same distribution as

$$-2\pi c_2 \sin 2\pi x_2 \int_0^1 B_2(s)\, ds, \tag{6.16}$$

which is $\mathscr{N}(0, (4\pi^2 c_2^2/3)\sin^2 2\pi x_2)$. Thus, under (6.14), the first two rows of Table 1 change to

Time scale	**Asymptotic law**
$(1)'$ $1 \ll t \ll a$	$\dfrac{X_1(t) - X_1(0) - t(c_0 + c_2 \cos 2\pi x_2)}{\sqrt{t}} \to_{\mathscr{L}} \mathscr{N}\left(0, 1 + \dfrac{c_1^2}{2\pi^2}\right)$
$(2)'$ $a \ll t \ll a^2$	$\dfrac{X_1(t) - X_1(0) - t(c_0 + c_2 \cos 2\pi x_2)}{t^{3/2}/a} \to_{\mathscr{L}} \mathscr{N}\left(0, \dfrac{4\pi^2 c_2^2}{3} \sin^2 2\pi x_2\right)$

Once again, if $a \to \infty$, $t/a \to r > 0$, then the asymptotic law in $(1)'$ cannot hold. To see this note that in the integral representation of $t^{-1/2}(X_1(t) - X_1(0) - t(c_0 + c_2 \cos 2\pi x_2))$ [see (6.3)], $t^{-1/2} \int_0^t c_1 \sin 2\pi(ax_2 + B_2(s))\, ds + t^{-1/2} \cdot$

$B_1(t)$ converges in law to $\mathscr{N}(0, 1 + c_1^2/2\pi^2)$, as in (1)$'$. However, here the middle term $t^{-1/2}\int_0^t c_2\{\cos(2\pi(x_2 + B_2(s)/a)) - \cos 2\pi x_2\}\,ds$ converges in law to $-r(2\pi c_2 \sin 2\pi x_2)\int_0^1 B_2(u)\,du = rZ_2$, say, by essentially the same argument as given for (2)$'$ above [see (6.15), (6.16)]. Thus if the asymptotic law in (1)$'$ is to hold for $t/a \to r > 0$ ($a \to \infty$), then one would have in the limit $Z_1 + rZ_2 =_{\mathscr{L}} Z_1$, where Z_1 and Z_2 are nondegenerate normal. This can not hold if r is sufficiently large. Therefore, the time scale given in Theorem 2.1(b) cannot be improved upon in general. The preciseness of the time scale $t \ll a^{2/3}$ in part (a) of Theorem 2.1 will be shown in Remark 6.2.1 below.

REMARK 6.1.3. The time scale for the final phase in Example 6.1 is $t \gg a^2$, whereas Theorem 5.2 gives a time scale $t \gg a^2(\log a)^2$ in the general case. We do not know if, in general, the logarithmic factor can be dropped altogether. Recall that our estimation for the time scale to equilibrium on the big torus is already $t \gg a^2 \log a$ (Theorem 4.5).

EXAMPLE 6.2. Consider the same equation for $X_1(t)$ as in Example 6.1 [see (6.1)], but for $X_2(t)$ take a Brownian motion with a *nonzero drift* δ,

$$dX_2(t) = \delta\,dt + dB_2(t). \tag{6.17}$$

The initial condition is as in (6.1), namely, $X(0) = ax = (ax_1, ax_2)$, but we assume $\sin 2\pi x_2 \neq 0$ [i.e., (6.14)]. With this seemingly minor change, the asymptotic behavior and time scales are dramatically different at larger scales, as shown in Table 2.

Case (1) follows from Theorem 2.1(a) and Theorem 2.2, or one can directly use the integral representation (6.3), but with a different denominator, namely,

TABLE 2
Phase changes in Example 6.2

Time scale	Asymptotic law
(1) $1 \ll t \ll a^{2/3}$	$\dfrac{X_1(t) - X_1(0) - t(c_0 + c_2\cos 2\pi x_2)}{\sqrt{t}} \to_{\mathscr{L}} \mathscr{N}\left(0, 1 + \dfrac{c_1^2}{2(\delta^2 + \pi^2)}\right)$
(2) $\dfrac{t}{a^{2/3}} \to r > 0$	$\dfrac{X_1(t) - X_1(0) - t(c_0 + c_2\cos 2\pi x_2)}{\sqrt{t}}$ $\to_{\mathscr{L}} \mathscr{N}\left(-c_2\delta r^{3/2}\pi \sin 2\pi x_2,\ 1 + \dfrac{c_1^2}{2(\delta^2 + \pi^2)}\right)$
(3) $t \gg a^2$	$\dfrac{X_1(t) - X_1(0) - tc_0}{\sqrt{t}} \to_{\mathscr{L}} \mathscr{N}\left(0, 1 + \dfrac{c_1^2}{2(\pi^2 + \delta^2)} + \dfrac{c_2^2}{2\delta^2}\right)$

$t^{1/2}$ (instead of t^2/a^2),

$$
\begin{aligned}
&\frac{X_1(t) - X_1(0) - t(c_0 + c_2 \cos 2\pi x_2)}{\sqrt{t}} \\
&\quad = \frac{c_1}{\sqrt{t}} \int_0^t \sin 2\pi(ax_2 + s\delta + B_2(s))\, ds \qquad (6.18) \\
&\qquad + \frac{c_2}{\sqrt{t}} \int_0^t \left\{ \cos\left(2\pi\left(x_2 + \frac{s\delta}{a} + \frac{B_2(s)}{a}\right)\right) - \cos 2\pi x_2 \right\} ds + \frac{B_1(t)}{\sqrt{t}}.
\end{aligned}
$$

Now $\sin 2\pi(ax_2 + s\delta + B_2(s)) = \sin 2\pi Z(s)$, where $Z(s) = X_2(s) \bmod 1$ is the Brownian motion on the unit circle with a drift. Since the distribution of $Z(s)$ approaches equilibrium (uniform distribution) exponentially fast in total variation distance, uniformly with respect to the initial state, it follows from a central limit theorem for Markov processes [see Bhattacharya (1982)] that the first term on the right in (6.18) converges in distribution to $\mathscr{N}(0, \sigma^2)$ where

$$
\sigma^2 = -2 \int_{[0,1)} f(y)u(y)\, dy, \qquad f(y) = c_1 \sin 2\pi y, \tag{6.19}
$$

$u(y)$ being the mean zero solution of

$$
\tfrac{1}{2}u''(y) + \delta u'(y) = f(y). \tag{6.20}
$$

A direct computation shows

$$
u(y) = -\frac{c_1\delta}{\pi^2 + \delta^2}\left\{ \frac{\cos 2\pi y}{2\pi} + \frac{1}{2\delta} \sin 2\pi y \right\}, \tag{6.21}
$$

leading to

$$
\sigma^2 = \frac{2c_1^2\delta}{\pi^2 + \delta^2}\left(\frac{1}{2\delta}\right) \int_{[0,1)} \sin^2 2\pi y\, dy = \frac{c_1^2}{2(\pi^2 + \delta^2)}. \tag{6.22}
$$

The third term on the right in (6.18) is independent of the first, and its distribution is $\mathscr{N}(0, 1)$. Thus the sum of the first and third terms converges in distribution to $\mathscr{N}(0, 1 + c_1^2/2(\pi^2 + \delta^2))$ as $t \to \infty$ (uniformly w.r.t. a). If, in addition $a \to \infty$, $t/a^{2/3} \to r > 0$ then, using a Taylor expansion such as in (6.5), the middle term on the right in (6.18) may be expressed as

$$
\frac{2\pi c_2}{\sqrt{t}} \int_0^t (-\sin 2\pi x_2)\left(\frac{B_2(s)}{a} + \frac{s\delta}{a}\right) ds + O\left(\frac{1}{\sqrt{t}} \int_0^t \left(\frac{B_2(s) + s\delta}{a}\right)^2 ds\right) \tag{6.23}
$$

The expected value of the O-term is of the order $O(t^3/\sqrt{t}\, a^2) = O(t^{5/2}/a^2) \to 0$, since $t^{5/2} = O(a^{5/3})$. Since $E|B_2(s)| = c' s^{1/2}$, the dominant contribution in the first term in (6.23) comes from

$$
\begin{aligned}
\frac{2\pi c_2}{\sqrt{t}}(-\sin 2\pi x_2)\frac{\delta}{a} \int_0^t s\, ds &= -\frac{2\pi c_2 \delta \sin 2\pi x_2}{\sqrt{t}\, a} \frac{t^2}{2} \\
&= (-c_2 \pi \delta \sin 2\pi x_2)\frac{t^{3/2}}{a} \to (-c_2\pi\delta \sin 2\pi x_2) r^{3/2}.
\end{aligned}
$$

Thus (2) is established. In particular, this shows, along with Example 6.1, that the time scales in Theorems 2.1, 2.2 are in general precise.

To derive case (3) in Table 2, we will use Itô's lemma to write

$$\begin{aligned}\frac{X_1(t)-X_1(0)-tc_0}{\sqrt{t}} &= \frac{w_1(X_2(t))-w_1(X_2(0))}{\sqrt{t}} \\ &\quad - \frac{1}{\sqrt{t}}\int_0^t w_1'(X_2(s))\,dB_2(s) + \frac{1}{\sqrt{t}}B_1(t),\end{aligned} \tag{6.24}$$

where w_1 is a periodic solution of

$$\tfrac{1}{2}w_1''(y) + \delta w_1'(y) = c_1 \sin 2\pi y + c_2 \cos(2\pi y/a). \tag{6.25}$$

By direct computation, w_1 is given by [apart from an additive constant which does not affect the right side of (6.24)]

$$\begin{aligned} w_1(y) &= -\frac{c_1\delta}{\pi^2+\delta^2}\left\{\frac{\cos 2\pi y}{2\pi} + \frac{\sin 2\pi y}{2\delta}\right\} \\ &\quad + \frac{c_2\delta a^3}{\delta^2 a^2+\pi^2}\left\{\frac{\sin(2\pi y/a)}{2\pi} - \frac{\cos(2\pi y/a)}{2\delta a}\right\}.\end{aligned} \tag{6.26}$$

Note that w_1 is $O(a)$. Therefore, if $t \gg a^2$, the first term on the right side in (6.24) goes to zero a.s. The integrand in the stochastic integral term is

$$\begin{aligned} w_1'(y) &= \frac{c_1\delta}{\pi^2+\delta^2}\left\{\sin 2\pi y - \frac{\pi\cos 2\pi y}{\delta}\right\} \\ &\quad + \frac{c_2\delta a^2}{\delta^2 a^2+\pi^2}\left\{\cos\left(\frac{2\pi y}{a}\right) + \frac{\pi \sin(2\pi y/a)}{\delta a}\right\}.\end{aligned} \tag{6.27}$$

Neglecting the $O(1/a)$ term whose contribution in the stochastic integral obviously goes to zero in probability, one may then write

$$\begin{aligned}\frac{X_1(t)-X_1(0)-tc_0}{\sqrt{t}} &\simeq -\frac{1}{\sqrt{t}}\int_0^t \left\{I_1(X_2(s)) + I_2(X_2(s))\right\} dB_2(s) \\ &\quad + \frac{B_1(t)}{\sqrt{t}},\end{aligned} \tag{6.28}$$

where

$$\begin{aligned} I_1(y) &= \frac{c_1\delta}{\pi^2+\delta^2}\left\{\sin 2\pi y - \frac{\pi\cos 2\pi y}{\delta}\right\}, \\ I_2(y) &= \frac{c_2\delta a^2}{\delta^2 a^2+\pi^2}\cos\left(\frac{2\pi y}{a}\right).\end{aligned} \tag{6.29}$$

The stochastic integral in (6.28) is a martingale and its quadratic variation (divided by t) is

$$\frac{1}{t}\int_0^t I_1^2(X_2(s))ds + \frac{1}{t}\int_0^t I_2^2(X_2(s))\,ds + \frac{2}{t}\int_0^t I_1(X_2(s))I_2(X_2(s))\,ds. \tag{6.30}$$

As argued for case (1), $I_1(X_2(s)) = I_1(Z(s))$ ($Z(s) := X_2(s) \bmod 1$), when $Z(s)$ is a Brownian motion on the unit circle with a constant drift δ, which approaches equilibrium exponentially fast in t,

$$\begin{aligned}\frac{1}{t}\int_0^t I_1^2(X_2(s))\,ds \to \int_{[0,1)} I_1^2(y)\,dy &= \frac{c_1^2\delta^2}{(\pi^2+\delta^2)^2}\left\{\frac{1}{2}+\frac{\pi^2}{2\delta^2}\right\}\\ &= \frac{c_1^2}{2(\pi^2+\delta^2)} \quad \text{as } t\to\infty.\end{aligned} \tag{6.31}$$

For the second term in (6.30), write

(6.32)

$$\begin{aligned}\frac{1}{t}\int_0^t I_2^2(X_2(s))\,ds &\simeq \left(\frac{c_2}{\delta}\right)^2\frac{1}{t}\int_0^t \cos^2 2\pi\left(\frac{B_2(s)}{a}+\frac{s\delta}{a}+x_2\right)ds\\ &=_{\mathscr{L}} \left(\frac{c_2}{\delta}\right)^2\frac{1}{t}\int_0^t \cos^2 2\pi\left(B_2(s/a^2)+\frac{s\delta}{a}+x_2\right)ds\\ &= \left(\frac{c_2}{\delta}\right)^2\frac{1}{t/a^2}\int_0^{t/a^2} \cos^2 2\pi\big(B_2(s')+as'\delta+x_2\big)\,ds'.\end{aligned}$$

Once again one may replace $B_2(s')+as'\delta+x_2$ by its value mod 1 and use the fact that the latter is a Brownian motion on the unit circle with a drift $a\delta$. This Brownian motion on the circle approaches equilibrium as $s'\to\infty$, uniformly w.r.t. the drift $a\delta$ since the Brownian motion on the unit circle *without drift* approaches equilibrium (exponentially fast in total variation distance) *uniformly* with respect to the initial state. Thus, as $t/a^2\to\infty$,

$$\frac{1}{t}\int_0^t I_2^2(X_2(s))\,ds \to \frac{c_2^2}{2\delta^2} \quad \text{in probability.} \tag{6.33}$$

We now show that the product term in (6.30) goes to zero in probability. For this note that

$$\begin{aligned}&\frac{1}{t}\int_0^t \sin 2\pi\big(B_2(s)+s\delta+ax_2\big)\cos 2\pi\left(\frac{B_2(s)}{a}+\frac{s\delta}{a}+x_2\right)ds\\ &\quad=_{\mathscr{L}} \frac{1}{t/a^2}\int_0^{t/a^2} \sin 2\pi(aB_2(s)+a^2s\delta+ax_2)\\ &\qquad\qquad \times \cos 2\pi(B_2(s)+as\delta+x_2)\,ds\\ &\quad= \frac{1}{t/a^2}\int_0^{t/a^2} \sin 2\pi(aZ_2(s))\cos 2\pi(Z_2(s))\,ds,\end{aligned} \tag{6.34}$$

where $Z_2(s) = (B_2(s)+sa\delta+x_2) \bmod 1$. Since, as argued earlier, $Z(s) := (B_2(s)+y) \bmod 1$ approaches equilibrium (exponentially fast in total variation distance) uniformly w.r.t. y, as $s\to\infty$, the last expression in (6.34) is asymptotically the same in distribution as

$$\frac{1}{t/a^2}\int_0^{t/a^2} \sin 2\pi(aZ(s))\sin 2\pi Z(s)\,ds, \tag{6.35}$$

where $\{Z(s)\colon s \geq 0\}$ is the stationary standard Brownian motion on the unit circle. One may rewrite (6.35) as

$$\frac{1}{2t/a^2}\int_0^{t/a^2}\left\{\sin(2\pi(a+1)Z(s)) + \sin(2\pi(a-1)Z(s))\right\}ds. \tag{6.36}$$

Now, uniformly for all $a = 1, 2, \ldots,$

$$E\left(\frac{1}{A}\int_0^A \sin(2\pi(a+1)Z(s))\,ds\right)^2 \to 0 \quad \text{as } A \to \infty. \tag{6.37}$$

To see this, note that $Z(s)$ is exponentially φ-mixing, and $z \to \sin(2\pi(a+1)z)$ is uniformly bounded. Hence the covariance $E\sin(2\pi(a+1)Z(s))\sin(2\pi(a+1)Z(s')) \to 0$ exponentially fast, uniformly in a, as $|s-s'| \to \infty$. One may replace $a+1$ in (6.37) by $a-1$ and thus show that (6.36) goes to zero in mean square as $t/a^2 \to \infty$. It follows that (6.34) $\to 0$ in mean square. The same proof applies to the other term in $I_1(X_2(s))I_2(X_2(s))$ involving $\cos 2\pi(X_2(s))$ $\cos 2\pi(X_2(s)/a)$ [see (6.29)]. Thus the average quadratic variation (6.30) converges in probability to $c_1^2/2(\pi^2+\delta^2) + c_2^2/\delta^2$ [see (6.31), (6.33)]. One may now easily apply the martingale central limit theorem to the stochastic integral term in (6.28), verifying the Lindeberg-type condition using the fact that $|I_1(y) + I_2(y)|$ is bounded (uniformly in a) [see Bhattacharya and Waymire (1990), page 508]. Alternatively, one may also show that the characteristic function of $t^{-\frac{1}{2}}\int_0^t I(X_2(s))\,dB_2(s)$ converges to that of the appropriate Gaussian, by using the exponential martingale property, as in the proof of case (4) of Example 6.1.

REMARK 6.2.1. To show that the time scale $t \gg a^2$ for the final phase in Example 6.2 is precise, let $a \to \infty$, $t/a^2 \to r > 0$. Then, if r is sufficiently large, there exists a positive constant c (independent of a, t and r) such that $E(t^{-1/2}(w_1(X_2(t)) - w_1(X_2(0)))^2 \geq cr$ [see (6.26)]. On the other hand, the mean square of the sum of the two remaining terms on the right side of (6.24) is bounded by an absolute constant c'. Therefore, the first term will be dominant for large r. This shows that (3) in Table 2 does not hold if the time scale is extended to include $t = O(a^2)$.

REMARK 6.2.2. The hypothesis of Theorem 5.4 is satisfied by Example 6.2, with $k_2 = 1$.

REMARK 6.2.3. The Gaussian convergences in Examples 6.1 and 6.2 may be strengthened to their functional versions (i.e., convergence to Brownian motions) by standard results such as given in Theorem 7.1.4 in Ethier and Kurtz (1986) [also see Hall and Heyde (1980), page 99]. The non-Guassian convergences in these examples may also be expressed in functional forms.

7. An application to solute transport in porous media. Suppose a chemical pollutant, or some solute, is injected at a point in a saturated aquifer —an underground water system. How will it spread over large times? There is a vast engineering literature on this subject [see Adams and Gelhar (1992); Bhattacharya and Gupta (1983); Cushman (1990); Dagan (1984); Fried and Combarnous (1971); Garabedian, LeBlanc, Gelhar and Celia (1991); Gelhar and Axness (1983); Gupta and Bhattacharya (1986); Guven and Molz (1986); LeBlanc, Garabedian, Hess, Gelhard, Quadri, Stollenwerk and Wood (1991); Sauty (1980); Sposito, Jury and Gupta (1986); Sudicky (1986)]. It is generally accepted [Fried and Combarnous (1971)], and laboratory scale experiments have confirmed it, that the solute concentration $c(t, y)$ at y at time t at a local scale, say the laboratory scale, satisfies a Fokker–Planck equation,

$$\frac{\partial c}{\partial t} = \frac{1}{2} \sum_{j, j'=1}^{3} \frac{\partial^2}{\partial y_j \partial y_{j'}} (D_{jj'} c) - \sum_{j=1}^{3} \frac{\partial}{\partial y_j} (v_j(y) c), \tag{7.1}$$

with $v(y) = (v_1(y), v_2(y), v_3(y))$ representing the velocity of water at y, and satisfying the *incompressibility* condition

$$\operatorname{div} v(y) = 0 \quad \forall\, y. \tag{7.2}$$

The positive definite symmetric matrix $((D_{jj'}))$ may represent something akin to Einstein's molecular diffusion $\sigma^2 I_3$ at a scale somewhat larger than the hydrodynamical scale [see, e.g., Bhattacharya and Gupta (1979), where this is erroneously called the "Darcy scale"], or an enhanced dispersion due to heterogeneities in the porous medium at the laboratory, or the so-called Darcy scale [Fried and Combarnous (1971)]. A commonly used experimental methodology is to fit Gaussians to the concentration $c(t, y)$ as a function of y, for successively larger scales of t. One may think of this as different Brownian motion approximations at different scales of time. It has been widely observed that the diagonal dispersion coefficients, or variances per unit time, increase steadily with the time scale, especially in the direction of flow. This phenomenon has been called the *scale effect* in dispersion. A different kind of study has focussed on the increase in dispersion at the laboratory-, or Darcy-, scale with the increase in the velocity magnitude of the flow [Fried and Combarnous (1971)].

As is well known [see, e.g., Friedman (1975), pages 144–150, or Bhattacharya and Waymire (1990), pages 377–380], the solution to (7.1) with a point initial input c_0 at x is given by the function $(t, y) \to c_0 p(t; x, y)$, where $p(t; x, y)$ is the transition probability density of a diffusion $X(t)$ with drift velocity v and diffusion coefficients $D_{jj'}$. In general, for an arbitrary compactly supported and continuous initial concentration $c_0(x)$, the solution to (7.1) is

$$c(t, y) = \int c_0(x) p(t; x, y)\, dx. \tag{7.3}$$

It follows that the asymptotic behavior of $c(t, y)$ for large t is given by the asymptotic distribution of $X(t)$. The present article provides these asymptotics assuming v to be periodic. For the physical problem at hand, the initial concentration is always taken to be *localized* at a point.

To study the *effect of velocity* on dispersion, let $v = u_0 \beta$ where u_0 is a scalar and β is periodic. It is shown in Section 3 that $X(t)$ is asymptotically Gaussian for large t [and therefore so is $c(t, \cdot)$], but with two extreme behaviors of *dispersivity* (i.e., asymptotic variance per unit time) depending on the nature of the flow velocity. If $\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j)$ has a nonzero component in the *null space* N of $S = \mathscr{D}^{-1}\beta\nabla$ in H^1, then the dispersivity of $X_j(t)$ grows quadratically with u_0 [Theorem 3.3(a)]. In the complementary case, $\mathscr{D}^{-1}(\beta_j - \overline{\beta_j})$ belong to the closure $\overline{\mathscr{R}}$ of the *range* $\mathscr{R}$ of S, since $H^1 = N \oplus \overline{\mathscr{R}}$. If $\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j) \in \mathscr{R}$, then the dispersivity of $X_j(t)$ grows from D_{jj} to a larger constant value, as u_0 increases [Theorem 3.3(b)]. The boundary case, where $\mathscr{D}^{-1}(\beta_j - \overline{\beta}_j) \in \overline{\mathscr{R}} \backslash \mathscr{R}$, seems difficult to analyze. For two-dimensional flows (i.e., $k = 2$), more information on this may be found in Fannjiang and Papanicolaou (1994). Figures 1 and 2 represent observed functional relationships between velocity and dispersivity in certain laboratory experiments as presented by Fried and Combarnous (1971).

We now turn to the *scale effect* in dispersion. As pointed out in Bhattacharya and Gupta (1983), different Gaussian approximations accompanied with in-

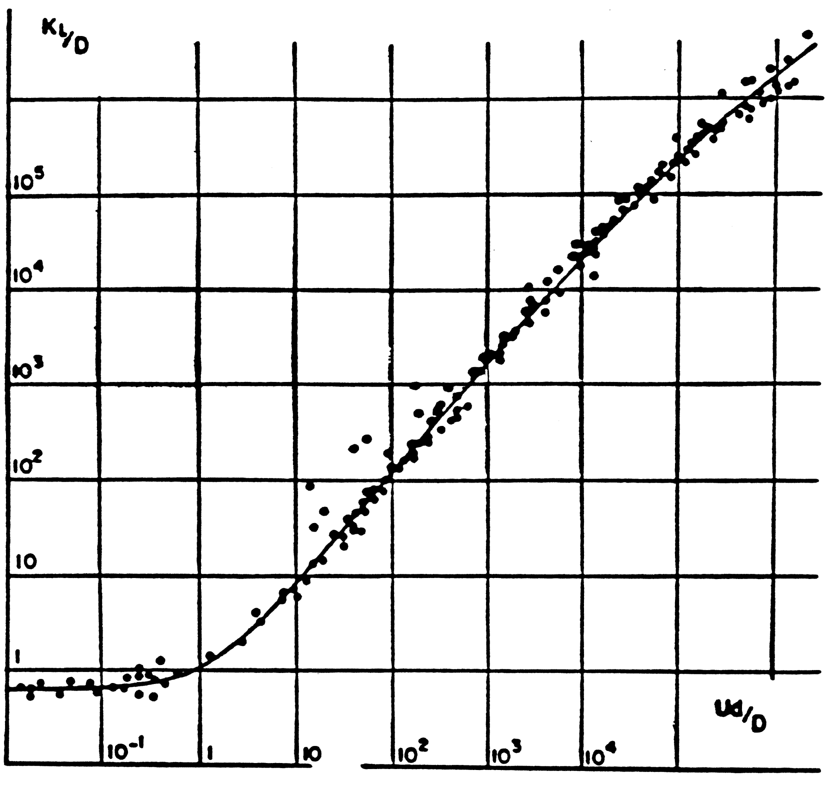

FIG. 1. *Laboratory experiments showing the growth in the dispersion coefficient K_L in the direction of flow with the velocity U. In order to make the coordinates dimensionaless, K_L/D is plotted against the Peclet number Ud/D, where D is the molecular diffusion coefficient and d is the diameter of a typical gain of the porous medium. Taken from Fried and Combarnous* (1971).

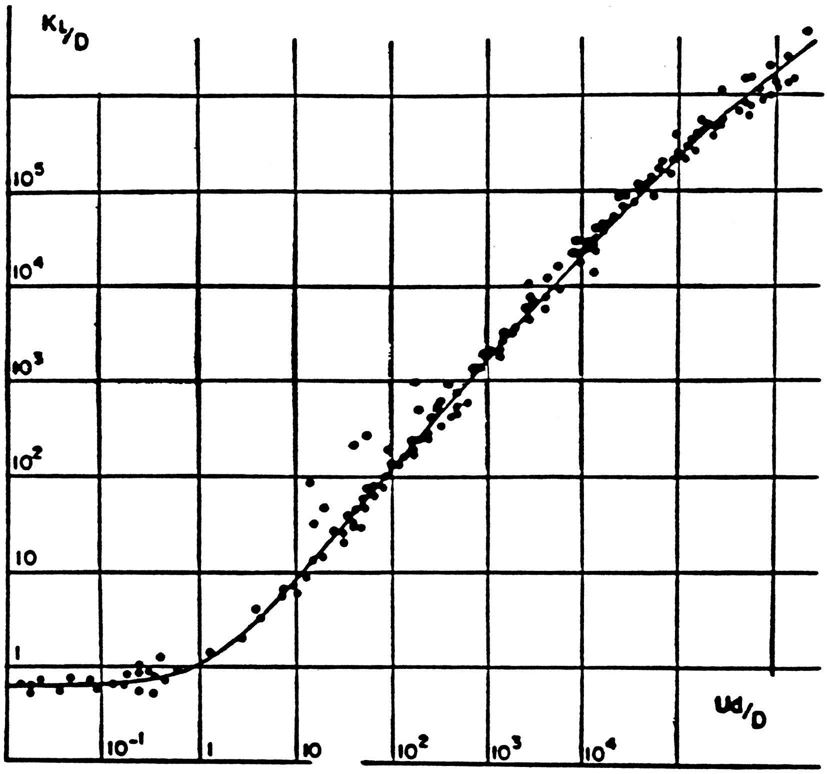

FIG. 2. *Laboratory experiments for the growth in the dispersion coefficient K_T in a direction transverse to the flow. The dotted line represents a fitted curve for K_L/D in the same experiment. Taken from Fried and Combarnous* (1971).

crease in dispersivity at successively larger time scales can only occur in a medium (aquifer) with new heterogeneities appearing at higher scales. To understand this, it is enough to consider two spatial scales of heterogeneity embodied in the flow velocity v,

$$v(y) = b(y) + \beta(y/a), \tag{7.4}$$

where "a" is a large scalar. Here fluctuations in b represent the effect of a local (or small) scale heterogeneity in the aquifer geometry and soil characteristics, while fluctuations in v which manifest only at a larger scale of distance (of the order a) are represented in $\beta(\cdot/a)$. Theorem 2.1 provides precise time scales ($t \ll a^{2/3}$, $t \ll a$ or $t \ll a^{4/3}$) over which the local scale b dominates and large scale fluctuations may be ignored. Here no specific assumptions are needed on $b(\cdot)$ or $\beta(\cdot)$, not even (7.2). The significance of this is that, irrespective of the nature of β, whenever a Gaussian approximation holds for the concentration corresponding to flow velocity $b(\cdot) + \beta(x_0)$ (assuming an initial point injection at ax_0) and diffusion matrix $D(x) \equiv ((D_{jj'}(x)))$, the same holds for the concentration with the actual flow velocity (7.4) and dispersion $D(x)$, provided $t \ll a^{2/3}$ (or $t \ll a$, $t \ll a^{4/3}$, as the case may be). Since Gaussian approximations for a single scale of fluctuations are known to be valid under assumptions of periodicity and almost periodicity of the coefficients, as well as under the assumption of their being an ergodic random field [see Bensoussan, Lions and Papanicolaou (1978); Bhattacharya (1985); Bhattacharya and Ramasubramanian (1988); Gelhar and Axness (1985); Kozlov (1979, 1980); Papanicolaou and Varadhan (1979); Winter, Newman and Neuman (1984)], a Gaussian approximation at the initial phase ($1 \ll t \ll a^{2/3}$) is expected to hold rather broadly in the present context (Theorem 2.2, Remark 2.1.3). Beyond this scale, as the effect of the large scale fluctuations gradually becomes manifest, this initial phase will break down. Under additional assumptions (on β), a different Gaussian approximation takes hold at a larger time scale. The latter approximation, along with its time scale, is provided in Theorems 5.2–5.4 for periodic flows satisfying assumptions (A1)–(A4), (A6) (Theorem 5.2), or (A1)–(A3), (A5), (A7) (Theorems 5.3, 5.4).

Under the hypotheses of Theorems 5.3, 5.4, the dispersivity grows from one constant D_{jj} to a larger constant, so that it is asymptotically a constant. Under the hypothesis of Theorem 5.2, the asymptotic growth in dispersivity is $O(a^2t/t) = O(a^2) = o(t)$ (since $t \gg a^2$ at the larger scale). Thus although dispersivity grows in the latter case, the growth is *sublinear* with time. In between the final and initial phases other *intermediate phases* appear. Examples 6.1, 6.2 in Section 6 illustrate this, along with a precise specification of the time scales for the initial, intermediate and final phases. A computation of dispersivity $d(t)$ in Example 6.1 through all these phases show a mostly sublinear growth,

$$\begin{gathered} d(t) = 1\,(t = O(1)), \qquad d(t) = 1 + \frac{c_1^2}{2\pi^2}(1 \ll t \ll a^{4/3}), \\ 1 \ll d(t) \ll t(a^{4/3} \ll t \ll a^2),\ \ d(t) \ll t(t \gg a^2). \end{gathered} \tag{7.5}$$

For the physical problem at hand, these are examples of multiscale versions of *stratified media* considered in Gupta and Bhattacharya (1986) and Guven and Molz (1986).

Because of the importance of the problem of solute transport in porous media in hydrology and environmental engineering, a number of field studies have been undertaken over the past two decades to monitor solute dispersion in aquifers [see, e.g., Adams and Gelhar (1992); Garabedian, LeBlanc, Gelhar and Celia (1991); LeBlanc Garbedian, Hess, Gelhar, Quadri, Stollenwerk and Wood (1991); Sauty (1980); Sudicky (1986)]. Such experiments are necessarily complex. They require the digging of many properly placed wells to monitor the solute concentration profile, often over a span of several years. The theoretical model most commonly fitted to the data is based on the important work of Gelhar and Axness (1983), where it is assumed that the coefficients of the Fokker–Planck equation governing solute concentration are ergodic random fields. An independent alternative mathematical approach under the same assumptions is given in Winter, Newman and Neuman (1984). Proofs of the validity of the

Gaussian approximation, along with a computation of its dispersion, may be found in Kozlov (1980) and Papanicolaou and Varadhan (1979) for the special case of the generator in divergence form. It seems that for the general case considered by Gelhar and Axness (1983) some mathematical details still need to be worked out, both for the CLT and for the analysis of the dispersion. As shown in Papanicolaou and Varadhan (1979), the periodic and almost periodic cases may be considered as special cases of the ergodic random field model.

The main thrust of the theoretical studies in the hydrology literature on solute dispersion in aquifers has been to explain the scale effect, that is, the increase in dispersivitiy with spatial scale. For example, the dispersivity at the field scales are observed to be larger by orders of magnitude from that at the laboratory scale. As pointed out in Bhattacharya and Gupta (1983), the validity of a hierarchy of Gaussian approximations at the laboratory and field scales, with increase in dispersivity with scale, can only be explained by the presence of multiple scales of heterogeneity in the medium. A single central limit theorem, such as mentioned in the preceding paragraph, cannot explain this phenomenon in a saturated aquifer whose dynamics are independent of time. The points of departure in the present article, following Bhattacharya and Götze (1995), are (1) the explicit introduction of multiple scales of heterogeneity in the velocity field and (2) determination of the time scales for changes from one Gaussian phase to the next. Although it is not claimed here that natural aquifers have periodic velocity fields, the detailed analysis of the periodic case with multiple scales provides a qualitative understanding of the scale effect in dispersion in general. Since under a random translation the periodic velocity field becomes an ergodic random field, the present study also provides an avenue for testing the validity of some of the informal theories and intuition on the nature of multiscale dispersion.

8. Final remarks. In the following series of remarks we mention some unresolved issues and research problems.

REMARK 8.1. The examples in Section 6 show that the time scale for the final Gaussian phase cannot in general be less than $t \gg a^2$ for divergence-free b and β. The additional logarithmic factors $(\log a)^2$ and $\log a$ in Theorems 5.2 and 5.4, respectively, are needed to offset the factor $a^{k/2}$ appearing in Theorem 4.5 in our estimate of the speed of convergence to equilibrium for diffusions on the big torus $\mathscr{T}_a$. We do not know if this factor $a^{k/2}$ can be removed in general. Among important recent methods for the estimation of the speed of convergence to equilibrium of Markov processes we would like to mention those of Chen and Wang (1994, 1997), and Diaconis and Saloff-Coste (1996).

The seemingly excessively large time scale $t \gg a^4(\log a)^2$ in Theorem 5.3 may be reduced to that given in Theorem 5.4, namely, $t \gg a^2 \log a$ if ag_j and $a \operatorname{grad} g_j$ can be shown to be bounded in sup norm rather than in the H^1-norm. We do not know if this is achievable in general.

REMARK 8.2. One may conjecture that the technical condition (A4)j in (3.68) is redundant for the validity of the conclusion of Theorem 5.2. We would also conjecture that, for Theorems 5.3 and 5.4, the assumption of continuity of the derivatives of p_j in (A5)j is redundant.

REMARK 8.3. It is easy to see that the condition that "$a \to \infty$ through integer values" may be relaxed to "$a \to \infty$ through a sequence of rational numbers with a bounded denominator." Can we relax this further in Theorem 5.2? Note that in Example 6.1 in Section 6 no restriction on "a" is needed (except that $a \to \infty$).

REMARK 8.4. As indicated by Theorems 5.5, 5.6 (also see Theorems 4.6, 4.7) for the one-dimensional case, multiscale multidimensional diffusions with periodic nondivergence-free velocity fields offer a rich diversity of behavior that needs to be explored further.

REMARK 8.5. An important problem, both from the point of view of mathematics and that of applications, is the analysis of multiscale diffusions whose coefficients constitute ergodic random fields. Methods employed in this article seem inapplicable for a general asymptotic analysis of such diffusions.

Acknowledgments. The author is grateful to Professor Hari Rajaram for advice on recent experimental studies on solute transport in aquifers and for bringing Adams and Gelhar (1992), Garabedian et al. (1991), LeBlanc et al. (1991) and Sauty (1980) to the author's notice. The author is also much indebted to two anonymous reviewers for their thorough and critical reading of the manuscript and for their helpful suggestions.

REFERENCES

ADAMS, E. E. and GELHAR, L. W. (1992). Field study of dispersion in a heterogeneous aquifer. 2. Spatial moments analysis. *Water Resour. Res.* **28** 3293–3307.

ARONSON, D. G. (1967). Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* **73** 890–896.

BENSOUSSAN, A., LIONS, J. L. and PAPANICOLAOU, G. C. (1978). *Asymptotic Analysis for Periodic Structures.* North-Holland, Amsterdam.

BHATTACHARYA, R. N. (1982). On the functional central limit theorem and the law of iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete.* **60** 185–201.

BHATTACHARYA, R. N. (1985). A central limit theorem for diffusions with periodic coefficients. *Ann. Probab.* **13** 385–396.

BHATTACHARYA, R. N., DENKER, M. and GOSWAMI, A. (1999). Speed of convergence to equilibrium and to normality for diffusions with multiple periodic scales. *Stochastic Process. Appl.* **80** 55–86.

BHATTACHARYA, R. N. and GÖTZE, F. (1995). Time-scales for Gaussian approximation and its break down under a hierarchy of periodic spatial heterogeneities. *Bernoulli* **1** 81–123.

BHATTACHARYA, R. N. and GUPTA, V. K. (1979). On a statistical theory of solute transport in porous media. *SIAM J. Appl. Math.* **37** 485–498.

BHATTACHARYA, R. N. and GUPTA, V. K. (1983). A theoretical explanation of solute dispersion in saturated porous media at the Darcy scale. *Water Resour. Res.* **19** 938–944.

BHATTACHARYA, R. N., GUPTA, V. K. and WALKER, H. F. (1989). Asymptotics of solute dispersion in periodic porous media. *SIAM J. Appl. Math.* **49** 86–98.

BHATTACHARYA, R. N. and RAMASUBRAMANIAN, S. (1988). On the central limit theorem for diffusions with almost periodic coefficients. *Sankhyā Ser. A* **50** 9–25.

BHATTACHARYA, R. N. and WAYMIRE, E. C. (1990). *Stochastic Processes with Applications.* Wiley, New York.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

CHEN, M. F. and WANG, F. Y. (1994). Applications of coupling method to the first eigenvalue on a manifold. *Sci. Sin. (A)* **37** 1–14. (English edition.)

CHEN, M. F. and WANG, F. Y. (1997). Estimates of the logarithmic Sobolev constant. An improvement of Bakry–Emery criterion. *J. Funct. Anal.* **144** 287–300.

CUSHMAN, J. (ed.) (1990). *Dynamics of Fluid in Hierarchical Porous Media.* Academic Press, New York.

DAGAN, G. (1984). Solute transport in heterogeneous porous formations. *J. Fluid Mech.* **145** 151–177.

DIACONIS, P. and STROOCK, D. W. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.

DIACONIS, P. and SALOFF–COSTE, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes. Characterization and Convergence.* Wiley, New York.

FANNJIANG, A. and PAPANICOLAOU, G. (1994). Convection enhanced diffusion for periodic flows. *SIAM J. Appl. Math.* **54** 333–408.

FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.

FRIED, J. J. and COMBARNOUS, M. A. (1971). Dispersion in porous media. *Adv. Hydrosci.* **7** 169–282.

FRIEDMAN, A. (1975). *Stochastic Differential Equations and Applications* **1**. Academic Press, New York.

GARABEDIAN, S. P., LEBLANC, D. R., GELHAR, L. W. and CELIA, M. A. (1991). Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts 2. Analysis of spatial moment for a nonreactive tracer. *Water Resour. Res.* **27** 911–924.

GELHAR, L. W. and AXNESS, C. L. (1983). Three-dimensional stochastic analysis of macrodispersion in aquifers. *Water Resour. Res.* **19** 161–180.

GUPTA, V. K. and BHATTACHARYA, R. N. (1986). Solute dispersion in multidimensional periodic saturated porous media. *Water Resour. Res.* **22** 156–164.

GUVEN, O. and MOLZ. F. J. (1986). Deterministic and stochastic analysis of dispersion in an unbounded stratified porous medium. *Water Resour. Res.* **22** 1565–1574.

HALL, P. G. and HEYDE, C. C. (1980). *Martingale Central Limit Theory and Its Applications.* Academic Press, New York.

HOLLEY, R. A., KUSUOKA, S. and STROOCK, D. W. (1989). Asymptotics of the spectral gap, with applications to simulated annealing. *J. Funct. Anal.* **83** 333–347.

IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland, New York.

KOZLOV, S. M. (1979). Averaging operators with almost periodic, rapidly oscillating coefficients. *Math. USSR–Sb*. **35** 481–498.

KOZLOV, S. M. (1980). Averaging of random operators. *Math. USSR–Sb*. **37** 167–180.

LEBLANC, R. D., GARABEDIAN, S. P., HESS, K. M., GELHAR, L. W., QUADRI, R. D., STOLLENWERK, K. G. and WOOD, W. W. (1991). Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts 1. Experimental design and observed tracer movement. *Water Resour. Res.* **27** 895–910.

NAGAEV, S. V. (1961). More exact statements of limit theorems for homogeneous Markov chains. *Theory Probab. Appl.* **6** 62–81.

PAPANICOLAOU, G. C. and VARADHAN, S. R. S. (1979). Boundary problems with rapidly oscillating coefficients. *Colloq. Math. Soc. János Bolyai* **27** 835–875.

REED, M. and SIMON, B. (1980). *Methods of Modern Mathematical Physics* **1**. *Functional Analysis*, rev. ed. Academic Press, New York.

SAUTY, J. P. (1980). An analysis of hydrodispersive transfer in aquifers. *Water Resour. Res.* **16** 145–158.

SPOSITO, G., JURY, W. A. and GUPTA, V. K. (1986). Fundamental problems in the stochastic convection–dispersion model of solute transport in aquifers and fields. *Water Resour. Res.* **22** 77–99.

SUDICKY, E. A. (1986). A natural gradient experiment on solute transport in a sand aquifer. *Water Resour. Res.* **22** 2069–2082.

TIKHOMIROV, A. N. (1980). On the rate of convergence in the central limit theorem for weakly dependent random variables. *Theory Probab. Appl.* **25** 800–818.

WINTER, C. L., NEWMAN, C. M. and NEUMAN, S. P. (1984). A perturbation expansion for diffusion in a random velocity field. *SIAM J. Appl. Math.* **44** 425–442.

DEPARTMENT OF MATHEMATICS
RAWLES HALL
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405
E-MAIL: bhattach@hamlet.ucs.indiana.edu