J. Fluid Mech. (2000), *vol.* 420, *pp.* 147–200. Printed in the United Kingdom

Passive scalar transport by travelling wave fields

By PETER B. WEICHMAN[1] AND ROMAN E. GLAZMAN[2]

[1]Blackhawk Geometrics, 301 Commercial Road, Suite B, Golden, CO 80401, USA

[2]Jet Propulsion Laboratory 300-323, California Institute of Technology, Pasadena, CA 91109, USA

(Received 20 July 1999 and in revised form 1 May 2000)

We study turbulent transport of passive tracers by random wave fields of a rather general nature. A formalism allowing for spatial inhomogeneity and anisotropy of an underlying velocity field (such as that caused by a latitudinally varying Coriolis parameter) is developed, with the aim of treating problems of large-scale ocean transport by long internal waves. For the special case of surface gravity waves on deep water, our results agree with the earlier theory of Herterich & Hasselmann (1982), though even in that case we discover additional, off-diagonal elements of the diffusion tensor emerging in the presence of a mean drift. An advective diffusion equation including all components of the diffusion tensor $\boldsymbol{D}$ plus a mean, Stokes-type drift $\boldsymbol{u}$ is derived and applied to the case of baroclinic inertia–gravity (BIG) waves. This application is of particular interest for ocean circulation and climate modelling, as the mean drift, according to our estimates, is comparable to ocean interior currents. Furthermore, while on the largest (100 km and greater) scales, wave-induced diffusion is found to be generally small compared to classical eddy-induced diffusion, the two become comparable on scales below 10 km. These scales are near the present limit on the spatial resolution of eddy-resolving ocean numerical models. Since we find that u_z and D_{zz} vanish identically, net vertical transport is absent in wave systems of this type. However, for anisotropic wave spectra the diffusion tensor can have non-zero *off-diagonal* vertical elements, D_{xz} and D_{yz}, and it is shown that their presence leads to non-positive definiteness of $\boldsymbol{D}$, and a negative diffusion constant is found along a particular principal axis. However, the simultaneous presence of a depth-dependent mean horizontal drift $\boldsymbol{u}(z)$ eliminates any potential unphysical behaviour.

1. Introduction

Passive scalar transport by turbulent velocity fields has been a subject of interest to fluid dynamicists for many years (Taylor 1921; Kraichnan 1970; Herterich & Hasselmann 1982; Gawedzki & Kupiainen 1995; Chertkov & Falkovich 1996; Chertkov, Falkovich & Lebedev 1996). The problem is of great importance in ocean and atmosphere dynamics where the transport of heat, moisture, and bio-geochemical quantities has short-term (weather) as well as long-term (climate) implications. Theories of passive scalar transport to date have focused mainly on the effects of vortical flows, as modelled, for example, by the Navier–Stokes equations in three dimensions or the shallow-water equations in two dimensions. The latter are most relevant to large-scale ('quasi-geostrophic') oceanic and atmospheric motions in which the horizontal scales of the motion are much larger than the oceanic or atmospheric depth. However, in addition to vortical motion, there exists a second major class of large-scale motions known as gravity waves. Both surface ('barotropic') and internal ('baroclinic,' also known as baroclinic inertia–gravity (BIG)) waves are detected in most measurements of the velocity, temperature, pressure and other fields (LeBlond & Mysak 1978). This paper is, then, devoted to a study of the possible effects of gravity waves of various types on tracer diffusion and drift.

The kinetic energy of the gravity-wave component is usually small in comparison to that of the vortical component (see e.g. Figures 2.1 and 15.1 in LeBlond & Mysak 1978), and the energy spectum is almost completely dominated by a large peak at frequencies near the local Coriolis (inertial) frequency. Thus, with respect to the much lower-frequency vortical motions (ocean currents, eddies, etc.) which are of primary interest for global ocean and climate studies, the gravity wave mode is generally viewed as high-frequency noise. These motions tend to be only weakly nonlinear, and are then characterized by a well-defined dispersion law, $\omega(\boldsymbol{k})$, which concentrates their spectra on *surfaces* in frequency–wavenumber space. These properties allow, as we shall see, an essentially exact analytical treatment of the passive tracer problem. A crucial fact, that will determine the overall order of magnitude of the transport coefficients, is that the frequency spectra tend to have vanishing weight near zero frequency.

A similar study of the effects of (yet higher frequency) deep-water wind-generated surface gravity waves on tracer diffusion was performed some time ago by Herterich & Hasselmann (1982, hereinafter referred to as HH) who derived the corresponding passive tracer diffusion equation. The well-known phenomenon of *Stokes drift* (Phillips 1977) provides a mechanism by which a monochromatic surface gravity wave induces a small current near the fluid surface. Roughly speaking, if both surface height field $\zeta(\boldsymbol{x}, t)$ (measured from the mean height h_0) and horizontal velocity field $\boldsymbol{v}(\boldsymbol{x}, t)$ near the surface vary sinusoidally, then the product $\boldsymbol{j} = \zeta\boldsymbol{v}$ measures the mass current near the surface and contains a $\sin^2$ term with non-zero mean, and direction of flow along the propagation direction of the wave. If $u_0 = \langle \boldsymbol{v}^2 \rangle^{1/2}$ is the r.m.s. particle velocity, then the resulting mean drift velocity is a factor of order $(u_0/c_0)^2$ smaller than c_0, where c_0 is the group velocity of the wave. The parameter u_0/c_0 is very small for small-amplitude waves, and is therefore a measure of the nonlinearity of the wave. A multichromatic surface gravity wave field also produces a net Stokes drift so long as the wavenumber spectrum is anisotropic. In addition, as shown by HH, the drift has a fluctuating component that leads to diffusion about the mean, and the corresponding diffusion tensor $\boldsymbol{D}$ is non-zero even when the spectrum is isotropic and the mean drift vanishes. Measured in natural units of $u_0^2\tau_0$, where τ_0 is the typical wave period, $\boldsymbol{D}$ is then of the same relative order $(u_0/c_0)^2$, and is therefore small in comparison to the classical eddy-induced turbulent diffusion.

In evaluating the importance of our results in an oceanographic context one must understand how various transport mechanisms operate on different length scales. As the mesh size of computational grids in modern eddy-resolving numerical models of ocean circulation is pushed ever lower, a question arises as to alternative mechanisms of horizontal transport that might become comparable to, or more effective than, the turbulent diffusion due to unresolved 'sub-grid' eddies. According to Richardson's empirical law (Richardson 1926, later derived by Batchelor 1952 and supported by laboratory and field experiments in the ocean: Richardson & Stommel 1948; Stommel 1949; Monin & Ozmidov 1981), the eddy diffusion coefficient $D(L)$ decreases with decreasing eddy size, L, according to

$$D(L) = B\epsilon^{1/3}L^{4/3}, \tag{1.1}$$

where B is a constant of order unity and ϵ is the rate of energy transfer to large scales (induced by the usual inverse cascade in two dimensions). This equation is then valid

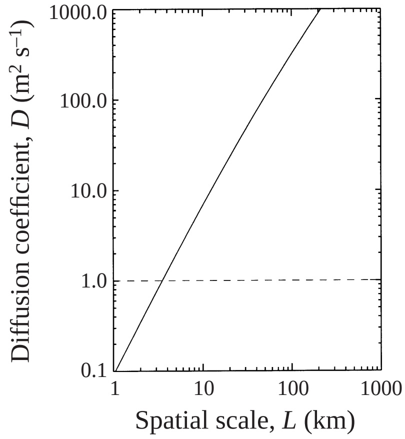

FIGURE 1. Schematic plot of the Richardson laws (1.1) and (1.2) interpolated using the empirical form $D(L) = C_1(L/L_{in})^{4/3}/[1 + C_2(L_{in}/L)^{2/3}]$, with (arbitrarily) $L_{in} = 100$ km, $C_1 = 10^{8/3}\,m^2\,s^{-1}$ in order to obtain $D(10^2\,km) \approx 300\,m^2\,s^{-1}$, and $C_2 = 10^{-1/3}$ in order to obtain $D(1\,km) \sim 0.1\,m^2\,s^{-1}$. These choices yield $\epsilon \sim 10^{-12}\,m^2\,s^{-3}$ if $B = O(1)$ and $\epsilon_\Omega \approx 10^{-21}\,s^{-3}$ if $B' = O(1)$, roughly consistent with oceanographic data. The dashed horizontal line is the estimated diffusion coefficient due to baroclinic inertia–gravity (BIG) waves (see § 6.4).

for length scales larger than the energy input scale, L_{in}. At scales below L_{in} the kinetic energy spectrum is controlled by the direct cascade of enstrophy (squared vorticity), and one has in place of (1.1),

$$D(L) = B'\epsilon_\Omega^{1/3} L^2, \tag{1.2}$$

where ϵ_Ω is the rate of enstrophy transfer to small scales. In figure 1 we show a schematic plot of (1.1) and (1.2), interpolated empirically and using physically motivated parameter choices as described in the caption. According to the Richardson law, the value of $D(L)$ used in these models to account for 'sub-grid' motions on scales shorter than 10 km should then be of order $10\,m^2\,s^{-1}$. For comparison, the molecular diffusion coefficient for clean water is of order $10^{-9}\,m^2\,s^{-1}$. Our estimates presented in § 6.4 indicate that the BIG wave-induced diffusion constant may attain $1\,m^2\,s^{-1}$, and hence may become an important, if not dominant, factor in horizontal turbulent diffusion on short scales, $L \ll 10$ km. One of the questions we address is: as L is pushed ever lower, would the actual diffusion eventually approach the molecular diffusion limit as eddies of all sizes are accounted for explicitly in an 'ideal' numerical model? Or, are there other significant contributions to D that must be accounted for? As we show in the present work, wave-induced diffusion, although usually small by comparison with eddy-induced transport, may well set a lower bound on turbulent diffusion in natural environments.

In § 2 we define the passive scalar problem formally. In §§ 3 and 4 a random walk representation is used to obtain formal Lagrangian expressions for the diffusion parameters. In § 5 the smallness of u_0/c_0 is used to derive explicit expressions for the diffusion parameters in terms of the Eulerian wavenumber–frequency spectrum. Our main result, equation (5.31), provides a very general, quantitatively accurate estimate for the diffusion coefficient, valid for any linear or weakly nonlinear statistically isotropic or anisotropic wave field. Explicit calculations and comparisons for the case of oceanic BIG waves (with length scales ranging from 10 to 1000 km) are performed in § 6. The results of HH for surface gravity waves (with scales ranging from 1 to

300 m) are rederived in § 7. However, in addition to the horizontal transport found by HH, in both these examples we find that for anisotropic spectra, in the presence of a non-zero mean drift, the diffusion tensor may also have non-zero *off-diagonal vertical elements*, D_{xz} and D_{yz}, though the diagonal vertical element D_{zz} vanishes identically. The latter implies that there is *no* net vertical transport, while the former leads to a non-positive-definite diffusion tensor, i.e. a *negative* diffusion coefficient along a particular principal axis. In § 8 we discuss these issues in detail, arguing that the absence of net vertical transport should be a general result for these types of wave systems, and showing that negative diffusion is permitted in the presence of a depth-varying mean drift and leads to no mathematical singularities in the diffusion equation. Two Appendices deal with details of mathematical derivations that would otherwise interrupt the flow of the paper.

2. Passive scalar dynamics and statistics

Our analysis begins from the passive scalar transport equation which takes the general form of a conservation law

$$\partial_t \psi + \nabla \cdot \boldsymbol{j}_\psi = 0, \tag{2.1}$$

where $\psi(\boldsymbol{x}, t)$ is the passive scalar (salinity, heat, phytoplankton, etc.) concentration field, and $\boldsymbol{j}_\psi$ is the conserved passive scalar current. The most common form for the current is $\boldsymbol{j}_\psi = \boldsymbol{v}\psi - \kappa \nabla \psi$, with microscopic diffusion constant κ and advecting velocity field $\boldsymbol{v}(\boldsymbol{x}, t)$. We will be concerned with cases in which turbulent transport induced by $\boldsymbol{v}$ is many orders of magnitude larger than that induced by microscopic diffusion, and so we simply set $\kappa = 0$ henceforth and study the equation

$$\partial_t \psi + \nabla \cdot (\boldsymbol{v}\psi) = 0. \tag{2.2}$$

Often the advecting velocity field $\boldsymbol{v}(\boldsymbol{x}, t)$, is incompressible, satisfying $\nabla \cdot \boldsymbol{v} = 0$. In this case (2.2) then takes the more standard form

$$\partial_t \psi + \boldsymbol{v} \cdot \nabla \psi = 0. \tag{2.3}$$

For general compressible $\boldsymbol{v}$, equations (2.2) and (2.3) are not equivalent, and are in fact adjoints of each other. The two solutions, which are then equal only when $\nabla \cdot \boldsymbol{v} = 0$, will be contrasted and compared further in § 3 below. Although the full three-dimensional velocity field is indeed incompressible, we will often be interested in an effective projected discription in which only horizontal components of the motion are considered, and $\boldsymbol{v}$ is two-dimensional. Oscillations of the fluid surface then make the effective $\boldsymbol{v}$ compressible, and we will consider this more general case in what follows. The velocity will be taken to be given *a priori* with Gaussian statistics and stationary two-point correlation matrix

$$C_{ij}(\boldsymbol{x}, \boldsymbol{x}', t - t') = \langle v_i(\boldsymbol{x}, t) v_j(\boldsymbol{x}', t') \rangle \tag{2.4}$$

(subscripts $i, j = 1, \ldots, d$ label the Cartesian components). For fixed $\boldsymbol{x}, \boldsymbol{x}'$ it will be assumed that C_{ij} decays exponentially in time on the scale of a decorrelation time τ which is a few times the dominant wave period τ_0. Similarly, for fixed $t - t'$, C_{ij} decays exponentially in $\Delta \boldsymbol{x} \equiv \boldsymbol{x} - \boldsymbol{x}'$ on a scale ξ which is a few times the dominant wavelength, λ_0. The argument $t - t'$ implies that we deal here only with stationary processes. For a process that is homogeneous, or translation invariant, along a subspace of directions $\boldsymbol{r}(= (x, y)$ for horizontal translation invariance along the ocean surface), only the difference $\boldsymbol{r} - \boldsymbol{r}'$ will appear in place of separate dependence on $\boldsymbol{r}$

and $\boldsymbol{r}'$. We will always consider situations where, at worst, the explicit dependence on the centre-of-mass variable $\boldsymbol{X} \equiv \frac{1}{2}(\boldsymbol{x}+\boldsymbol{x}')$ is slow on the scale of ξ. Thus we consider situations where the probability distribution of $\boldsymbol{v}$ changes only very slowly with $\boldsymbol{X}$, giving rise to slowly varying effective diffusion and drift parameters. An equivalent statement is that locally defined drift and diffusion parameters make sense only so long as the decorrelation time τ of C_{ij} is much smaller than the time T for a tracer to be transported a distance $L \gg \xi$ over which the $\boldsymbol{X}$-dependence of C_{ij} is significant. Typically, the major $\boldsymbol{X}$-dependence will be on the *depth* of the observation point. Due to their evanescent character, for surface gravity waves this dependence is on the scale of the dominant wavelength of the waves. For baroclinic inertia–gravity waves, whose wavelength is much larger than the ocean depth, this dependence is on the scale of the thermocline depth. Horizontal $\boldsymbol{X}$-dependence, coming, for example, from variations in wind patterns over the surface, or latitude dependence of the wave dispersion relation, is generally on the scale of many dominant wavelengths. Similar relaxation of the stationarity assumption, via generalizations to slow variations in $\bar{t} \equiv \frac{1}{2}(t+t')$, are clearly also straightforward in this same kind of approximation. In all that follows we will account explicitly only for the slow inhomogeneities in the vertical direction.

Since, in the absence of damping and strong nonlinear interactions between waves, a travelling wave field mode μ will have a well-defined dispersion law $\omega_\mu(\boldsymbol{k})$, the velocity field may in general be decomposed in the form

$$\boldsymbol{v}(\boldsymbol{x},t) = \sum_\mu \int \frac{\mathrm{d}^2k}{(2\pi)^2}\,[a_\mu(\boldsymbol{k})\hat{\boldsymbol{e}}_\mu(\boldsymbol{k};z)\,\mathrm{e}^{\mathrm{i}[\boldsymbol{k}\cdot\boldsymbol{r}-\omega_\mu(\boldsymbol{k})t]} + \text{c.c.}], \tag{2.5}$$

where c.c. stands for complex conjugate, and for purposes of generality we are allowing for the existence of an arbitrary number of modes $\mu = 1,2,3,\ldots$ with amplitude $a_\mu(\boldsymbol{k})$. The vertical profile $\hat{\boldsymbol{e}}_\mu(\boldsymbol{k},z)$ decays exponentially with depth for surface gravity waves, but in general could have more complicated non-monotonic behaviour throughout the depth of the fluid. One may normalize the vertical profiles so that, say, $\int_{-H_0}^{0} \mathrm{d}z\hat{\boldsymbol{e}}_\mu(\boldsymbol{k},z)\cdot\hat{\boldsymbol{e}}_\nu(\boldsymbol{k},z)^* = \delta_{\mu\nu}$, where $z=0$ is the fluid surface and H_0 is its depth.

In order to simplify the notation we have assumed in (2.5) that wave excitations propagate in two dimensions, $\boldsymbol{r}=(x,y)$, and have some non-propagating structure in the third dimension. This case is appropriate to all of the oceanographic examples treated later in this paper, but may be trivially extended to more general problems in which a d-dimensional space of $\boldsymbol{x}=(\boldsymbol{r},\boldsymbol{z})$ is divided into a $\hat{d} \leqslant d$ dimensional 'horizontal' subspace of $\boldsymbol{r}=(x_1,x_2,\ldots,x_{\hat{d}})$ (and hence $\boldsymbol{k}=(k_1,k_2,\ldots,k_{\hat{d}})$) containing the wave excitations and a $\bar{d}=d-\hat{d}$ dimensional 'vertical' subspace of $\boldsymbol{z}=(x_{\hat{d}+1},x_{\hat{d}+2},\ldots,x_d)$ containing the wave profile $\boldsymbol{e}_\mu(\boldsymbol{k},\boldsymbol{z})$. For surface gravity and baroclinic inertia–gravity waves, treated in §§6 and 7, one has $\hat{d}=2$ and $\bar{d}=1$. The important problem of waves in the β-plane, which we plan to treat in future work, is characterized by a wave amplitude decaying exponentially with latitude away from the equator. Therefore, for this problem, $\hat{d}=1$ and $\bar{d}=2$. Acoustic waves in d-dimensions (typically $d=2$ or 3) have $\hat{d}=d$ and $\bar{d}=0$, an example that will be treated in §§5.5 and 5.6.

The important property of (2.5) is that the spatio-temporal spectrum is restricted to certain *surfaces* in the three-dimensional $(\boldsymbol{k},\omega)$-space. For example, for inertia–gravity waves one has $\omega(\boldsymbol{k})=\sqrt{f^2+(kc)^2}$, where the Coriolis parameter $f=2\Omega\sin(\phi)$, Ω is the Earth's rotation frequency, ϕ the latitude, and c is the phase speed of sufficiently short waves for which the Coriolis force is negligible. The amplitudes $a_\mu(\boldsymbol{k})$ are taken to be independent Gaussian random variables:

$$\langle a_\mu(\boldsymbol{k}) a_\nu^*(\boldsymbol{k}')\rangle = \hat{f}_\mu(\boldsymbol{k})\delta_{\mu\nu}(2\pi)^2\delta(\boldsymbol{k}-\boldsymbol{k}') \tag{2.6}$$

where $\hat{f}_\mu(\boldsymbol{k})$ is real and positive, and physically is strongly peaked at some characteristic wavevector k_0 and decays fairly rapidly on either side. The appearance of the δ-function constraint implies that we are considering here only the homogeneous case: lack of translation invariance would lead to a broadening of the δ-function, or, equivalently, a residual dependence of $\hat{f}_\mu$ on $\boldsymbol{X}$. As discussed above, we will always compute local quantities ignoring this horizontal $\boldsymbol{X}$-dependence. The Fourier transform of $C_{ij}(\boldsymbol{r}-\boldsymbol{r}', t-t'; z, z')$ is given by

$$\left.\begin{aligned}
\Phi_{ij}(\boldsymbol{k},\omega;z,z') &\equiv \int \mathrm{d}^2r \int \mathrm{d}t C_{ij}(\boldsymbol{r},t;z,z')\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} \\
&= \sum_\mu \{F_{ij}^\mu(\boldsymbol{k};z,z')2\pi\delta[\omega-\omega_\mu(\boldsymbol{k})] \\
&\quad + F_{ji}^\mu(-\boldsymbol{k};z',z)2\pi\delta[\omega+\omega_\mu(-\boldsymbol{k})]\} \\
&= \Phi_{ji}(-\boldsymbol{k},-\omega;z',z) = \Phi_{ij}(-\boldsymbol{k},-\omega;z,z')^*, \\
F_{ij}^\mu(\boldsymbol{k},z,z') &\equiv \hat{f}_\mu(\boldsymbol{k})\hat{e}_{\mu,i}(\boldsymbol{k},z)\hat{e}_{\mu,j}(\boldsymbol{k},z')^* = F_{ji}^\mu(\boldsymbol{k},z',z)^*.
\end{aligned}\right\} \tag{2.7}$$

Nonlinear interactions between waves will broaden the frequency delta functions here. One finds in the case of inertia–gravity waves, for example, a broadening corresponding to a nonlinear decorrelation time $\tau_{\mathrm{nl}} \gg \tau$ of order 10–30 wave periods (Glazman 1996*b*). This time would be exhibited in the decay of C_{ij} along space–time trajectories $|\boldsymbol{x}-\boldsymbol{x}'| \propto c_0|t-t'|$ which travel with the dominant wave speed. This will become important for passive scalar *correlations* at separations less than 10–30 wavelengths, but is not important for the diffusion constant. HH have also shown, at least in the case of surface gravity waves, that the nonlinear additions to (2.5) do not contribute to the transport coefficients to leading non-trivial order.

Given the above characterization of the statistics of the velocity field, $\boldsymbol{v}$, we would now like to compute the statistics of the passive scalar field, ψ. This paper will be concerned with deriving an equation of motion for the average, $\bar{\psi}(\boldsymbol{x},t) \equiv \langle\psi(\boldsymbol{x},t)\rangle$. We shall see that under certain conditions a diffusion equation emerges:

$$\partial_t\bar{\psi}(\boldsymbol{x},t) + \nabla\cdot[\boldsymbol{u}(\boldsymbol{x})\bar{\psi}(\boldsymbol{x},t)] = \nabla\cdot[\mathbf{D}(\boldsymbol{x})\cdot\nabla\bar{\psi}(\boldsymbol{x},t)]. \tag{2.8}$$

Here $\boldsymbol{u}(\boldsymbol{x})$ is a local steady-state drift velocity, and $\mathbf{D}(\boldsymbol{x})$ is a local diffusion tensor. In terms of the scale-dependent $D(L)$ discussed in the Introduction, this equation should be viewed as describing diffusion *relative* to the frame of reference of any finite overall drift due to much larger-scale eddy motions. In the examples we treat we find (in disagreement with HH) that $\mathbf{D}$ has non-zero components in both the horizontal and vertical directions, though because $D_{zz} \equiv 0$ no actual net vertical transport results. This feature will be discussed in detail in § 8.

3. Random walk representation

The computations are based on the following random walk representation for $\psi(\boldsymbol{x},t)$ (Monin & Yaglom 1971; Piterbarg 1997). Let $\boldsymbol{Z}_{\boldsymbol{x}t}(s)$ be the Lagrangian trajectory of a particle freely advected by the flow that is constrained to be at the point $\boldsymbol{x}$ at time

t, and hence satisfying the equation

$$\partial_s \boldsymbol{Z}_{xt}(s) = \boldsymbol{v}(\boldsymbol{Z}_{xt}(s), s) \tag{3.1}$$

with the boundary condition $\boldsymbol{Z}_{xt}(t) = \boldsymbol{x}$. The subscripts on $\boldsymbol{Z}_{xt}(s)$ therefore define the unique fluid particle that passes through the point $\boldsymbol{x}$ at time t. The point from which this particle began its motion at time zero is then $\boldsymbol{Z}_{xt}(0)$, and at time s this particle will be (or was) at point $\boldsymbol{Z}_{xt}(s)$. It is then straightforward to show that a formal solution to the passive scalar equation (2.2) is

$$\begin{aligned}\psi(\boldsymbol{x}, t) &= \int \mathrm{d}^3x' \psi(\boldsymbol{x}', s)\delta[\boldsymbol{x} - \boldsymbol{Z}_{x's}(t)] \\ &= \int \mathrm{d}^3x' \psi(\boldsymbol{x}', s)\det[\partial \boldsymbol{Z}_{x's}(t)/\partial \boldsymbol{x}']^{-1}\delta[\boldsymbol{x}' - \boldsymbol{Z}_{xt}(s)] \\ &= \psi[\boldsymbol{Z}_{x,t}(s), s]\det[\partial \boldsymbol{Z}_{xt}(s)/\partial \boldsymbol{x}],\end{aligned} \tag{3.2}$$

for any $t > s$. This follows, by virtue of (3.1), from the fact that $\partial_t \delta[\boldsymbol{x} - \boldsymbol{Z}_{x',s}(t)] = -\boldsymbol{v}(\boldsymbol{Z}_{x',s}(t), t)\cdot\nabla\delta[\boldsymbol{x} - \boldsymbol{Z}_{x',s}(t)] = -\nabla\cdot\{\boldsymbol{v}(\boldsymbol{x}, t)\delta[\boldsymbol{x} - \boldsymbol{Z}_{x',s}(t)]\}$. The rules for manipulating derivatives of delta-functions must be followed meticulously here. The Jacobian factor makes up for local compression and dilatation of the mapping $\boldsymbol{x} \to \boldsymbol{Z}_{xt}(s)$ which maps the distribution at time t back to that at time s (and is then the inverse of the mapping $\boldsymbol{x}' \to \boldsymbol{Z}_{x's}(t)$ from the distribution at time s to that at time t). Given an initial condition $\psi_0(\boldsymbol{x}) \equiv \psi(\boldsymbol{x}, 0)$, one has the special case

$$\begin{aligned}\psi(\boldsymbol{x}, t) &= \int \mathrm{d}^3x' \psi_0(\boldsymbol{x}')\delta[\boldsymbol{x} - \boldsymbol{Z}_{x'0}(t)] \\ &= \psi_0[\boldsymbol{Z}_{xt}(0)]\det[\partial \boldsymbol{Z}_{xt}(0)/\partial \boldsymbol{x}].\end{aligned} \tag{3.3}$$

It is worth noting that the factor

$$\phi(\boldsymbol{x}, t) \equiv \psi_0[\boldsymbol{Z}_{xt}(0)] \tag{3.4}$$

satisfies (2.3) even for compressible $\boldsymbol{v}$. This may be verified explicitly by noting that if $\phi(\boldsymbol{x}, t)$ satisfies (2.3), then $(\mathrm{d}/\mathrm{d}s)\phi(\boldsymbol{Z}_{xt}(s), s) = 0$, and hence that $\phi[\boldsymbol{Z}_{xt}(s), s] = \phi[\boldsymbol{Z}_{xt}(s'), s']$ for any s, s'. Setting $s = t$ and $s' = 0$ yields (3.4). However, because (2.3) is not in the form of a conservation law if $\boldsymbol{v}$ is compressible, ψ is not a conserved density. Taking as an intial condition $\phi(\boldsymbol{x}, s) = x_i$ to be any one of the Cartesian coordinates, one obtains the useful relation, valid for any $\boldsymbol{v}$,

$$\partial_t \boldsymbol{Z}_{xt}(s) + [\boldsymbol{v}(\boldsymbol{x}, t)\cdot\nabla]\boldsymbol{Z}_{xt}(s) = 0. \tag{3.5}$$

The $\boldsymbol{x}, t$ dependence of $\boldsymbol{Z}_{xt}(s)$, at any fixed s, therefore satisfies the adjoint equation (2.3). This gives a convenient relation between t- and $\boldsymbol{x}$-derivatives of $\boldsymbol{Z}_{xt}(s)$ that complements the s-derivative given in (3.1). Intuitively, (3.5) states that a particle at $\boldsymbol{x}$ at time $t+\mathrm{d}t$ must have been at $\boldsymbol{x}-\boldsymbol{v}(\boldsymbol{x}, t)\,\mathrm{d}t$ at time t, i.e. that $\boldsymbol{Z}_{x,t+\mathrm{d}t}(s) = \boldsymbol{Z}_{x-v(x,t)\,\mathrm{d}t,t}(s)$ define precisely the same trajectory.

Equation (3.1) may also be written in the integral form

$$\boldsymbol{Z}_{xt}(s) = \boldsymbol{x} + \int_t^s \mathrm{d}s' \boldsymbol{v}(\boldsymbol{Z}_{xt}(s'), s'), \tag{3.6}$$

which makes the boundary condition more explicit.

If $\boldsymbol{v}$ is incompressible, then the two mappings discussed below (3.2) are area preserving and the Jacobian is simply unity. The formal solution to (2.3) is then (3.4), and the fields ϕ and ψ are exactly the same. The concentration at $(\boldsymbol{x}, t)$ is therefore now completely determined by the initial concentration at the point at time zero which evolves into $(\boldsymbol{x}, t)$ under the flow.

4. Statistics of the concentration field

We will be concerned with deriving an equation of motion for $\bar{\psi}(\boldsymbol{x}, t) = \langle \psi(\boldsymbol{x}, t) \rangle$, where the average is over the ensemble of velocity fields $\boldsymbol{v}$ defined by (2.4). From (2.2) we have

$$\partial_t \bar{\psi} + \nabla \cdot \langle \boldsymbol{j}_\psi \rangle = 0, \qquad \langle \boldsymbol{j}_\psi \rangle = \langle \psi \boldsymbol{v} \rangle. \tag{4.1}$$

If ψ has spatial variations only on a scales l much larger than the correlation length ξ (but much smaller than the homogeneity length L), then we expect to be able to perform a gradient expansion

$$\langle \boldsymbol{j}_\psi \rangle = \boldsymbol{u}\bar{\psi} - (\boldsymbol{D} \cdot \nabla)\bar{\psi} + (\boldsymbol{E} : \nabla\nabla)\bar{\psi} + \cdots, \tag{4.2}$$

where $\boldsymbol{u}(\boldsymbol{x})$ is a vector, and $\boldsymbol{D}(\boldsymbol{x})$, $\boldsymbol{E}(\boldsymbol{x}), \ldots$ are second, third, etc. rank tensors, respectively. The notation $(\boldsymbol{E} : \nabla\nabla)_i = \sum_{j,k} E_{ijk} \partial_j \partial_k$ is used here. All may be slow functions of $\boldsymbol{x}$. One obtains then the diffusion equation (2.8) if one keeps terms only to linear order in the spatial derivatives. Using the Lagrangian representation, formal expressions for these quantities will now be derived. All considerations in this section and the following one are general, and no particular properties of $\boldsymbol{v}$ associated with waves are used.

4.1. *Markov property*

From (3.3) one obtains

$$\bar{\psi}(\boldsymbol{x}, t) = \int \mathrm{d}^3 x' P(\boldsymbol{x}, t | \boldsymbol{x}', 0) \psi_0(\boldsymbol{x}') \tag{4.3}$$

in which the transition probability is given by

$$P(\boldsymbol{x}, t | \boldsymbol{x}', s) = \langle \delta[\boldsymbol{x} - \boldsymbol{Z}_{\boldsymbol{x}'s}(t)] \rangle. \tag{4.4}$$

The object is now to convert (4.3) into an equation of motion for $\bar{\psi}$, i.e. into one that relates $\bar{\psi}$ to itself at two nearby times. Note that directly averaging (3.2) for s near t does not work because $\psi(\boldsymbol{x}', s)$ itself is a random function that is correlated with $\boldsymbol{Z}_{\boldsymbol{x}'s}(t)$. To accomplish the purpose one must make use of the decorrelation time τ: on time scales greater than τ the velocity field $\boldsymbol{v}$ loses memory of its previous history. Unfortunately this does not mean that $\boldsymbol{Z}_{\boldsymbol{x}'s}(t)$ becomes a Markov process on time scales larger than τ. A Markov process, by definition, has no memory: the statistics of its future evolution depend only on its present position, and are independent of the particular path taken to get to its present position. In order for this property to hold, $\boldsymbol{Z}_n \equiv \boldsymbol{Z}_{\boldsymbol{x}'s}(s + n\tau)$, $n = 0, 1, 2, \ldots$, would have to have the character of a random walk with independent increments $\boldsymbol{W}_n \equiv \boldsymbol{Z}_n - \boldsymbol{Z}_{n-1}$, $n = 1, 2, 3, \ldots$. This is clearly not necessarily the case because $\boldsymbol{v}$ has zero mean, and the leading behaviour of $\boldsymbol{Z}_{\boldsymbol{x}'s}(t)$ is to oscillate about the starting point $\boldsymbol{x}$, irrespective of the value of τ. The Markov property only emerges on time scales t_D where the residual diffusive part of $\boldsymbol{Z}_{\boldsymbol{x}'s}(t)$ accumulates a net motion larger than the typical distance $d_0 = u_0 \tau$ travelled during a decoherence time. This is especially true for waves, which are inherently oscillatory in nature, where we shall find that t_D is at least an order of magnitude larger than τ. This is less true for Navier–Stokes turbulence in which one finds t_D and τ to be of the same order. We shall see that the key distinctions between the two lie in the high degree of nonlinearity of the latter, and in the suppression of the low-frequency part of the velocity frequency spectrum in the former.

In the following discussion we make no assumption about the size of t_D relative to τ. The key observation that allows one to take advantage of finite τ is that a suitable moving average allows one to resolve the diffusive part of $\boldsymbol{Z}_{x's}(t)$ on time scales greater than τ even when $\tau \ll t_D$. Thus, let $\phi(t) \geqslant 0$ be a smooth (e.g. Gaussian) weight with $\int_{-\infty}^{\infty} \phi(t)\,\mathrm{d}t = 1$ and width larger than, but of the same order as, τ, and let

$$\left.\begin{aligned} \bar{\boldsymbol{Z}}_{x's}(t) &= \int \mathrm{d}t'\phi(t')\boldsymbol{Z}_{x's}(t+t'), \\ \boldsymbol{\eta}_{x's}(t) &= \boldsymbol{Z}_{x's}(t) - \bar{\boldsymbol{Z}}_{x's}(t). \end{aligned}\right\} \tag{4.5}$$

Since the smoothing eliminates from $\boldsymbol{Z}_{x's}(t)$ the dominant oscillations $\boldsymbol{\eta}_{x's}(t)$ over a time scale larger than that over which they are correlated, one expects its residual motions to reflect only the underlying diffusion process. Thus $\bar{\boldsymbol{Z}}_{x's}(t)$ *is expected to be a Markov process* on time scales larger than τ, in the sense that its increments $\bar{\boldsymbol{W}}_n$ (defined analogously to $\boldsymbol{W}_n$ above) should be independent. In addition, the oscillatory part $\boldsymbol{\eta}_{x's}(t)$ is bounded and essentially independent of $\bar{\boldsymbol{Z}}_{x's}(t)$, and may therefore be thought of as an additive random noise, correlated over a time scale τ, acting on the underlying Markov process. To the extent that the statistics of $\boldsymbol{v}$ vary only on a length scale $L_0 \gg d_0$, the statistics of $\bar{\boldsymbol{Z}}_{x's}(t)$ will be independent of $t-s$ for $\tau < t-s < T_0$, where $T_0 \gg t_0$ is the time required for $\bar{\boldsymbol{Z}}_{x's}(t)$ to travel a distance of order L_0.

The object of the analysis now is to derive the underlying transport properties of $\bar{\boldsymbol{Z}}_{x's}(t)$. The following theorem allows us to do this in a convenient way directly from the probability P defined in (4.3) and (4.4).

Let $\bar{\boldsymbol{Y}}_{x's}(t)$ be a stationary, homogeneous Markov process with transition probability

$$\Pi(\boldsymbol{x},t|\boldsymbol{x}',s) = \langle \delta(\boldsymbol{x} - \bar{\boldsymbol{Y}}_{x's}(t))\rangle, \tag{4.6}$$

which is then actually a function only of $\boldsymbol{x}-\boldsymbol{x}'$ and $t-s$. Let $\boldsymbol{\theta}_{x's}(t)$ be stationary, homogeneous and independent of $\bar{\boldsymbol{Y}}_{xs}(t)$, and let $\boldsymbol{\theta}_{x's}(t)$ be independent of $\boldsymbol{\theta}_{x's}(t')$ for $|t-s| > \tau_\theta$, where τ_θ is the assumed finite decoherence time of the 'noise' process $\boldsymbol{\theta}$. Let $\boldsymbol{Y} = \bar{\boldsymbol{Y}} + \boldsymbol{\theta}$ have transition function $\mathscr{P}$, defined analogously to P in (4.3). The independence and translation invariance conditions imply that

$$\begin{aligned} \mathscr{P}(\boldsymbol{x},t|\boldsymbol{x}',0) &= \langle \delta(\boldsymbol{x} - \bar{\boldsymbol{Y}}_{x'0}(t) - \boldsymbol{\theta}_{x'0}(t))\rangle \\ &= \langle \Pi(\boldsymbol{x} - \boldsymbol{\theta}_{x'0}(t), t|\boldsymbol{x}',0)\rangle \\ &= \langle \Pi(\boldsymbol{x}, t|\boldsymbol{x}' + \boldsymbol{\theta}_{x'0}(t),0)\rangle. \end{aligned} \tag{4.7}$$

The Markov property implies that the function Π obeys for any $s < t$ the Chapman–Kolmogorov identity

$$\Pi(\boldsymbol{x},t|\boldsymbol{x}',0) = \int \mathrm{d}^3y\,\Pi(\boldsymbol{x},t|\boldsymbol{y},s)\Pi(\boldsymbol{y},s|\boldsymbol{x}',0). \tag{4.8}$$

If $\boldsymbol{Y}_{x's}(t)$ has a finite memory time τ (which need not in general be related to τ_η), validity of (4.8) requires $t-s > \tau$. One obtains then from (4.7)

$$\begin{aligned} \mathscr{P}(\boldsymbol{x},t|\boldsymbol{x}',0) &= \int \mathrm{d}^3y\,\Pi(\boldsymbol{x},t|\boldsymbol{y},s)\langle \Pi(\boldsymbol{y},s|\boldsymbol{x}' + \boldsymbol{\theta}_{x'0}(t),0)\rangle \\ &= \int \mathrm{d}^3y\,\Pi(\boldsymbol{x},t|\boldsymbol{y},s)\langle \Pi(\boldsymbol{y} - \boldsymbol{\theta}_{x'0}(t),s|\boldsymbol{x}',0)\rangle. \end{aligned} \tag{4.9}$$

Finally we use homogeneity and independence to obtain,

$$\mathscr{P}(\boldsymbol{x}, t|\boldsymbol{x}', 0) = \int \mathrm{d}^3 y \Pi(\boldsymbol{x}, t|\boldsymbol{y}, s)\langle \Pi(\boldsymbol{y} - \boldsymbol{\theta}_{\boldsymbol{x}'0}(s), s|\boldsymbol{x}', 0)\rangle$$
$$= \int \mathrm{d}^3 y \Pi(\boldsymbol{x}, t|\boldsymbol{y}, s)\mathscr{P}(\boldsymbol{y}, s|\boldsymbol{x}', 0), \tag{4.10}$$

in which the replacement of $\boldsymbol{\theta}_{\boldsymbol{x}'0}(t)$ by $\boldsymbol{\theta}_{\boldsymbol{x}'0}(s)$ is valid so long as $t, s > \tau$ so that memory of the starting time $t = 0$ is lost. Equation (4.10) is the fundamental result we seek.

We make now the key assumption that the elements $\bar{\boldsymbol{Z}}$ and $\boldsymbol{\eta}$ in decomposition (4.5) of $\boldsymbol{Z}$ indeed have the properties, at least in some minimal asymptotic sense, required of $\bar{\boldsymbol{Y}}$ and $\boldsymbol{\theta}$ in the theorem. A rigorous proof (which we do not at present have) of the validity of this key assumption, i.e. a specification and proof of whatever minimal asymptotic criterion is needed, would be required to establish our formalism rigorously. Such a property, for example, must allow for the fact that $\bar{\boldsymbol{Z}}$ and $\boldsymbol{\eta}$ as defined in (4.5) are not fully independent. Weak violations of translation invariance should also be allowed so long as one has a broad separation between the inhomogeneity time scale T_0 and the decorrelation time scale τ.

The key assumption implies that we may insert the transition probability (4.4) in place of $\mathscr{P}$ in (4.10). Inserting the result into (4.3), we infer the existence of a Markov transition probability Π such that

$$\bar{\psi}(\boldsymbol{x}, t) = \int \mathrm{d}^3 y \Pi(\boldsymbol{x}, t|\boldsymbol{y}, s)\bar{\psi}(\boldsymbol{y}, s), \quad t > s > \tau. \tag{4.11}$$

If $\bar{\boldsymbol{Z}}_{\boldsymbol{x}'s}(t)$ has finite memory time τ, validity of (4.11) again requires $t - s > \tau$ as well. Thus, from the key assumption follows the result that after a transient time τ, the subsequent evolution of the mean concentration field is Markovian with a transition probability determined entirely in terms of the Markovian part of $\boldsymbol{Z}_{\boldsymbol{x}'s}(t)$. The 'noise' part $\boldsymbol{\eta}_{\boldsymbol{x}'s}(t)$, no matter how large, affects the evolution of $\bar{\psi}(\boldsymbol{x}, t)$ only for $t < \tau$. Notice that (4.11) represents a kind of factorization of the average of the exact relation (3.2),

$$\bar{\psi}(\boldsymbol{x}, t) = \int \mathrm{d}^3 y \langle \delta(\boldsymbol{x} - \boldsymbol{Z}_{\boldsymbol{y}s}(t))\psi(\boldsymbol{y}, s)\rangle, \tag{4.12}$$

in which the theorem above shows that the 'noise' $\boldsymbol{\eta}$ should simply be dropped before performing the factorization.

In the explicit calculations to follow we will make use of (4.10) by computing the time derivative of $P(\boldsymbol{x}, t|\boldsymbol{x}', 0)$ for $t > \tau$. The theorem above shows that this time derivative is completely determined by Π. This will be confirmed explicity by showing that the short-time oscillatory dynamics indeed disappear for $t > \tau$, and expressing the time derivative as a differential operator, representing the infinitesimal generator of Π, acting on the $\boldsymbol{x}$-dependence of $P(\boldsymbol{x}, t|\boldsymbol{x}', 0)$. Through (4.3), the result is the requisite diffusion equation for $\bar{\psi}$.

The fact that all computations go through consistently, at least for the lowest few orders in perturbation theory in the small parameter u_0/c_0, is strong support for the validity of the assumption and leads us to believe that rigorous criteria can indeed be established. For completeness, however, we note the following, possibly related, problem. Recent attempts to confirm the existence of wave-induced diffusion by numerical simulation (Balk & McLaughlin 1999) have succeeded in two horizontal dimensions, but not in one dimension (though Stokes drift in one dimension has been confirmed). This points either to subtle numerical problems in one dimension, or to a breakdown in our mathematical formalism (which predicts finite diffusion in one dimension). We have not yet attempted to reproduce the numerical results, and so can offer no insight there. It would be interesting to attempt to construct exact solutions in one dimension. If there are, for example, subtle non-perturbative, non-Markovian correlations in one dimension that lead to a breakdown of the key assumption above, such a solution could provide much needed insight. Since all of the physical examples treated in this paper are in two dimensions where there appears to be no problem, we will leave such an investigation to future work.

4.2. *Formal expression for transport parameters*

It is useful to consider the Fourier representation

$$P(\boldsymbol{x},t|\boldsymbol{x}',0) = \int \frac{d^3K}{(2\pi)^3} e^{i\boldsymbol{K}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \langle e^{-i\boldsymbol{K}\cdot[\boldsymbol{Z}_{\boldsymbol{x}'0}(t)-\boldsymbol{x}']}\rangle = \int \frac{d^3K}{(2\pi)^3} e^{i\boldsymbol{K}\cdot(\boldsymbol{x}-\boldsymbol{x}')-\lambda(\boldsymbol{K};\boldsymbol{x}',t)}, \tag{4.13}$$

in which $\boldsymbol{K}$ is a full three-dimensional wavevector, and

$$\lambda(\boldsymbol{K};\boldsymbol{x},t) = -\ln\langle e^{-i\boldsymbol{K}\cdot[\boldsymbol{Z}_{\boldsymbol{x}0}(t)-\boldsymbol{x}]}\rangle = -\sum_{n=1}^{\infty}\frac{1}{n!}\sum_{l_1,l_2,\ldots,l_n}\lambda_{l_1 l_2\ldots l_n}(\boldsymbol{x},t)(-iK_{l_1})(-iK_{l_2})\ldots(-iK_{l_n}). \tag{4.14}$$

Defining $\Delta\boldsymbol{Z}_{\boldsymbol{x}s}(t) = \boldsymbol{Z}_{\boldsymbol{x}s}(t) - \boldsymbol{x}$, the coefficients in the expansion, which by construction are symmetric in all of their indices, are given by the usual cummulant averages:

$$\lambda_l^{(1)}(\boldsymbol{x},t) = \langle \Delta Z_{\boldsymbol{x}0}^l(t)\rangle, \tag{4.15}$$

$$\lambda_{lm}^{(2)}(\boldsymbol{x},t) = \langle \Delta Z_{\boldsymbol{x}0}^l(t)\Delta Z_{\boldsymbol{x}0}^m(t)\rangle - \langle \Delta Z_{\boldsymbol{x}0}^l(t)\rangle\langle \Delta Z_{\boldsymbol{x}0}^m(t)\rangle, \tag{4.16}$$

and so on. The first of these is interpreted as producing a systematic drift (which would vanish for isotropic flows). If $\Delta\boldsymbol{Z}_{\boldsymbol{x}0}(t)$ were (for fixed $\boldsymbol{x}$) purely Gaussian, then all higher-order terms would vanish identically. Even when the field $\boldsymbol{v}(\boldsymbol{x},t)$ is Gaussian, however, the transformation to Lagrangian coordinates brings in non-Gaussian corrections. For isotropic flows one has $\lambda_{lm}^{(2)} = \lambda^{(2)}\delta_{lm}$, and $\lambda^{(2)}(\boldsymbol{x},t) = \frac{1}{3}\langle|\Delta\boldsymbol{Z}_{\boldsymbol{x}0}(t)|^2\rangle$. More generally, for flows isotropic in $\hat{d} < d$ directions, $\lambda^{(2)}$ will be proportional to the $\hat{d}\times\hat{d}$ identity matrix when the indices l and m are restricted to those directions. This diffusion term represents therefore the mean-square distance travelled by a Lagrangian particle in time t, relative to any systematic drift. In general $\lambda_{lm}^{(2)}$ is not simply a multiple of the identity matrix. However, it may always be diagonalized by an orthogonal transformation to yield a set of principal axes whose eigenvalues correspond to the (different) rates of diffusion along these axes. This general case has important geophysical applications, for example to flows in the β-plane (where latitudinal variations of the Coriolis force are taken into account). In the translation-invariant case, $\lambda_{lm}^{(2)}$ and $\lambda_l^{(1)}$ will be independent of the position $\boldsymbol{x}$.

One obtains now from (4.3) and (4.14),

$$\partial_t\bar{\psi}(\boldsymbol{x},t) = -\int\frac{d^3K}{(2\pi)^3}\int d^3x'\,\partial_t\lambda(\boldsymbol{K};\boldsymbol{x}',t)\,e^{i\boldsymbol{K}\cdot(\boldsymbol{x}-\boldsymbol{x}')-\lambda(\boldsymbol{K};\boldsymbol{x}',t)}\psi_0(\boldsymbol{x}'). \tag{4.17}$$

We shall see explicitly below that under rather general conditions

$$\rho(\boldsymbol{K};\boldsymbol{x}) \equiv \partial_t\lambda(\boldsymbol{K};\boldsymbol{x},t) \tag{4.18}$$

is independent of t for $t > \tau$. It is this result that confirms the theorem proven in the previous subsection. The expansion (4.14) for λ then implies an expansion

$$\rho(\boldsymbol{K};\boldsymbol{x}) = -\sum_{n=1}^{\infty} \frac{1}{n!} \sum_{l_1,l_2,\ldots,l_n} \rho^{(n)}_{l_1 l_2 \ldots l_n}(\boldsymbol{x})(-\mathrm{i}K_{l_1})(-\mathrm{i}K_{l_2})\ldots(-\mathrm{i}K_{l_n}), \tag{4.19}$$

in which, from (3.1),

$$\left.\begin{aligned} \rho^{(1)}_l(\boldsymbol{x}) &= \langle v_l(\boldsymbol{Z}_{x0}(t),t)\rangle, \\ \rho^{(2)}_{lm}(\boldsymbol{x}) &= \langle v_l(\boldsymbol{Z}_{x0}(t),t)\Delta Z^m_{x0}(t)\rangle - \langle v_l(\boldsymbol{Z}_{x0}(t),t)\rangle\langle \Delta Z^m_{x0}(t)\rangle + (l \leftrightarrow m), \end{aligned}\right\} \tag{4.20}$$

and so on. These coefficients are also symmetric in all indices. We obtain then

$$\begin{aligned} \partial_t \bar{\psi}(\boldsymbol{x},t) &= -\sum_{n=1}^{\infty} \frac{1}{n!} \sum_{l_1,l_2,\ldots,l_n} \int \mathrm{d}^3x' \int \frac{\mathrm{d}^3K}{(2\pi)^3} \mathrm{e}^{\mathrm{i}\boldsymbol{K}\cdot(\boldsymbol{x}-\boldsymbol{x}')-\lambda(\boldsymbol{K};\boldsymbol{x}',t)} \psi_0(\boldsymbol{x}') \\ &\quad \times \rho^{(n)}_{l_1 l_2 \ldots l_n}(\boldsymbol{x}')(-\mathrm{i}K_{l_1})(-\mathrm{i}K_{l_2})\ldots(-\mathrm{i}K_{l_n}) \\ &= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \sum_{l_1,l_2,\ldots,l_n} \partial_{l_1}\partial_{l_2}\ldots\partial_{l_n} \\ &\quad \times \int \mathrm{d}^3x' P(\boldsymbol{x},t|\boldsymbol{x}',0)\psi_0(\boldsymbol{x}')\rho^{(n)}_{l_1 l_2 \cdots l_n}(\boldsymbol{x}'), \end{aligned} \tag{4.21}$$

in which the products of wavector components have been eliminated via the replacement $\mathrm{i}K_l \to \partial_l$ and (4.13) has then been used to eliminate the $\boldsymbol{K}$-integration. We clearly may now identify the averaged tracer current (4.1):

$$\langle j^l_\psi \rangle = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n!} \sum_{l_2,\ldots,l_n} \partial_{l_2}\ldots\partial_{l_n} \int \mathrm{d}^3x' P(\boldsymbol{x},t|\boldsymbol{x}',0)\psi_0(\boldsymbol{x}')\rho^{(n)}_{l l_2 \ldots l_n}(\boldsymbol{x}'). \tag{4.22}$$

We finally make use of the assumption that the spatial variation of the cummulants $\rho^{(n)}(\boldsymbol{x}')$ is very slow on the scale of the dependence of P on $\boldsymbol{x}-\boldsymbol{x}'$. We may then, to an excellent approximation, replace $\boldsymbol{x}'$ by $\boldsymbol{x}$ in $\rho^{(n)}$ to obtain the local form

$$\langle j^l_\psi \rangle = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n!} \sum_{l_2,\ldots,l_n} \partial_{l_2}\cdots\partial_{l_n}[\rho^{(n)}_{l l_2 \ldots l_n}(\boldsymbol{x})\bar{\psi}(\boldsymbol{x},t)], \tag{4.23}$$

in which (4.3) has been used to eliminate the $\boldsymbol{x}'$-integration. We have now explicitly derived the gradient expansion (4.2), and the diffusion equation (2.8) now follows with the explicit identifications

$$\left.\begin{aligned} u_l(\boldsymbol{x}) &= \rho^{(1)}_l(\boldsymbol{x}) - \frac{1}{2}\sum_m \partial_m \rho^{(2)}_{lm}(\boldsymbol{x}) + \frac{1}{6}\sum_{m,n} \partial_m\partial_n \rho^{(3)}_{lmn}(\boldsymbol{x}) \mp \cdots, \\ D_{lm}(\boldsymbol{x}) &= \tfrac{1}{2}\rho^{(2)}_{lm}(\boldsymbol{x}) - \frac{1}{6}\sum_n \partial_n \rho^{(3)}_{lmn}(\boldsymbol{x}) + \frac{1}{24}\sum_{n,p} \partial_n\partial_p \rho^{(4)}_{lmnp}(\boldsymbol{x}) \mp \cdots, \\ E_{lmn} &= \tfrac{1}{6}\rho^{(3)}_{lmn}(\boldsymbol{x}) - \frac{1}{24}\sum_p \partial_p \rho^{(4)}_{lmnp}(\boldsymbol{x}) \pm \cdots, \end{aligned}\right\} \tag{4.24}$$

and so on. It will be seen below that $\rho^{(n)} = O(u_0^n \tau^{n-1})$ so that these expansions terminate at any given order in u_0/c_0.

To summarize, then, one has

$$\left.\begin{aligned}\rho^{(1)}_l(\boldsymbol{x}) &= \langle v_l(\boldsymbol{Z}_{x0}(t),t)\rangle,\\ \rho^{(2)}_{lm}(\boldsymbol{x}) &= \int_0^t \mathrm{d}t'[G^{(2)}_{lm}(\boldsymbol{x},t,t') + G^{(2)}_{ml}(\boldsymbol{x},t,t')],\\ \rho^{(3)}_{lmn}(\boldsymbol{x}) &= \int_0^t \mathrm{d}t' \int_0^t \mathrm{d}t''[G^{(3)}_{lmn}(\boldsymbol{x},t,t',t'') + G^{(3)}_{lmn}(\boldsymbol{x},t',t,t'') + G^{(3)}_{lmn}(\boldsymbol{x},t',t'',t)],\end{aligned}\right\} \quad (4.25)$$

in which it should be recalled that $t > \tau$, and, using (3.6), we have defined the single-particle two-time and three-time Lagrangian correlation tensors based at $\boldsymbol{x}, s$,

$$\left.\begin{aligned}G^{(2)}_{lm}(\boldsymbol{x},t-s,t'-s) &= G^{(2)}_{ml}(\boldsymbol{x},t'-s,t-s)\\ &= \langle v_l(\boldsymbol{Z}_{xs}(t),t)v_m(\boldsymbol{Z}_{xs}(t'),t')\rangle\\ &\quad -\langle v_l(\boldsymbol{Z}_{xs}(t),t)\rangle\langle v_m(\boldsymbol{Z}_{xs}(t'),t')\rangle,\\ G^{(3)}_{lmn}(\boldsymbol{x},t-s,t'-s,t''-s) &= \langle v_l(\boldsymbol{Z}_{xs}(t),t)v_m(\boldsymbol{Z}_{xs}(t'),t')v_n(\boldsymbol{Z}_{xs}(t'),t'')\rangle_c,\end{aligned}\right\} \quad (4.26)$$

in which the subscript c indicates that appropriate products of lower-order averages should be subtracted. These expressions make clear the required symmetry of the $\rho^{(n)}$ under permutation of their indices. These tensors become functions only of $t-t'$, $t-t''$, etc., only if all of $t-s, t'-s, t''-s, \ldots > \tau$. The integration in (4.25), however, is dominated by $t'-s < \tau$. In this latter range no such simplification occurs due to strong correlations between $\boldsymbol{Z}_{xs}(t')$ and $\boldsymbol{v}(\boldsymbol{Z}_{xs}(t'),t')$ (it is these same correlations, for example, that make the average in (4.24) defining $\boldsymbol{u}$ non-zero). These expressions will form the basis of the systematic computation of $\boldsymbol{u}$ and $\boldsymbol{D}$ that follow.

5. Systematic computation of turbulent transport coefficients

The aim now is to compute $\boldsymbol{u}$ and $\boldsymbol{D}$ systematically. The computation is based on the exact solution to the problem in which $\boldsymbol{v}(\boldsymbol{Z}_{xs}(t),t) \approx \boldsymbol{v}(\boldsymbol{x},t)$ is taken to vary so slowly in space that the dependence on the Lagrangian coordinate may be dropped. In essence perturbation theory will be performed about the limit in which $\boldsymbol{v}$ is spatially uniform, though fluctuating in time. This limit is appropriate because $u_0 \ll c_0$: the tracer particle travels a distance $d_0 \ll \lambda_0$ much smaller than the dominant wavelength λ_0 over a decorrelation time τ, and therefore explores the spatial dependence of $\boldsymbol{v}(\boldsymbol{x},t)$ only on scales over which it is essentially spatially (but not temporally) constant.

5.1. *Expansion of Lagrangian quantities in terms of Eulerian quantities*

For reasons that will become clear below, the basic starting point is the adjoint equation (2.3). To simplify the notation, let us write this equation in the operator form

$$\partial_t \phi = \mathrm{L}\phi, \quad (5.1)$$

in which, for the purposes of the following derivation, the linear operator $\mathrm{L}(\boldsymbol{x},t)$ may be quite general, but will eventually be replaced by $\mathrm{L} = -\boldsymbol{v}\cdot\nabla$ appropriate to (2.3). In integral form, (5.1) then reads

$$\phi(\boldsymbol{x},t) = \phi(\boldsymbol{x},s) + \int_s^t \mathrm{d}s_1 \mathrm{L}(s_1)\phi(\boldsymbol{x},s_1). \quad (5.2)$$

One may now iterate this equation:

$$\begin{aligned}\phi(\boldsymbol{x},t) &= \phi(\boldsymbol{x},s) + \int_s^t \mathrm{d}s_1 \mathrm{L}(s_1)\phi(\boldsymbol{x},s) + \int_s^t \mathrm{d}s_1 \int_s^{s_1} \mathrm{d}s_2 \mathrm{L}(s_1)\mathrm{L}(s_2)\phi(\boldsymbol{x},s_2) \\ &= \cdots = \sum_{n=0}^{\infty} \int_s^t \mathrm{d}s_1 \int_s^{s_1} \mathrm{d}s_2 \ldots \int_s^{s_{n-1}} \mathrm{d}s_n \mathrm{L}(s_1)\mathrm{L}(s_2) \ldots \mathrm{L}(s_n)\phi(\boldsymbol{x},s). \quad (5.3)\end{aligned}$$

We have therefore generated the usual time-ordered product expansion with the formal solution

$$\begin{aligned}\phi(\boldsymbol{x},t) &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_s^t \mathrm{d}s_1 \int_s^t \mathrm{d}s_2 \ldots \int_s^t \mathrm{d}s_n \mathrm{T}_t[\mathrm{L}(s_1)\mathrm{L}(s_2)\ldots\mathrm{L}(s_n)]\phi(\boldsymbol{x},s) \\ &\equiv \boldsymbol{U}(\boldsymbol{x};t,s)\phi(\boldsymbol{x},s), \quad (5.4)\end{aligned}$$

in which the evolution operator may be written formally as

$$\begin{aligned}\boldsymbol{U}(\boldsymbol{x};t,s) &\equiv \mathrm{T}_t \exp\left[\int_s^t \mathrm{d}s' \mathrm{L}(s')\right] \\ &= \lim_{M\to\infty} \left[\prod_{N=0}^{M-1} \mathrm{e}^{\Delta t \mathrm{L}(s+N\Delta t)}\right], \quad \Delta t \equiv \frac{t-s}{M}, \quad (5.5)\end{aligned}$$

where T_t is the time-ordering operator which puts larger times to the left. The product form on the second line of (5.5) makes it clear that $\boldsymbol{U}$ obeys the semigroup property $\boldsymbol{U}(t_3,t_2)\boldsymbol{U}(t_2,t_1) = \boldsymbol{U}(t_3,t_1)$ for any $t_3 \geqslant t_2 \geqslant t_1$. The basic expression (5.3) will suffice for most of our purposes, however. Explicitly, one then has

$$\begin{aligned}\phi(\boldsymbol{x},t) = \sum_{n=0}^{\infty}(-1)^n \int_s^t \mathrm{d}s_1 \int_s^{s_1} \mathrm{d}s_2 \ldots \int_s^{s_{n-1}} \mathrm{d}s_n [\boldsymbol{v}(\boldsymbol{x},s_1)\cdot\nabla][\boldsymbol{v}(\boldsymbol{x},s_2)\cdot\nabla] \ldots \\ \times[\boldsymbol{v}(\boldsymbol{x},s_n)\cdot\nabla]\phi(\boldsymbol{x},s), \quad (5.6)\end{aligned}$$

in which the gradients operate on all $\boldsymbol{x}$-dependence to the right.

The key observation now is that, by (3.5), the Lagrangian trajectory has the formal Eulerian expansion

$$\begin{aligned}\boldsymbol{Z}_{xt}(s) - \boldsymbol{x} = \sum_{n=1}^{\infty}(-1)^n \int_s^t \mathrm{d}s_1 \int_s^{s_1} \mathrm{d}s_2 \ldots \int_s^{s_{n-1}} \mathrm{d}s_n [\boldsymbol{v}(\boldsymbol{x},s_1)\cdot\nabla][\boldsymbol{v}(\boldsymbol{x},s_2)\cdot\nabla] \ldots \\ \times[\boldsymbol{v}(\boldsymbol{x},s_{n-1})\cdot\nabla]\boldsymbol{v}(\boldsymbol{x},s_n), \quad (5.7)\end{aligned}$$

in which the identities $\boldsymbol{Z}_{x,s}(s) = \boldsymbol{x}$ and $[\boldsymbol{v}(\boldsymbol{x},s_n)\cdot\nabla]\boldsymbol{x} = \boldsymbol{v}(\boldsymbol{x},s_n)$ have been used. This equation is valid for both $s > t$ and $s < t$. Assuming the former, then interchanging the roles of s and t, and reparameterizing the domain of integration from $s \leqslant s_1 \leqslant s_2 \leqslant \cdots \leqslant s_n \leqslant t$ to $s \leqslant s_n \leqslant s_{n-1} \leqslant \cdots \leqslant s_1 \leqslant t$ one may write this in the more convenient form

$$\begin{aligned}\boldsymbol{Z}_{xs}(t) - \boldsymbol{x} &= \sum_{n=1}^{\infty} \int_s^t \mathrm{d}s_1 \int_s^{s_1} \mathrm{d}s_2 \ldots \int_s^{s_{n-1}} \mathrm{d}s_n [\boldsymbol{v}(\boldsymbol{x},s_n)\cdot\nabla][\boldsymbol{v}(\boldsymbol{x},s_{n-1})\cdot\nabla] \ldots \\ &\quad \times[\boldsymbol{v}(\boldsymbol{x},s_2)\cdot\nabla]\boldsymbol{v}(\boldsymbol{x},s_1) \\ &\equiv \int_s^t \mathrm{d}s_1 \boldsymbol{U}^{-1}(\boldsymbol{x};s_1,s)\boldsymbol{v}(\boldsymbol{x},s_1), \quad (5.8)\end{aligned}$$

in which the inverse of the evolution operator is given formally by

$$U^{-1}(x;t,s) \equiv \mathrm{T}_{-t} \exp\left[-\int_s^t \mathrm{d}s' \mathrm{L}(s')\right]$$

$$= \lim_{M\to\infty} \left[\prod_{N=1}^{M} \mathrm{e}^{-\Delta t \mathrm{L}(t-N\Delta t)}\right], \quad \Delta t \equiv \frac{t-s}{M}, \tag{5.9}$$

in which T_{-t} puts *smaller* times to the left. The product form makes it clear that $U^{-1}(t,s)$ is indeed the inverse of $U(t,s)$. More generally, one then also has $U^{-1}(t_3,t_2)$ $U(t_3,t_1) = U(t_2,t_1)$ and $U(t_3,t_1)U^{-1}(t_2,t_1) = U(t_3,t_2)$ for any $t_3 \geqslant t_2 \geqslant t_1$. Finally, using (3.1), one obtains the corresponding expansion for the Lagrangian velocity:

$$v(Z_{xs}(t),t) = \sum_{n=0}^{\infty} \int_s^t \mathrm{d}s_1 \int_s^{s_1} \mathrm{d}s_2 \ldots \int_s^{s_{n-1}} \mathrm{d}s_n [v(x,s_n)\cdot\nabla][v(x,s_{n-1})\cdot\nabla]\ldots$$

$$\times[v(x,s_1)\cdot\nabla]v(x,t)$$

$$\equiv U^{-1}(x;s,t)v(x,t). \tag{5.10}$$

Equation (5.10) is the basic result that we require for input into the expressions (4.25) and (4.26). This equation is again completely general: no specific properties of waves have been used. In the case of waves, however, one may at this point consider the following heuristic estimate of the size of the terms in this expansion. Since $v = O(u_0)$, and since v varies on the scale of the dominant wavelength, λ_0, one will have $\nabla v = O(u_0/\lambda_0)$. The time-integration is over an interval of order τ, so that $\int_s^t \mathrm{d}s' \nabla v(s') = O(u_0\tau/\lambda_0) = O(u_0\tau/c_0\tau_0)$, where τ_0 is the dominant wave period, and $c_0 = O(\lambda_0/\tau_0)$ is the dominant phase speed of the waves. Assuming $\tau/\tau_0 = O(1)$ (the decorrelation time is a few wave periods), we find that the nth term in (5.8) is $O[u_0(u_0/c_0)^n]$. This estimate will be confirmed explicitly below. The expansion is thus well controlled for small u_0/c_0, as mentioned in the Introduction. For travelling waves, u_0 is proportional to the wave amplitude, while c_0 is fixed by the dispersion relation and is independent of amplitude. Thus u_0/c_0 is a measure of the nonlinearity of the wave field, and is indeed small, typically of order 0.1, under the conditions of weak turbulence normally encountered in the ocean. For velocity fields generated by eddy turbulence there is no dispersion relation, and c_0 is not defined. Conditions of strong turbulence generally operate and one finds that all of the terms in the expansion (5.10) are of the same order. The perturbation theory presented below is then inapplicable in that case and other approximation schemes must be developed (Chertkov & Falkovich 1996; Chertkov *et al.* 1996; Gawedzki & Kupiainen 1995).

5.2. *Lowest-order results*

Consider first the lowest-order results for the $\rho^{(n)}$. According to (5.7), one simply replaces $Z_{xs}(t)$ by x everywhere in the arguments of the Lagrangian velocity. Since v is assumed Gaussian, the Lagrangian position in this approximation will also be Gaussian, and all cummulants of order three and higher vanish identically. One then obtains

$$\left.\begin{aligned} \rho_l^{(1)}(\boldsymbol{x}) &= \langle v_l(\boldsymbol{x},t)\rangle = 0, \\ \rho_{lm}^{(2)}(\boldsymbol{x}) &= \int_0^t \mathrm{d}t'[\langle v_l(\boldsymbol{x},t)v_m(\boldsymbol{x},t')\rangle + \langle v_m(\boldsymbol{x},t)v_l(\boldsymbol{x},t')\rangle] \\ &= \int_0^t \mathrm{d}s_1[g_{lm}(\boldsymbol{x},s_1) + g_{ml}(\boldsymbol{x},s_1)], \\ \rho^{(n)} &\equiv 0, \quad n \geqslant 3. \end{aligned}\right\} \tag{5.11}$$

in which $g_{lm}(\boldsymbol{x}, t-t') = g_{ml}(\boldsymbol{x}, t'-t) \equiv C_{lm}(\boldsymbol{x}, \boldsymbol{x}, t-t')$ is the lowest-order estimate of $G_{lm}^{(2)}(\boldsymbol{x}, t-s, t'-s)$ and is in fact independent of s. The mean drift therefore vanishes, and the lowest-order diffusion tensor is given by the Kubo-type formula,

$$D_{lm}^{(0)}(\boldsymbol{x}) = \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}s_1 g_{lm}(\boldsymbol{x},s_1) = \tfrac{1}{2}\hat{S}_{lm}(\boldsymbol{x},0), \tag{5.12}$$

where the fact that the argument of the integrand essentially vanishes for $s > \tau$ allows the extension of the integration to infinity (explicitly demonstrating, at least at this order, the independence of the result of the precise choice of $t-s$), and the frequency spectrum is given by

$$\hat{S}_{lm}(\boldsymbol{x},\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t} g_{lm}(\boldsymbol{x},t). \tag{5.13}$$

In the case of wave systems one has the general result that the frequency spectrum vanishes identically in a neighbourhood of zero frequency. In the case of BIG waves, with dispersion relation $\omega(\boldsymbol{k}) = \sqrt{f^2 + (ck)^2}$, the Coriolis parameter f provides a low-frequency cutoff to the spectrum. For surface gravity waves, with dispersion relation $\omega(\boldsymbol{k}) = \sqrt{gk}$, the frequency vanishes only when the wavenumber does. However, in natural environments the longest waves directly generated by wind travel at or below the wind speed U. The corresponding resonant wavenumber, based on the gravity wave dispersion law, is then $k_0 \approx U^2/g$, and the corresponding wave frequency is $\omega_0 \approx g/U$. While nonlinear wave–wave interactions do transfer some wave energy from this 'generation range' up the spectrum to yet lower wavenumbers and frequencies, this transfer has its intrinsic limitations which are ultimately responsible for a finite low-frequency cutoff. Thus $\hat{S}_{lm}(\omega)$ again vanishes in a neighbourhood of (or, at the very least, vanishes extremely rapidly near) $\omega = 0$. One obtains therefore the result that $\boldsymbol{D}^{(0)} = 0$, and the process appears to be subdiffusive. Physically this result arises from the fact that a superposition of perfectly periodic motions can lead to net motion only on time scales less than the maximum period in the system. If the spectrum of periods is strictly bounded from above, equivalent to the statement that $\hat{S}_{lm}(\omega \to 0) \to 0$, $\boldsymbol{D}^{(0)}$ must vanish: no diffusion occurs on time scales larger than the decorrelation time τ. It is also this result that makes the time scale t_D defined in §4.1 much larger than τ: it expresses the fact that the oscillatory motion is very inefficient at generating net transport, which then becomes visible only at higher order in u_0/c_0, and hence only on time scales much larger than τ.

5.3. *Perturbation expansion for the mean drift: Stokes drift*

In order to obtain finite estimates for the transport coefficients, slow spatial variation in $\boldsymbol{v}(\boldsymbol{x}, t)$ must be taken into account. This can generate drift and diffusion because the periodic motions at neighbouring points will be slightly out of synchrony and the orbit of a fluid particle will no longer close. Such higher-order corrections are accounted for precisely by the higher-order terms in the expansion (5.10).

Consider first the mean drift. To first order one has, using (4.25) and (5.10),

$$\begin{aligned}\rho_l^{(1)}(\boldsymbol{x}) &= \int_0^t \mathrm{d}s_1 \langle [\boldsymbol{v}(\boldsymbol{x}, s_1) \cdot \nabla] v_l(\boldsymbol{x}, t) \rangle \\ &= \lim_{\boldsymbol{x}' \to \boldsymbol{x}} \sum_m \int_0^t \mathrm{d}s_1 \partial_m' C_{lm}(\boldsymbol{x}', \boldsymbol{x}; s_1) \\ &= \sum_m \int \frac{\mathrm{d}\omega}{2\pi} \frac{\hat{R}_{lm}^m(\boldsymbol{x}, \omega)}{\omega - \mathrm{i}\eta},\end{aligned} \tag{5.14}$$

in which $\eta \to 0^+$ is a convergence factor and

$$\hat{R}_{lm}^i(\boldsymbol{x}, \omega) \equiv -\mathrm{i} \lim_{\boldsymbol{x}' \to \boldsymbol{x}} \int_{-\infty}^{\infty} \mathrm{d}s\, \mathrm{e}^{\mathrm{i}\omega s} \partial_i' C_{lm}(\boldsymbol{x}', \boldsymbol{x}; s). \tag{5.15}$$

The standard identity $(\omega \mp \mathrm{i}\eta)^{-1} = P(1/\omega) \pm \mathrm{i}\pi\delta(\omega)$, where P denotes principal value, together with the odd and even combinations

$$\hat{R}_{lm}^{i\pm}(\boldsymbol{x}, \omega) \equiv \tfrac{1}{2}[\hat{R}_{lm}^i(\boldsymbol{x}, \omega) \pm \hat{R}_{lm}^i(\boldsymbol{x}, -\omega)], \tag{5.16}$$

and the first line of (4.24) yields

$$u_l(\boldsymbol{x}) = \sum_m \left[\frac{\mathrm{i}}{2} \hat{R}_{lm}^m(\boldsymbol{x}, 0) + \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi\omega} \hat{R}_{lm}^{m-}(\boldsymbol{x}, \omega) - \tfrac{1}{2}\partial_m \hat{S}_{lm}(\boldsymbol{x}, 0) \right] [1 + O(u_0/c_0)], \tag{5.17}$$

where the last term is the contribution from $\rho^{(2)}$. For waves only the second term survives (and one may in fact drop the $-$ superscript as convergence of the integral at $\omega = 0$ is no longer an issue), and using the Fourier representation (2.7) one obtains

$$\begin{aligned}u_l^{(1)}(z) &= \lim_{z' \to z} \sum_{m=1}^{2} \sum_\mu \int \frac{\mathrm{d}^2 k}{(2\pi)^2} \frac{1}{\omega_\mu(\boldsymbol{k})} \left[(\boldsymbol{k}, -\mathrm{i}\partial_z')_m F_{lm}^\mu(\boldsymbol{k}; z', z) + (\boldsymbol{k}, \mathrm{i}\partial_z')_m F_{ml}^\mu(\boldsymbol{k}; z', z) \right], \\ &= \sum_\mu \int \frac{\mathrm{d}^2 k}{(2\pi)^2} \frac{1}{\omega_\mu(\boldsymbol{k})} \left\{ \sum_{m=1}^{2} k_m [F_{lm}^\mu(\boldsymbol{k}; z, z) + F_{ml}^\mu(\boldsymbol{k}; z, z)] \right. \\ &\quad \left. - \mathrm{i} \lim_{z' \to z} \partial_z [F_{lz}^\mu(\boldsymbol{k}; z, z') - F_{zl}^\mu(\boldsymbol{k}; z', z)] \right\}.\end{aligned} \tag{5.18}$$

Notice that vertical gradients in the velocity correlations give contributions to the horizontal drift. In a projected description in which all z, z' dependence is ignored, only the first term in the integrand would be present. In the case of surface gravity waves we will see that the second term actually doubles the magnitude of the drift. Intuitively this occurs because 'vertical' gradients enhance the small imbalance between the forward and backward motions during a wave cycle.

Equation (5.18) represents (to lowest non-trivial order) precisely the net Stokes drift due to a spectrum of waves (HH). If $\omega(\boldsymbol{k}) = \omega(k)$ is isotropic then it is clear that $\boldsymbol{u}$ will be non-zero only if the spectrum is anisotropic. In particular one requires that $F_{lm}^\mu(\boldsymbol{k}) \neq F_{lm}^\mu(-\boldsymbol{k})$ for some range of $\boldsymbol{k}$: net drift results only if the wavenumber spectrum or the dispersion relation embodies a definite spatial direction.

Higher-order contributions to the drift, which we will not consider any further here, arise from higher-order terms in (5.8) and (5.10). By symmetry, these all must vanish as well for an isotropic spectrum.

5.4. *Perturbation expansion for the diffusion tensor*

From the second line of (4.24), the relative order $(u_0/c_0)^2$ corrections to (5.12) contain contributions from both $\rho^{(2)}$ and $\rho^{(3)}$. Since they are simpler, and turn out to vanish identically for waves, we consider the $\rho^{(3)}$ contributions first. From (4.25), (4.26) and (5.10) one obtains to leading non-trivial order

$$
\begin{aligned}
\rho^{(3)}_{lmn}(\boldsymbol{x}) = &\int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^t \mathrm{d}s_3 \langle [\boldsymbol{v}(\boldsymbol{x},s_3)\cdot\nabla] v_l(\boldsymbol{x},t) v_m(\boldsymbol{x},s_1) v_n(\boldsymbol{x},s_2)\rangle_c \\
&+ \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^{s_1} \mathrm{d}s_3 \langle v_l(\boldsymbol{x},t)[\boldsymbol{v}(\boldsymbol{x},s_3)\cdot\nabla] v_m(\boldsymbol{x},s_1) v_n(\boldsymbol{x},s_2)\rangle_c \\
&+ \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^{s_2} \mathrm{d}s_3 \langle v_l(\boldsymbol{x},t) v_m(\boldsymbol{x},s_1)[\boldsymbol{v}(\boldsymbol{x},s_3)\cdot\nabla] v_n(\boldsymbol{x},s_2)\rangle_c \\
&+ (l \leftrightarrow m) + (l \leftrightarrow n) \\
= &\sum_j \partial'_j \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^t \mathrm{d}s_3 [C_{lm}(\boldsymbol{x}',\boldsymbol{x},t-s_1) C_{jn}(\boldsymbol{x},\boldsymbol{x},s_3-s_2) \\
&+ C_{ln}(\boldsymbol{x}',\boldsymbol{x},t-s_2) C_{jm}(\boldsymbol{x},\boldsymbol{x},s_3-s_1)] \\
&+ \sum_j \partial'_j \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^{s_1} \mathrm{d}s_3 [C_{lm}(\boldsymbol{x},\boldsymbol{x}',t-s_1) C_{jn}(\boldsymbol{x},\boldsymbol{x},s_3-s_2) \\
&+ C_{lj}(\boldsymbol{x},\boldsymbol{x},t-s_3) C_{mn}(\boldsymbol{x}',\boldsymbol{x},s_1-s_2)] \\
&+ \sum_j \partial'_j \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^{s_2} \mathrm{d}s_3 [C_{ln}(\boldsymbol{x},\boldsymbol{x}',t-s_2) C_{mj}(\boldsymbol{x},\boldsymbol{x},s_1-s_3) \\
&+ C_{lj}(\boldsymbol{x},\boldsymbol{x},t-s_3) C_{mn}(\boldsymbol{x},\boldsymbol{x}',s_1-s_2)] \\
&+ (l \leftrightarrow m) + (l \leftrightarrow n),
\end{aligned} \tag{5.19}
$$

in which one should set $\boldsymbol{x}' = \boldsymbol{x}'' = \boldsymbol{x}$ after all derivatives have been taken. Wick's theorem has been used to decompose the four-point correlator into a sum of products of two-point correlators, and the cummulant average eliminates one term from each. The triple time integrations may be simplified using the definitions

$$
\begin{aligned}
I_{lm}(\boldsymbol{x},\boldsymbol{x}',t) &\equiv \int_t^{\infty} \mathrm{d}s C_{lm}(\boldsymbol{x},\boldsymbol{x}',s) \\
\tilde{I}_{lm}(\boldsymbol{x},\boldsymbol{x}';t) &\equiv \int_{-\infty}^{-t} \mathrm{d}s C_{lm}(\boldsymbol{x},\boldsymbol{x}',s) = I_{ml}(\boldsymbol{x}',\boldsymbol{x};t) \\
&= I_{lm}(\boldsymbol{x},\boldsymbol{x}';-\infty) - I_{lm}(\boldsymbol{x},\boldsymbol{x}';-t)
\end{aligned} \tag{5.20}
$$

which then both essentially vanish for $t > \tau$. The required manipulations for the six different types of triple integral that appear in (5.19) are outlined in Appendix A. For example, the integral over the product $C_{lm}(\boldsymbol{x}',\boldsymbol{x},t-s_1)C_{jn}(\boldsymbol{x},\boldsymbol{x},s_3-s_2)$ is of the form $J_9(t)$ in Appendix A, and so on. One may divide the contributions into two parts: those that are apparently divergent and scale linearly in t for $t > \tau$ and those that approach a finite limit. The former terms may be reduced to the form

$$
\begin{aligned}
\rho^{(3)\mathrm{div}}_{lmn}(\boldsymbol{x}) &= t \lim_{\boldsymbol{x}'\to\boldsymbol{x}} \partial'_j \sum_j [I_{lm}(\boldsymbol{x}',\boldsymbol{x}';-\infty) I_{jn}(\boldsymbol{x}',\boldsymbol{x}';-\infty) + (l \leftrightarrow n) + (m \leftrightarrow n)] \\
&= t \sum_j [\hat{S}_{jn}(\boldsymbol{x};0)\partial'_j \hat{S}_{lm}(\boldsymbol{x};0) + \hat{S}_{jl}(\boldsymbol{x};0)\partial'_j \hat{S}_{nm}(\boldsymbol{x};0) + \hat{S}_{jm}(\boldsymbol{x};0)\partial'_j \hat{S}_{ln}(\boldsymbol{x};0)],
\end{aligned} \tag{5.21}
$$

and in the general inhomogeneous case represent slow variations in the transport parameters with the length scale (varying as $\sqrt{t}$) explored by the diffusion process – recall that $S_{lm}(\boldsymbol{x};0)$ is proportional to the zeroth-order result for the diffusion constant so that (5.21) is finite only if there are spatial gradients in the zeroth-order diffusion constant. The result will clearly vanish if the system is translation invariant. For waves, despite the explicit lack of translation invariance in the 'vertical' directions, and even if 'horizontal' inhomogeneities exist, the vanishing of the frequency spectrum at zero frequency ensures that $\rho^{(3)\mathrm{div}}_{lmn}(\boldsymbol{x}) \equiv 0$. It is probable, however, that non-vanishing terms linear in t, presumably proportional to gradients in the drift velocity (5.18), will manifest at higher order in u_0/c_0.

The convergent terms may be reduced to the form

$$
\begin{aligned}
\rho^{(3)\mathrm{con}}_{lmn}(\boldsymbol{x}) = &-\lim_{\boldsymbol{x}'\to\boldsymbol{x}}\sum_j \partial'_j \int_0^\infty \mathrm{d}s\, \{I_{lm}(\boldsymbol{x}',\boldsymbol{x}';-\infty)I_{jn}(\boldsymbol{x},\boldsymbol{x};s) \\
&+I_{nj}(\boldsymbol{x},\boldsymbol{x};-\infty)[I_{lm}(\boldsymbol{x}',\boldsymbol{x};s)+I_{lm}(\boldsymbol{x},\boldsymbol{x}';s)] + (l\leftrightarrow n) + (m\leftrightarrow n)\} \\
= &-\sum_j \partial_j \hat{S}_{lm}(\boldsymbol{x};0)\left\{\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi\omega^2}[\hat{S}_{jn}(\boldsymbol{x};0)-\hat{S}^{+}_{jn}(\boldsymbol{x},\omega)] - \frac{\mathrm{i}}{2}\hat{S}'_{jn}(\boldsymbol{x};0)\right\} \\
&-\sum_j \hat{S}_{nj}(\boldsymbol{x};0)\left\{\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi\omega^2}[\partial_j\hat{S}_{ml}(\boldsymbol{x};0)-\partial_j\hat{S}^{+}_{ml}(\boldsymbol{x};\omega)]\right. \\
&\left.-\frac{\mathrm{i}}{2}\partial_j\hat{S}'_{ml}(\boldsymbol{x};0) - \hat{R}^{j\prime}_{ml}(\boldsymbol{x};0)\right\} \\
&+(l\leftrightarrow n)+(m\leftrightarrow n),
\end{aligned}
\tag{5.22}
$$

in which $\hat{S}^{\pm}_{lm}(\boldsymbol{x},\omega) = \frac{1}{2}[\hat{S}_{lm}(\boldsymbol{x},\omega) \pm \hat{S}_{lm}(\boldsymbol{x},-\omega)]$, we have used $\lim_{\boldsymbol{x}'\to\boldsymbol{x}}\partial'_j I_{ml}(\boldsymbol{x},\boldsymbol{x}';s) = \lim_{\boldsymbol{x}'\to\boldsymbol{x}}[\partial'_j I_{ml}(\boldsymbol{x}',\boldsymbol{x}';s) - \partial'_j I_{ml}(\boldsymbol{x}',\boldsymbol{x};s)]$, and where primes denote differentiation with respect to ω. This again vanishes identically for waves, and

$$
\rho^{(3)}_{lmn}(\boldsymbol{x}) \equiv O(u_0^5\tau^2/c_0^2),
\tag{5.23}
$$

in that case. We conclude then that for waves only $\rho^{(2)}$ contributes to the diffusion tensor at lowest non-trivial order.

We turn then to the study of $\rho^{(2)}$ at $O(u_0^4\tau/c_0^2)$. The expansion (5.10) yields

$$
\begin{aligned}
\rho^{(2)}_{lm}(\boldsymbol{x}) = &\int_0^t \mathrm{d}s_1 \int_0^{s_1}\mathrm{d}s_2\int_0^{s_2}\mathrm{d}s_3\langle v_l(\boldsymbol{x},t)[\boldsymbol{v}(\boldsymbol{x},s_3)\cdot\nabla][\boldsymbol{v}(\boldsymbol{x},s_2)\cdot\nabla]v_m(\boldsymbol{x},s_1)\rangle \\
&+\int_0^t \mathrm{d}s_1 \int_0^{t}\mathrm{d}s_2\int_0^{s_2}\mathrm{d}s_3\langle[\boldsymbol{v}(\boldsymbol{x},s_3)\cdot\nabla][\boldsymbol{v}(\boldsymbol{x},s_2)\cdot\nabla]v_l(\boldsymbol{x},t)v_m(\boldsymbol{x},s_1)\rangle \\
&+\int_0^t \mathrm{d}s_1 \int_0^{t}\mathrm{d}s_2\int_0^{s_1}\mathrm{d}s_3\langle[\boldsymbol{v}(\boldsymbol{x},s_2)\cdot\nabla]v_l(\boldsymbol{x},t)[\boldsymbol{v}(\boldsymbol{x},s_3)\cdot\nabla]v_m(\boldsymbol{x},s_1)\rangle_c + (l\leftrightarrow m) \\
= &\sum_{i,j}(\partial'_i+\partial''_i)\partial''_j\int_0^t\mathrm{d}s_1\int_0^{s_1}\mathrm{d}s_2\int_0^{s_2}\mathrm{d}s_3[C_{li}(\boldsymbol{x},\boldsymbol{x},t-s_3)C_{mj}(\boldsymbol{x}'',\boldsymbol{x}',s_1-s_2) \\
&+C_{lj}(\boldsymbol{x},\boldsymbol{x}',t-s_2)C_{mi}(\boldsymbol{x}'',\boldsymbol{x},s_1-s_3)+C_{lm}(\boldsymbol{x},\boldsymbol{x}'',t-s_1)C_{ji}(\boldsymbol{x}',\boldsymbol{x},s_2-s_3)] \\
&+\sum_{i,j}(\partial'_i+\partial''_i)\partial''_j\int_0^t\mathrm{d}s_1\int_0^{t}\mathrm{d}s_2\int_0^{s_2}\mathrm{d}s_3[C_{li}(\boldsymbol{x}'',\boldsymbol{x},t-s_3)C_{mj}(\boldsymbol{x},\boldsymbol{x}',s_1-s_2) \\
&+C_{lj}(\boldsymbol{x}'',\boldsymbol{x}',t-s_2)C_{mi}(\boldsymbol{x},\boldsymbol{x},s_1-s_3)+C_{lm}(\boldsymbol{x}'',\boldsymbol{x},t-s_1)C_{ji}(\boldsymbol{x}',\boldsymbol{x},s_2-s_3)]
\end{aligned}
$$

$$+\sum_{i,j}\partial_i'\partial_j''\int_0^t \mathrm{d}s_1\int_0^t \mathrm{d}s_2\int_0^{s_1}\mathrm{d}s_3[C_{lj}(\boldsymbol{x}',\boldsymbol{x},t-s_3)C_{mi}(\boldsymbol{x}'',\boldsymbol{x},s_1-s_2)$$
$$+C_{lm}(\boldsymbol{x}',\boldsymbol{x}'',t-s_1)C_{ij}(\boldsymbol{x},\boldsymbol{x},s_2-s_3)]+(l\leftrightarrow m), \quad (5.24)$$

in which, as usual, the limit $\boldsymbol{x}',\boldsymbol{x}''\to\boldsymbol{x}$ should be taken at the end. The reduction of the time integrations is derived again in Appendix A. The integrals are of the form, respectively, J_7, J_1, J_2, J_3, J_8, J_4, J_6, and J_5. Once again there appear for large t apparently divergent terms linear in t and terms independent of t. The former may be reduced to the form

$$\rho_{lm}^{(2)\mathrm{div}}(\boldsymbol{x})=t\sum_{i,j}\big\{\partial_j''I_{lm}(\boldsymbol{x}'',\boldsymbol{x}'';-\infty)\partial_i'I_{ji}(\boldsymbol{x}',\boldsymbol{x};0)$$
$$+\tfrac{1}{2}\partial_i''\partial_j''I_{lm}(\boldsymbol{x}'',\boldsymbol{x}'';-\infty)I_{ji}(\boldsymbol{x},\boldsymbol{x};-\infty)$$
$$+(\partial_i'+\partial_i'')\partial_j''[I_{lj}(\boldsymbol{x}'',\boldsymbol{x}';0)I_{mi}(\boldsymbol{x},\boldsymbol{x};-\infty)+(l\leftrightarrow m)]\big\}$$
$$=t\sum_{i,j}\Bigg(\partial_j\hat{S}_{lm}(\boldsymbol{x};0)\left[\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi\omega}\hat{R}_{ji}^{i-}(\boldsymbol{x};\omega)+\frac{\mathrm{i}}{2}\hat{R}_{ji}^{i}(\boldsymbol{x};0)\right]$$
$$+\left\{\hat{S}_{mi}(\boldsymbol{x};0)\left[\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi\omega}\partial_i\hat{R}_{lj}^{j-}(\boldsymbol{x};\omega)+\frac{\mathrm{i}}{2}\partial_i\hat{R}_{lj}^{i}(\boldsymbol{x};0)\right]+(l\leftrightarrow m)\right\}$$
$$+\tfrac{1}{2}\partial_i\partial_j\hat{S}_{lm}(\boldsymbol{x};0)\hat{S}_{ji}(\boldsymbol{x};0)\Bigg). \quad (5.25)$$

This vanishes identically for waves.

The convergent terms may be simplified to

$$\rho_{lm}^{(2)\mathrm{con}}(\boldsymbol{x})=\sum_{i,j}(\partial_i'+\partial_i'')\partial_j''\int_0^{\infty}\mathrm{d}s[I_{mi}(\boldsymbol{x}'',\boldsymbol{x};s)I_{lj}(\boldsymbol{x},\boldsymbol{x}';-\infty)$$
$$-I_{im}(\boldsymbol{x},\boldsymbol{x};-\infty)I_{lj}(\boldsymbol{x}'',\boldsymbol{x}';s)-I_{im}(\boldsymbol{x},\boldsymbol{x};s)I_{lj}(\boldsymbol{x}'',\boldsymbol{x}';0)$$
$$-I_{lm}(\boldsymbol{x},\boldsymbol{x}'';-\infty)I_{ji}(\boldsymbol{x}',\boldsymbol{x};s)-I_{lm}(\boldsymbol{x},\boldsymbol{x}'';s)I_{ji}(\boldsymbol{x}',\boldsymbol{x};0)]$$
$$+\sum_{i,j}\partial_i'\partial_j''\int_0^{\infty}\mathrm{d}s\{[I_{mi}(\boldsymbol{x}'',\boldsymbol{x};-\infty)-I_{im}(\boldsymbol{x},\boldsymbol{x}'';s)]I_{lj}(\boldsymbol{x}',\boldsymbol{x};s)$$
$$-[I_{ij}(\boldsymbol{x},\boldsymbol{x};-\infty)-I_{ij}(\boldsymbol{x},\boldsymbol{x};s)]I_{lm}(\boldsymbol{x}',\boldsymbol{x}'';s)$$
$$-[I_{ij}(\boldsymbol{x},\boldsymbol{x};s)+I_{ji}(\boldsymbol{x},\boldsymbol{x};s)]I_{lm}(\boldsymbol{x}',\boldsymbol{x}'';0)\}+(l\leftrightarrow m). \quad (5.26)$$

Using the definitions

$$\left.\begin{aligned}
\hat{T}_{lm}^{ij}(\boldsymbol{x};\omega)&=-\lim_{\boldsymbol{x}'\to\boldsymbol{x}}\int_{-\infty}^{\infty}\mathrm{d}s\,\mathrm{e}^{\mathrm{i}\omega s}\partial_i'\partial_j'C_{lm}(\boldsymbol{x}',\boldsymbol{x};s),\\
\bar{T}_{lm}^{ij}(\boldsymbol{x};\omega)&=\lim_{\boldsymbol{x}'\to\boldsymbol{x}}\int_{-\infty}^{\infty}\mathrm{d}s\,\mathrm{e}^{\mathrm{i}\omega s}\partial_i'\partial_jC_{lm}(\boldsymbol{x}',\boldsymbol{x};s)\\
&=\mathrm{i}\partial_j\hat{R}_{lm}^{i}(\boldsymbol{x};\omega)+\hat{T}_{lm}^{ij}(\boldsymbol{x};\omega)=\bar{T}_{ml}^{ji}(\boldsymbol{x},-\omega),\\
\hat{T}_{lm}^{ij\pm}(\boldsymbol{x};\omega)&=\tfrac{1}{2}[\hat{T}_{lm}^{ij}(\boldsymbol{x};\omega)\pm\hat{T}_{lm}^{ij}(\boldsymbol{x};-\omega)],\\
\bar{T}_{lm}^{ij\pm}(\boldsymbol{x};\omega)&=\tfrac{1}{2}[\bar{T}_{lm}^{ij}(\boldsymbol{x};\omega)\pm\bar{T}_{lm}^{ij}(\boldsymbol{x};-\omega)],
\end{aligned}\right\} \quad (5.27)$$

and by taking advantage of the symmetry of the operator $\partial_i''\partial_j''$ under the interchange $i \leftrightarrow j$ and of the operator $\partial_i'\partial_j''$ under the simultaneous interchange $i \leftrightarrow j$, $\boldsymbol{x}' \leftrightarrow \boldsymbol{x}''$, $\rho^{(2)\mathrm{con}}$ may be simplified to the form

$$\rho_{lm}^{(2)\mathrm{con}}(\boldsymbol{x}) \equiv \rho_{lm,1}^{(2)\mathrm{con}}(\boldsymbol{x}) + \rho_{lm,2}^{(2)\mathrm{con}}(\boldsymbol{x}) + \rho_{lm,3}^{(2)\mathrm{con}}(\boldsymbol{x}) + (l \leftrightarrow m), \tag{5.28a}$$

$$\begin{aligned}\rho_{lm,1}^{(2)\mathrm{con}}(\boldsymbol{x}) = \frac{1}{2}\sum_{i,j}\Big\{&-\hat{S}_{im}(\boldsymbol{x};0)\partial_i\hat{R}_{lj}^{j\prime}(\boldsymbol{x};0) + \mathrm{i}\hat{S}_{lj}(\boldsymbol{x};0)\hat{T}_{mi}^{ij\prime}(\boldsymbol{x};0)\\
&-\frac{\mathrm{i}}{2}\hat{S}_{ij}(\boldsymbol{x};0)\hat{T}_{ml}^{ij\prime}(\boldsymbol{x};0) + \mathrm{i}\hat{R}_{mi}^{j\prime}(\boldsymbol{x};0)[\hat{R}_{jl}^{i}(\boldsymbol{x};0) + \hat{R}_{lj}^{i}(\boldsymbol{x};0)]\\
&-\tfrac{1}{2}\partial_j\hat{S}_{ml}(\boldsymbol{x};0)\hat{R}_{ji}^{i\prime}(\boldsymbol{x};0)\Big\}\\
&+\frac{1}{2}\sum_j[\mathrm{i}\hat{S}_{jm}'(\boldsymbol{x};0)\partial_j\rho_l^{(1)}(\boldsymbol{x}) + \hat{R}_{ml}^{j\prime}(\boldsymbol{x};0)\rho_j^{(1)}(\boldsymbol{x})],\end{aligned} \tag{5.28b}$$

$$\begin{aligned}\rho_{lm,2}^{(2)\mathrm{con}}(\boldsymbol{x}) = -\sum_{i,j}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi\omega^2}\Big\{&\mathrm{i}\hat{S}_{im}(\boldsymbol{x};0)\partial_i[\hat{R}_{lj}^{j}(\boldsymbol{x};0) - \hat{R}_{lj}^{j+}(\boldsymbol{x};\omega)]\\
&+\hat{S}_{lj}(\boldsymbol{x};0)[\hat{T}_{mi}^{ij}(\boldsymbol{x};0) - \hat{T}_{mi}^{ij+}(\boldsymbol{x};\omega)]\\
&+[\hat{R}_{lj}^{i}(\boldsymbol{x};0) + \hat{R}_{jl}^{i}(\boldsymbol{x};0)][\hat{R}_{mi}^{j}(\boldsymbol{x};0) - \hat{R}_{mi}^{j+}(\boldsymbol{x};\omega)]\\
&+\mathrm{i}\partial_j\hat{R}_{ml}(\boldsymbol{x};0)[\hat{S}_{ij}(\boldsymbol{x};0) - \hat{S}_{ij}^{+}(\boldsymbol{x};\omega)]\\
&+\frac{\mathrm{i}}{2}\partial_j\hat{S}_{ml}(\boldsymbol{x};0)[\hat{R}_{ji}^{i}(\boldsymbol{x};0) - \hat{R}_{ji}^{i+}(\boldsymbol{x};\omega)]\\
&+\tfrac{1}{2}\hat{S}_{ij}(\boldsymbol{x};0)[\hat{T}_{ml}^{ij}(\boldsymbol{x};0) - \hat{T}_{ml}^{ij+}(\boldsymbol{x};\omega)]\\
&+\hat{S}_{ij}(\boldsymbol{x};0)[\bar{T}_{ml}^{ij}(\boldsymbol{x};0) - \bar{T}_{ml}^{ij+}(\boldsymbol{x};\omega)]\Big\},\end{aligned} \tag{5.28c}$$

$$\begin{aligned}\rho_{lm,3}^{(2)\mathrm{con}}(\boldsymbol{x}) = \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{4\pi\omega^2}\Big(&-\sum_i\{2\partial_i\rho_l^{(1)}(\boldsymbol{x})[\hat{S}_{im}(\boldsymbol{x};0) - \hat{S}_{im}^{+}(\boldsymbol{x};\omega)]\\
&+\rho_i^{(1)}(\boldsymbol{x})\partial_i[\hat{S}_{ml}(\boldsymbol{x};0) - \hat{S}_{ml}^{+}(\boldsymbol{x};\omega)]\}\\
&+\sum_{i,j}\{\hat{R}_{lj}^{i-}(\boldsymbol{x};\omega)\hat{R}_{mi}^{j-}(\boldsymbol{x};\omega)\\
&+[\hat{R}_{lj}^{i}(\boldsymbol{x};0) - \hat{R}_{lj}^{i+}(\boldsymbol{x};\omega)][\hat{R}_{mi}^{j+}(\boldsymbol{x};0) - \hat{R}_{mi}^{j+}(\boldsymbol{x};\omega)]\\
&-\hat{S}_{ij}^{-}(\boldsymbol{x},\omega)\bar{T}_{lm}^{ij-}(\boldsymbol{x};\omega) + [\hat{S}_{ij}(\boldsymbol{x};0)\\
&-\hat{S}_{ij}^{+}(\boldsymbol{x};\omega)][\bar{T}_{lm}^{ij}(\boldsymbol{x};0) - \bar{T}_{lm}^{ij+}(\boldsymbol{x};\omega)]\}\Big),\end{aligned} \tag{5.28d}$$

where the primes again stand for frequency derivatives, and, from (5.14), we use here the lowest-order result $\rho_l^{(1)}(\boldsymbol{x}) = \lim_{\boldsymbol{x}'\to\boldsymbol{x}}\sum_m \partial_m' I_{lm}(\boldsymbol{x}',\boldsymbol{x};0)$, i.e. the first two terms on the right-hand side of (5.17).

The symmetrization of the indices l, m is aided with the identitities

$$\left.\begin{aligned}
&\hat{S}_{ml}(\boldsymbol{x};\omega) = \hat{S}_{lm}(\boldsymbol{x};-\omega) \Rightarrow \hat{S}^{\pm}_{ml}(\boldsymbol{x};\omega) = \pm\hat{S}^{\pm}_{lm}(\boldsymbol{x};\omega),\\
&\hat{R}^{i}_{ml}(\boldsymbol{x};\omega) = -\hat{R}^{i}_{lm}(\boldsymbol{x};-\omega) - \mathrm{i}\partial_i\hat{S}_{ml}(\boldsymbol{x};\omega)\\
&\qquad \Rightarrow \hat{R}^{i\pm}_{ml}(\boldsymbol{x};\omega) = \mp[\hat{R}^{i\pm}_{lm}(\boldsymbol{x};\omega) + \mathrm{i}\partial_i\hat{S}^{\pm}_{lm}(\boldsymbol{x};\omega)],\\
&\bar{T}^{ij}_{ml}(\boldsymbol{x};\omega) = \bar{T}^{ji}_{lm}(\boldsymbol{x};\omega) = \bar{T}^{ij}_{lm}(\boldsymbol{x};-\omega) + \mathrm{i}[\partial_j\hat{R}^{i}_{ml}(\boldsymbol{x};\omega) - \partial_i\hat{R}^{j}_{ml}(\boldsymbol{x};\omega)]\\
&\Rightarrow \bar{T}^{ij\pm}_{ml}(\boldsymbol{x};\omega) = \pm\{\bar{T}^{ij\pm}_{lm}(\boldsymbol{x};\omega) - \mathrm{i}[\partial_j\hat{R}^{i\pm}_{lm}(\boldsymbol{x};\omega) - \partial_i\hat{R}^{j\pm}_{lm}(\boldsymbol{x};\omega)]\},\\
&\hat{T}^{ij}_{ml}(\boldsymbol{x};\omega) = \hat{T}^{ij}_{lm}(\boldsymbol{x};-\omega) - \mathrm{i}[\partial_j\hat{R}^{i}_{ml}(\boldsymbol{x};\omega) + \partial_i\hat{R}^{j}_{ml}(\boldsymbol{x};\omega)] + \partial_i\partial_j\hat{S}_{ml}(\boldsymbol{x};\omega)\\
&\Rightarrow \hat{T}^{ij\pm}_{ml}(\boldsymbol{x};\omega) = \pm\{\hat{T}^{ij\pm}_{lm}(\boldsymbol{x};-\omega) + \mathrm{i}[\partial_j\hat{R}^{i\pm}_{lm}(\boldsymbol{x};\omega)\\
&\qquad +\partial_i\hat{R}^{j\pm}_{lm}(\boldsymbol{x};\omega)] - \partial_i\partial_j\hat{S}^{\pm}_{lm}(\boldsymbol{x};\omega)\},
\end{aligned}\right\} \tag{5.29}$$

from which one obtains

$$\left.\begin{aligned}
&\hat{S}^{+}_{lm}(\boldsymbol{x},\omega) + \hat{S}^{+}_{ml}(\boldsymbol{x},\omega) = 2S^{+}_{lm}(\boldsymbol{x},\omega),\\
&\hat{S}^{-}_{lm}(\boldsymbol{x},\omega) + \hat{S}^{-}_{ml}(\boldsymbol{x},\omega) = 0,\\
&\hat{R}^{i+}_{lm}(\boldsymbol{x},\omega) + \hat{R}^{i+}_{ml}(\boldsymbol{x},\omega) = -\mathrm{i}\partial_i\hat{S}^{+}_{lm}(\boldsymbol{x};\omega),\\
&\hat{R}^{i-}_{lm}(\boldsymbol{x},\omega) + \hat{R}^{i-}_{ml}(\boldsymbol{x},\omega) = 2\hat{R}^{i-}_{lm}(\boldsymbol{x},\omega) - \mathrm{i}\partial_i\hat{S}^{-}_{lm}(\boldsymbol{x};\omega),\\
&\bar{T}^{ij+}_{lm}(\boldsymbol{x};\omega) + \bar{T}^{ij+}_{ml}(\boldsymbol{x};\omega) = 2\bar{T}^{ij+}_{lm}(\boldsymbol{x};\omega) - \mathrm{i}[\partial_j\hat{R}^{i+}_{lm}(\boldsymbol{x},\omega) - \partial_i\hat{R}^{j+}_{lm}(\boldsymbol{x},\omega)],\\
&\bar{T}^{ij-}_{lm}(\boldsymbol{x};\omega) + \bar{T}^{ij-}_{ml}(\boldsymbol{x};\omega) = \mathrm{i}[\partial_j\hat{R}^{i-}_{lm}(\boldsymbol{x},\omega) - \partial_i\hat{R}^{j-}_{lm}(\boldsymbol{x},\omega)],\\
&\hat{T}^{ij+}_{lm}(\boldsymbol{x};\omega) + \hat{T}^{ij+}_{ml}(\boldsymbol{x};\omega) = 2\hat{T}^{ij+}_{lm}(\boldsymbol{x};\omega) + \mathrm{i}[\partial_j\hat{R}^{i+}_{lm}(\boldsymbol{x},\omega) + \partial_i\hat{R}^{j+}_{lm}(\boldsymbol{x},\omega)]\\
&\qquad -\partial_i\partial_j\hat{S}^{+}_{lm}(\boldsymbol{x},\omega),\\
&\hat{T}^{ij-}_{lm}(\boldsymbol{x};\omega) + \hat{T}^{ij-}_{ml}(\boldsymbol{x};\omega) = -\mathrm{i}[\partial_j\hat{R}^{i-}_{lm}(\boldsymbol{x},\omega) + \partial_i\hat{R}^{j-}_{lm}(\boldsymbol{x},\omega)] + \partial_i\partial_j\hat{S}^{-}_{lm}(\boldsymbol{x},\omega).
\end{aligned}\right\} \tag{5.30}$$

We shall make explicit use of these only in the case of waves, for which great simplifications occur: only the final expression, $\rho^{(2)\mathrm{con}}_{lm,3}(\boldsymbol{x})$ remains non-zero, and is then in fact the *only* surviving contribution to the diffusion tensor at leading non-trivial order. From (5.29) and (5.30) one obtains in that case the $O(u_0^4\tau/c_0^2)$ contribution

$$\left.\begin{aligned}
D^{(2)}_{lm}(\boldsymbol{x}) &= \tfrac{1}{2}\rho^{(2)}_{lm}(\boldsymbol{x}) = \tfrac{1}{2}\left[\rho^{(2)\mathrm{con}}_{lm,3}(\boldsymbol{x}) + \rho^{(2)\mathrm{con}}_{ml,3}(\boldsymbol{x})\right]\\
&= \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{4\pi\omega^2}\Bigg(\sum_i\Big\{\partial_i\rho^{(1)}_l(\boldsymbol{x})\hat{S}^{+}_{im}(\boldsymbol{x};\omega)\\
&\quad +\partial_i\rho^{(1)}_m(\boldsymbol{x})\hat{S}^{+}_{il}(\boldsymbol{x};\omega) + \rho^{(1)}_i(\boldsymbol{x})\partial_i\hat{S}^{+}_{ml}(\boldsymbol{x};\omega)\Big\}\\
&\quad +\sum_{i,j}\Big\{\hat{R}^{i+}_{lj}(\boldsymbol{x};\omega)\hat{R}^{j+}_{mi}(\boldsymbol{x};\omega) + \hat{R}^{i-}_{lj}(\boldsymbol{x};\omega)\hat{R}^{j-}_{mi}(\boldsymbol{x};\omega)\\
&\quad +\hat{S}^{+}_{ij}(\boldsymbol{x};\omega)\bar{T}^{ij+}_{lm}(\boldsymbol{x};\omega)\\
&\quad -\frac{\mathrm{i}}{2}\hat{S}^{-}_{ij}(\boldsymbol{x};\omega)\left[\partial_j\hat{R}^{i-}_{lm}(\boldsymbol{x};\omega) - \partial_i\hat{R}^{j-}_{lm}(\boldsymbol{x};\omega)\right]\Big\}\Bigg),\\
\rho^{(1)}_l(\boldsymbol{x}) &= u_l(\boldsymbol{x}) = \sum_m\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi\omega}\hat{R}^{m-}_{lm}(\boldsymbol{x};\omega),
\end{aligned}\right\} \tag{5.31}$$

in which the last term arises from the application of the sixth line of (5.30) to the combination $\hat{S}^{-}\bar{T}^{-}$. Equation (5.31) represents the fundamental result of this section which will form the basis for all subsequent explicit calculations for various special cases.

5.5. *Fully translation-invariant case*

In later sections realistic wave systems will be treated. Here let us consider the simplest case in which the system is fully translation-invariant in all directions (i.e. $\bar{d} = 0$ and $\hat{d} = d$ in the notation introduced below equation (2.5)). Only the difference variable $\boldsymbol{x}' - \boldsymbol{x}$ enters the definitions of $\hat{S}$, $\hat{R}$, $\hat{T}$ and $\bar{T}$. In this case all explicit $\boldsymbol{x}$-dependence disappears, and all spatial derivatives acting on such dependence vanish identically. In particular, for example, all extra terms on the right-hand sides of the identities (5.30) disappear. The spectral form (2.7) now contains no z, z'-dependence, and from (5.18) (generalized trivially from two to d horizontal dimensions), the drift velocity takes the form

$$u_l = \sum_{\mu} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi\omega} \sum_{m} \hat{R}^{m}_{lm}(\omega) = \sum_{\mu} \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{1}{\omega_\mu(\boldsymbol{k})} \sum_{m} k_m \left[F^{\mu}_{lm}(\boldsymbol{k}) + F^{\mu}_{ml}(\boldsymbol{k})\right]. \quad (5.32)$$

The diffusion tensor simplifies to the form

$$D^{(2)}_{lm} = \sum_{i,j} \int \frac{\mathrm{d}\omega}{4\pi\omega^2} [\hat{T}^{ij+}_{lm}(\omega)\hat{S}^{+}_{ij}(\omega) + \hat{R}^{i+}_{lj}(\omega)\hat{R}^{j+}_{mi}(\omega) + \hat{R}^{i-}_{lj}(\omega)\hat{R}^{j-}_{mi}(\omega)]. \quad (5.33)$$

Substituting (2.7), one obtains the more explicit result

$$\begin{aligned} D^{(2)}_{lm} = \sum_{\mu,\nu} \sum_{i,j} \int \frac{\mathrm{d}^d k}{(2\pi)^d} \int \frac{\mathrm{d}^d k'}{(2\pi)^d} \frac{k_i}{2\omega_\mu(\boldsymbol{k})^2} \\ \times(\{k_j [F^{\mu}_{lm}(\boldsymbol{k}) + F^{\mu}_{ml}(\boldsymbol{k})] F^{\nu}_{ij}(\boldsymbol{k}') + k'_j [F^{\mu}_{mj}(\boldsymbol{k}) F^{\nu}_{li}(\boldsymbol{k}') + F^{\mu}_{jm}(\boldsymbol{k}) F^{\nu}_{il}(\boldsymbol{k}')]\} \\ \times 2\pi\delta[\omega_\mu(\boldsymbol{k}) - \omega_\nu(\boldsymbol{k}')] \\ + \{k_j [F^{\mu}_{lm}(\boldsymbol{k}) + F^{\mu}_{ml}(\boldsymbol{k})] F^{\nu}_{ij}(\boldsymbol{k}') - k'_j [F^{\mu}_{mj}(\boldsymbol{k}) F^{\nu}_{il}(\boldsymbol{k}') + F^{\mu}_{jm}(\boldsymbol{k}) F^{\nu}_{li}(\boldsymbol{k}')]\} \\ \times 2\pi\delta[\omega_\mu(\boldsymbol{k}) + \omega_\nu(\boldsymbol{k}')]). \quad (5.34) \end{aligned}$$

The symmetry of (5.33) under interchange of l and m follows from the fact that the $\hat{R}^{i\pm}_{lm}(\omega)$ are, respectively, symmetric and antisymmetric in l and m. To establish this symmetry in (5.34) the combined symmetries of the integrand under interchange of i and j and of $\boldsymbol{k}$ and $\boldsymbol{k}'$ must be used.

5.6. *Isotropic spectrum*

The general result (5.34) is not very illuminating, so we specialize here to the yet simpler case of an isotropic dispersion relation $\omega(\boldsymbol{k}) = \omega(k)$, and an isotropic spectrum,

$$\boldsymbol{F}(\boldsymbol{k}) = F_L(k)\hat{\boldsymbol{k}}\hat{\boldsymbol{k}} + F_T(k)[\boldsymbol{I} - \hat{\boldsymbol{k}}\hat{\boldsymbol{k}}], \quad (5.35)$$

in which $\boldsymbol{I}$ is the identity matrix and L and T denote longitudinal and transverse parts, respectively. The diffusion tensor, $\boldsymbol{D} = D\boldsymbol{I}$ will then be diagonal. Clearly $\hat{R}^{k}_{ij}(\omega) \equiv 0$, and the Stokes drift velocity (5.17) vanishes to the order calculated so far. Purely on symmetry grounds, this result clearly must generalize to all orders: $\boldsymbol{u} \equiv 0$ under isotropic conditions.

One may substitute (5.35) directly into (5.34), but a simpler route is to work from

(5.33). First, one obtains then $\hat{\mathbf{S}}(\omega) = (1/d)\hat{S}(\omega)\boldsymbol{I}$ (so that $\hat{S}(\omega) = \mathrm{tr}\hat{\mathbf{S}}(\omega)$) with

$$\begin{aligned}\hat{S}(\omega) &= \int_{-\infty}^{\infty} \mathrm{d}t\langle \boldsymbol{v}(\boldsymbol{x},t)\cdot\boldsymbol{v}(\boldsymbol{x},0)\rangle \mathrm{e}^{\mathrm{i}\omega t} \\ &= \int \frac{\mathrm{d}^d k}{(2\pi)^d}[F_L(k) + (d-1)F_T(k)]\{\delta[\omega-\omega(k)] + \delta[\omega+\omega(k)]\}, \end{aligned} \tag{5.36}$$

in which the results of Appendix B have been used to compute the angular average of $k_i k_j$ over the unit sphere. Second, one has

$$\begin{aligned}\hat{T}^{ij}_{lm}(\omega) &= \int \frac{\mathrm{d}^d k}{(2\pi)^d} k_i k_j F_{lm}(k)\{\delta[\omega-\omega(k)] + \delta[\omega+\omega(k)]\} \\ &= k(\omega)^2 \int \frac{\mathrm{d}^d k}{(2\pi)^d}\left\{\frac{1}{d(d+2)}[F_L(k) - F_T(k)][\delta_{lm}\delta_{ij} + \delta_{lj}\delta_{mi} + \delta_{li}\delta_{mj}]\right. \\ &\quad \left. + \frac{1}{d}F_T(k)\delta_{lm}\delta_{ij}\right\}\{\delta[\omega-\omega(k)] + \delta[\omega+\omega(k)]\}, \end{aligned} \tag{5.37}$$

in which $k(\omega)$ is the inverse of the function $\omega(k)$ (for inertia gravity waves – see below – one has simply $k(\omega)^2 = (\omega^2 - f^2)/c^2$), and the results of Appendix B have again been used to compute the angular average of $k_i k_j k_l k_m$ over the unit sphere. One obtains then

$$\sum_{i,j} \hat{S}_{ij}(\omega)\hat{T}^{ij}_{lm}(\omega) = \frac{1}{d^2}k(\omega)^2\hat{S}(\omega)^2. \tag{5.38}$$

Since all quantities are already even in ω, this may be substituted directly into (5.33) to obtain the remarkably simple result $D^{(2)}_{lm} = D^{(2)}\delta_{lm}$ with

$$D^{(2)} = \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{4\pi d^2}\frac{k(\omega)^2}{\omega^2}\hat{S}(\omega)^2. \tag{5.39}$$

To exhibit the important dependence on physical, dimensional parameters, it is useful to write this expression in a scaled form. Let f_0 be a characteristic frequency of the system and let $x = \omega/f_0$ be the non-dimensional frequency. Since $u_0^2 = \int(\mathrm{d}\omega/2\pi)\hat{S}(\omega)$, one may write

$$\hat{S}(\omega) = \frac{u_0^2}{f_0}s(x); \quad \int_{-\infty}^{\infty}\frac{\mathrm{d}x}{2\pi}s(x) = 1. \tag{5.40}$$

Similarly, one writes $k(\omega)^2/\omega^2 = K(x)^2/c^2$, where c is a characteristic phase velocity and $K(x)$ is dimensionless. For inertia–gravity waves one has $K(x)^2 = 1 - x^{-2}$ (with the choice $f_0 = f$). The diffusion constant then may be expressed as

$$D^{(2)} = \frac{u_0^4}{f_0 c_0^2}B_d, \tag{5.41}$$

where all of the detailed spectral properties are characterized by the dimensionless quantity

$$B_d = \int_{-\infty}^{\infty}\frac{\mathrm{d}x}{4\pi d^2}K(x)^2 s(x)^2. \tag{5.42}$$

Equation (5.41) exhibits explicitly the fact that the estimate $D \propto u_0^2/f_0$ that naively follows from the zeroth-order result (5.12) is reduced by the factor $(u_0/c_0)^2$. Since $s(x)$ has unit integral, roughly speaking B_d will be small compared to unity if $s(x)$ is very broad, and large compared to unity if it is strongly peaked. In the following sections, numerical estimates, based on oceanographic data, will be presented for baroclinic inertia–gravity waves and surface gravity waves.

6. Applications to inertia–gravity waves

Large-scale oceanic and atmospheric motions (as occur in a thin layer of a rotating fluid) satisfy the hydrostatic approximation for the pressure field which then leads to a simplified set of equations known as the shallow-water equations (LeBlond & Mysak 1978; Gill 1982). These equations contain, in addition to the usual horizontal vortical flows, in the linear approximation, a set of oscillating solutions known as inertia–gravity (IG) or Poincaré waves. While the time scale of the former is measured in weeks (being limited in principle only by the size of the ocean), the period of IG waves actually has a lower bound determined by the latitude. Of main interest in oceanography are IG waves occurring at the interface, known as the *thermocline depth*, between two horizontal layers with slightly different temperatures, and hence slightly different densities. These long internal waves, known as baroclinic inertia–gravity (BIG) waves, account for most of the energy in oceanic motions with time scales less than one day. The corresponding amplitude of thermocline depth oscillations may attain 10 m while the horizontal fluid velocity scale is about $10\,\mathrm{cm\,s^{-1}}$. Being weakly to moderately nonlinear, BIG waves are characterized by a broad frequency spectrum resulting from the Kolmogorov type cascades of wave energy and wave action (Falkovich 1992; Falkovich & Medvedev 1992; Glazman 1996*a*, *b*). Among known causes of BIG waves in the ocean are the instability of shear flows with respect to gravity wave perturbations (Ford 1994), the scattering of semi-diurnal barotropic tides by topographic features on the ocean floor, fluctuations of wind stress and atmospheric pressure at the ocean surface, and amplification of internal waves by mesoscale-eddy fields and shear flows (Fabrikant 1991; Stepanyants & Fabrikant 1989; Troitskaya & Fabrikant 1989).

6.1. *Model equations and their solutions*

BIG waves are periodic oscillations in a static background vertical density profile $\rho_0(z)$. The corresponding background vertical pressure profile is determined by $\partial_z p_0(z) = -g\rho_0(z)$ with the boundary condition that $p_0(z=0) = p_a$, where p_a is the ambient atmospheric pressure at the ocean surface. Stability clearly requires that $\rho_0(z)$ increase with depth. The linearized shallow-water equations yield solutions of the form

$$\left.\begin{aligned} \delta p_\mu(\boldsymbol{r},z,t) &\equiv p_\mu(\boldsymbol{r},z,t) - p_0(z) = \delta\hat{p}_\mu(z;\boldsymbol{k})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}-\mathrm{i}\omega_\mu(\boldsymbol{k})t}, \\ \delta\rho_\mu(\boldsymbol{r},z,t) &\equiv \rho_\mu(\boldsymbol{r},z,t) - \rho_0(z) = \delta\hat{\rho}_\mu(z;\boldsymbol{k})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}-\mathrm{i}\omega_\mu(\boldsymbol{k})t}, \\ \boldsymbol{v}_\mu(\boldsymbol{r},z,t) &= \hat{\boldsymbol{v}}_\mu(z;\boldsymbol{k})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}-\mathrm{i}\omega_\mu(\boldsymbol{k})t}, \end{aligned}\right\} \tag{6.1}$$

where $\mu = 1, 2, 3, \ldots$ is a mode index. The pressure profiles $\delta\hat{p}_\mu(z;\boldsymbol{k})$ and dispersion relations $\omega_\mu(\boldsymbol{k})$ are determined by the eigenvalue equation

$$(\omega^2 - f^2)\partial_z\left[\frac{\partial_z\delta\hat{p}}{N^2}\right] + \frac{\rho_* k^2}{\rho_0}\delta\hat{p} = 0, \tag{6.2}$$

in which ρ_* is some fixed characteristic density (its vertical mean, say), and $N(z) \equiv \sqrt{(g/\rho_*)|\partial_z\rho_0(z)|}$ is the Brunt–Väisälä frequency. The remaining profiles are then determined in terms of these solutions by

$$\left.\begin{aligned}&\delta\hat{\rho} = -\frac{1}{g}\partial_z\delta\hat{p}, \quad \hat{v}_z = \frac{\mathrm{i}\omega}{\rho^* N^2}\partial_z\delta\hat{p},\\ &\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{v}^\perp = -\frac{\mathrm{i}\omega}{\rho_*}\partial_z\left[\frac{\partial_z\delta\hat{p}}{N^2}\right], \quad \mathrm{i}\hat{\boldsymbol{z}}\cdot\boldsymbol{k}\times\boldsymbol{v}^\perp = -\frac{\omega^2}{\rho_* f}\partial_z\left[\frac{\partial_z\delta\hat{p}}{N^2}\right] - \frac{k^2}{f\rho_0}\delta\hat{p}.\end{aligned}\right\} \tag{6.3}$$

The vertical and the horizontal velocities are related through the incompressibility relation $\partial_z\hat{v}_z + \mathrm{i}\boldsymbol{k}\cdot\hat{\boldsymbol{v}}^\perp = 0$. The boundary condition that v_z must vanish on the ocean floor, $z = -H_0$, is imposed, which then implies that $\partial_z\delta\hat{p}(-H_0) = 0$. The fact that the pressure just beneath the surface, $\delta p(z = 0) = g\rho_*\eta$, differs from p_a by the perturbation in the hydrostatic pressure due to the wave amplitude $\eta(x, y)$, leads to the additional boundary condition $(\partial_z + N^2/g)\delta\hat{p}(z = 0) = 0$ at the free surface.

In general the eigenvalue equation (6.2) must be solved numerically. In the present work we will treat only the exactly soluble case of a linear density profile in the Boussinesq approximation. Thus one sets $N(z) \equiv N_0$ a constant, and one sets $\rho_0(z)/\rho^* \equiv 1$ in the coefficient of the second term in (6.2). The pressure profiles are then solutions to the harmonic oscillator equation with the above boundary conditions:

$$\left.\begin{aligned}&\delta\hat{p}(z) = \delta\hat{p}_0\cos(\kappa_z\pi z/H_0 + \phi),\\ &\tan(\phi) = \tan(\kappa_z\pi) = N_0^2 H_0/\pi g\kappa_z,\\ &\omega(\boldsymbol{k})^2 = f^2 + c^2k^2, \quad c \equiv N_0H_0/\kappa_z\pi.\end{aligned}\right\} \tag{6.4}$$

Wave solutions with frequency smaller than f therefore do not exist. For typical oceanographic applications, $H_0 \sim 1\,\mathrm{km}$, $2\pi/N_0 \sim 600\,\mathrm{s}$, and hence $N_0^2H_0/\pi g \sim 0.003$ is very small (note that $N_0^2H_0/g = H_0|\partial_z\rho_0|/\rho_* = [\rho_0(-H_0) - \rho_0(0)]/\rho_*$ is just the fractional change in density through the water column). The solutions to the second line of (6.4) are then $\kappa_z \simeq n$, $n = 0, 1, 2, \ldots$. The lowest-order mode, $n = 0$, corresponds to shallow-water surface waves in which the entire water column moves in phase. The leading corrections in this case to $\kappa_z = 0$ yield $\phi = \kappa_z\pi \simeq \sqrt{N_0^2H_0/g}$, and $c = \sqrt{gH_0} \sim 100\,\mathrm{m\,s^{-1}}$. These waves are insensitive to the density gradient.

We will work with the lowest-order internal mode, $n = 1$, as the simplest possible model of BIG waves. To fit the observed water column best, one should in this case really interpret $H_0/2 \sim 500\,\mathrm{m}$ to be the thermocline depth. The speed is then $c = N_0H_0/\pi \sim 3\,\mathrm{m\,s^{-1}}$, in good agreement with observed values. The depth profiles are given by

$$\left.\begin{aligned}&\delta\hat{p}(z) = \delta\hat{p}_0\cos(\pi z/H_0), \quad \delta\hat{\rho}_0 = \frac{\pi}{gH_0}\delta\hat{p}_0\sin(\pi z/H_0),\\ &v_z(z) = -\mathrm{i}\omega(\boldsymbol{k})\rho_* N_0^2 H_0\delta\hat{p}_0\sin(\pi z/H_0),\\ &\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{v}^\perp = \frac{\mathrm{i}\omega(\boldsymbol{k})}{\rho_* c^2}\delta\hat{p}_0\cos(\pi z/H_0), \quad \mathrm{i}\hat{\boldsymbol{z}}\cdot\boldsymbol{k}\times\boldsymbol{v}^\perp = \frac{f}{\rho_* c^2}\delta\hat{p}_0\cos(\pi z/H_0).\end{aligned}\right\} \tag{6.5}$$

Notice that the ratio of the amplitudes of the longitudinal (compressional) to transverse (vortical) part of the horizontal velocity field is $\omega(\boldsymbol{k})/f = \sqrt{1 + R_0^2k^2} = O(1)$, where $R_0 \equiv c/f$ is the Rossby radius of deformation. Thus despite the fact that the vertical velocity essentially vanishes at the fluid surface, the horizontal velocity has a strong compressional component: since incompressibility demands only that $\partial_z v_z = -\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{v}^\perp$, it is sufficient that the vertical velocity have a finite gradient at the surface.

From the last two lines of (6.5), the velocity profile appearing in (2.5) takes the form

$$\hat{\boldsymbol{e}}(\boldsymbol{k};z) = \frac{1}{\mathcal{N}}[(\hat{\boldsymbol{k}} - \mathrm{i}\gamma\hat{\boldsymbol{z}} \times \hat{\boldsymbol{k}})\cos(\pi z/H_0) - \mathrm{i}\kappa\hat{\boldsymbol{z}}\sin(\pi z/H_0)] \tag{6.6}$$

in which $\gamma \equiv f/\omega(\boldsymbol{k})$ and $\kappa \equiv H_0 k/\pi$. If the depth integral of $|\hat{\boldsymbol{e}}(\boldsymbol{k};z)|^2$ is normalized to unity, then $\mathcal{N} = \sqrt{(1+\gamma^2+\kappa^2)H_0/2}$. From (2.5) and (2.6) one then has the relationship

$$a(\boldsymbol{k}) = \frac{\omega(\boldsymbol{k})\mathcal{N}}{\rho_* k c^2}\delta\hat{p}_0(\boldsymbol{k}) = \frac{g\omega(\boldsymbol{k})\mathcal{N}}{kc^2}\hat{\eta}_0(\boldsymbol{k}), \tag{6.7}$$

where $\hat{\eta}_0(\boldsymbol{k}) = (\rho_* g)^{-1}\delta\hat{p}_0(\boldsymbol{k})$ is the corresponding Fourier amplitude of the surface height variation. More generally, away from the surface the wave height, defined by the oscillation amplitude of a given constant-density surface at a given depth z, is given to linear order by $\delta\hat{\rho}_0(z)/|\partial_z\rho_0(z)| = g\delta\hat{p}_0(z)/\rho_* N_0^2$. This formula fails near $z = 0$ due to the neglect of the N^2/g term in the surface pressure boundary condition. From the second line of (6.5), the wave height at the thermocline depth $z = -H_0/2$ is then a factor $\pi g/N_0^2 H_0 \sim 300$ times larger than the surface value, $\hat{\eta}_0$.

From (2.7), the frequency spectrum, which vanishes for $|\omega| < f$, is given by

$$\hat{S}_{lm}(z;\omega) = \frac{|\omega|}{c^2}[\langle\hat{f}(\boldsymbol{k})\hat{e}_l(\boldsymbol{k},z)\hat{e}_m^*(\boldsymbol{k},z)\rangle_S\theta(\omega) + \langle\hat{f}(\boldsymbol{k})\hat{e}_l^*(\boldsymbol{k},z)\hat{e}_m(\boldsymbol{k},z)\rangle_S\theta(-\omega)], \tag{6.8}$$

in which $\boldsymbol{k} = k(\omega)\hat{\boldsymbol{k}}$, with $k(\omega) = \sqrt{(\omega^2 - f^2)/c^2}$, $\gamma = f/\omega[k(\omega)] = f/|\omega|$, and $\langle\cdot\rangle_S$ denotes an angular average over the unit sphere, i.e. over the two-dimensional unit vector $\hat{\boldsymbol{k}}$. The horizontal and vertical traces are given by

$$\left.\begin{aligned} \hat{S}(z;\omega) &\equiv \hat{S}_{xx}(z;\omega) + \hat{S}_{yy}(z;\omega) = \hat{S}(\omega)\cos^2(\pi z/H_0), \\ \hat{S}_{zz}(z;\omega) &= \frac{\kappa^2}{1+\gamma^2}\hat{S}(\omega)\sin^2(\pi z/H_0), \\ \hat{S}(\omega) &\equiv \frac{(1+\gamma^2)|\omega|}{c^2\mathcal{N}^2}\langle\hat{f}(\boldsymbol{k})\rangle_S = \frac{1+\gamma^2}{1-\gamma^2}\frac{g^2}{c^2}\hat{S}_\eta(\omega), \end{aligned}\right\} \tag{6.9}$$

where $\hat{S}_\eta(\omega) = (|\omega|/c^2)\langle\hat{f}_\eta(\boldsymbol{k})\rangle_S$ is the surface height frequency spectrum. The quantity $\hat{S}(\omega)$ represents the (entirely horizontal in this case) frequency spectrum at the surface of the fluid.

6.2. *Stokes drift*

Consider now the computation of the Stokes drift. We require as input the tensor

$$\begin{aligned} \hat{R}^i_{lm}(z;\omega) &= \lim_{z'\to z}\int\frac{\mathrm{d}^2k}{(2\pi)^2}(\boldsymbol{k}, -\mathrm{i}\partial_z)_i\hat{\boldsymbol{\Phi}}(\boldsymbol{k},\omega;z,z') \\ &= \begin{cases} \dfrac{|\omega|k(\omega)}{c^2}\langle\hat{k}_i f(\boldsymbol{k})[\hat{e}_l(\boldsymbol{k};z)\hat{e}_m^*(\boldsymbol{k};z)\theta(\omega) \\ \qquad\qquad -\hat{e}_l^*(\boldsymbol{k};z)\hat{e}_m(\boldsymbol{k};z)\theta(-\omega)]\rangle_S, \quad i = x \text{ or } y \\ \dfrac{\pi|\omega|}{\mathrm{i}c^2H_0}\langle f(\boldsymbol{k})[\hat{\varepsilon}_l(\boldsymbol{k};z)\hat{e}_m^*(\boldsymbol{k};z)\theta(\omega) + \hat{\varepsilon}_l^*(\boldsymbol{k};z)\hat{e}_m(\boldsymbol{k};z)\theta(-\omega)]\rangle_S, \quad i = z, \end{cases} \end{aligned} \tag{6.10}$$

where

$$\hat{\varepsilon}(\boldsymbol{k};z) = \frac{H_0}{\pi}\partial_z \hat{\boldsymbol{e}}(\boldsymbol{k};z)$$
$$= \frac{1}{\mathcal{N}}[(-\hat{\boldsymbol{k}} + \mathrm{i}\gamma\hat{\boldsymbol{z}} \times \hat{\boldsymbol{k}})\sin(\pi z/H_0) - \mathrm{i}\kappa\hat{\boldsymbol{z}}\cos(\pi z/H_0)]. \tag{6.11}$$

One obtains then, from (5.31),

$$u_l(z) = \sum_m \int_0^\infty \frac{\mathrm{d}\omega}{2\pi\omega}[\hat{R}^m_{lm}(z;\omega) - \hat{R}^m_{lm}(z;-\omega)]$$
$$= \int_0^\infty \frac{k(\omega)\,\mathrm{d}\omega}{2\pi c^2\mathcal{N}}\Big\{\langle\hat{f}(\boldsymbol{k})[\hat{e}_l(\boldsymbol{k};z) + \hat{e}^*_l(\boldsymbol{k};z)]\rangle_S \cos(\pi z/H_0)$$
$$+\langle\hat{f}(\boldsymbol{k})[\hat{\varepsilon}_l(\boldsymbol{k};z) + \hat{\varepsilon}^*_l(\boldsymbol{k};z)]\rangle_S \sin(\pi z/H_0)\Big\}$$
$$= \cos(2\pi z/H_0)\int_{-\infty}^\infty \frac{\mathrm{d}\omega}{2\pi}\frac{k(\omega)}{(1+\gamma^2)|\omega|}\hat{S}(\omega)U_l(\omega), \tag{6.12}$$

where $\hat{S}(\omega)$ was defined in (6.8), and

$$\boldsymbol{U}(\omega) = \frac{\langle\hat{f}(\boldsymbol{k})(\hat{\boldsymbol{k}},0)\rangle_S}{\langle\hat{f}(\boldsymbol{k})\rangle_S} \tag{6.13}$$

quantifies the anisotropy of the spectrum at wavenumber magnitude $|\boldsymbol{k}| = k(\omega)$. The integrand again vanishes for $|\omega| < f$. The factor $1/(1+\gamma^2)$ indicates that it is only the longitudinal (compressive) part of the horizontal motion that contributes to the drift.

The depth profile of the drift velocity is quite remarkable: it is parallel to the anisotropy direction near the top and bottom of the fluid, antiparallel in the neighbourhood of the thermocline depth $z = -H_0/2$, and has nodes at $z = -H_0/4, -3H_0/4$. The vertical average of the drift velocity therefore always vanishes. The reason for this structure is apparent if one thinks about the structure of the particle trajectories. There are two types of contribution to the drift. One, which is apparent in (5.32), is intrinsic to the back-and-forth horizontal motion and arises even in the absence of any vertical motion. The second is intrinsic to the vertical motion and arises from the fact that the horizontal velocity varies with depth. Thus, if, for example, the horizontal velocity decreases with depth, then during a *clockwise* tracer trajectory the forward horizontal velocity is slightly larger at the top than is the reverse horizontal velocity at the bottom of the particle's trajectory; there will be a slight net forward motion during each wave cycle. This will be seen in the next section to be the case for surface gravity waves, and the effect *adds* to the purely horizontal part of the drift.

Now, from (6.6) one sees that above $z = -H_0/2$ one actually has the opposite circumstance: the horizontal velocity indeed decreases with depth, but the phase of the horizontal and vertical motions is such that the projections of the trajectories onto the $(\hat{\boldsymbol{k}}, z)$ vertical plane are *counterclockwise*. This means that this effect yields a *negative* addition to the overall drift. Due to the cosine profile, close to the surface the depth gradient of the horizontal velocity is very small, and due to the sine profile the vertical velocity is very small. The vertical effect is then very small, the positive horizontal contribution dominates, and the drift is along $\hat{\boldsymbol{k}}$. In fact, at $z = 0$ the result (6.12) and (6.13) matches exactly the purely translation-invariant result (5.32) with $d = 2$. However, as the depth increases, both the depth gradient of the horizontal velocity and the amplitude of the vertical velocity increase. The forward horizontal contribution therefore decreases while at the same time the backward vertical contribution increases. At $z = -H_0/4$ there is a precise cancellation and the net drift vanishes. Below $z = -H_0/4$ the vertical contribution dominates and the net drift is opposite to $\hat{\boldsymbol{k}}$. At $z = -H_0/2$ where the horizontal velocity vanishes and the vertical velocity is maximal, the reverse drift maximizes. Below $z = -H_0/2$ the entire scenario occurs in reverse. The drift becomes positive at $z = -3H_0/4$, and is maximally forward once again at the bottom, $z = -H_0$.

6.3. *The diffusion tensor*

Let us write the diffusion tensor in the form

$$D^{(2)}_{lm}(z) = D^{(2)}_{1,lm}(z) + D^{(2)}_{2,lm}(z), \tag{6.14}$$

in which $D^{(2)}_{1,lm}(z)$ contains the contributions in (5.31) involving $\rho^{(1)}(z)$, and $D^{(2)}_{2,lm}(z)$ contains the remaining terms. Noting that only derivatives with respect to z survive, and using the result that $u_z(z) \equiv 0$, one obtains

$$D^{(2)}_{1,lm}(z) = \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{4\pi\omega^2} [\partial_z u_l(z) \hat{\mathsf{S}}^{+}_{zm}(z;\omega) + (l \leftrightarrow m)]. \tag{6.15}$$

From (6.8) it follows that

$$\hat{\mathsf{S}}^{+}_{zm}(z;\omega) = \frac{\kappa}{1+\gamma^2} \hat{S}(\omega) \sin(\pi z/H_0) \left[\gamma \cos(\pi z/H_0) \hat{\boldsymbol{z}} \times \boldsymbol{U}(\omega) + \kappa \hat{\boldsymbol{z}} \sin(\pi z/H_0)\right]_m, \tag{6.16}$$

and one obtains then from (6.12) and (6.16):

$$\begin{aligned} D^{(2)}_{1,lm}(z) = -&\left\{ \int \frac{\mathrm{d}\omega}{\pi\omega^2} \frac{k(\omega)}{(1+\gamma^2)|\omega|} \hat{S}(\omega) \boldsymbol{U}(\omega) \right\}_l \\ &\times \left\{ \sin^2(2\pi z/H_0) \hat{\boldsymbol{z}} \times \int \frac{\mathrm{d}\omega}{4\pi\omega^2} \frac{k(\omega)\gamma}{1+\gamma^2} \hat{S}(\omega) \boldsymbol{U}(\omega) \right. \\ &\left. + \sin^3(\pi z/H_0) \cos(\pi z/H_0) \hat{\boldsymbol{z}} \int \frac{\mathrm{d}\omega}{\pi\omega^2} \frac{\kappa k(\omega)}{1+\gamma^2} \hat{S}(\omega) \right\}_m + (l \leftrightarrow m) \end{aligned} \tag{6.17}$$

For the computation of $D^{(2)}_{2,lm}(z)$ we require in addition to (6.8) and (6.10) the tensor

$$\bar{T}^{ij}_{lm}(z;\omega) = \lim_{z'\to z}\int \frac{\mathrm{d}^2k}{(2\pi)^2}(\boldsymbol{k},-\mathrm{i}\partial_z)_i(\boldsymbol{k},\mathrm{i}\partial'_z)_j\Phi(\boldsymbol{k},\omega;z,z')$$

$$= \frac{|\omega|}{c^2}\begin{cases} k(\omega)^2[\langle\hat{f}(\boldsymbol{k})\hat{k}_i\hat{k}_j\hat{e}_l(\boldsymbol{k};z)\hat{e}^*_m(\boldsymbol{k};z)\rangle_S\theta(\omega) \\ \qquad +\langle\hat{f}(\boldsymbol{k})\hat{k}_i\hat{k}_j\hat{e}^*_l(\boldsymbol{k};z)\hat{e}_m(\boldsymbol{k};z)\rangle_S\theta(-\omega)], \quad i,j = x \text{ or } y \\ \mathrm{i}(\pi/H_0)k(\omega)[\langle\hat{f}(\boldsymbol{k})\hat{k}_i\hat{e}_l(\boldsymbol{k};z)\hat{\varepsilon}^*_m(\boldsymbol{k};z)\rangle_S\theta(\omega) \\ \qquad -\langle\hat{f}(\boldsymbol{k})\hat{k}_i\hat{k}_j\hat{e}^*_l(\boldsymbol{k};z)\hat{\varepsilon}_m(\boldsymbol{k};z)\rangle_S\theta(-\omega)], \quad i = x \text{ or } y;\ j = z \\ -\mathrm{i}(\pi/H_0)k(\omega)[\langle\hat{f}(\boldsymbol{k})\hat{k}_j\hat{\varepsilon}_l(\boldsymbol{k};z)\hat{e}^*_m(\boldsymbol{k};z)\rangle_S\theta(\omega) \\ \qquad -\langle\hat{f}(\boldsymbol{k})\hat{k}_j\hat{\varepsilon}^*_l(\boldsymbol{k};z)\hat{e}_m(\boldsymbol{k};z)\rangle_S\theta(-\omega)], \quad i = z; j = x \text{ or } y \\ (\pi/H_0)^2[\langle\hat{f}(\boldsymbol{k})\hat{\varepsilon}_l(\boldsymbol{k};z)\hat{\varepsilon}^*_m(\boldsymbol{k};z)\rangle_S\theta(\omega) \\ \qquad -\langle\hat{f}(\boldsymbol{k})\hat{\varepsilon}^*_l(\boldsymbol{k};z)\hat{\varepsilon}_m(\boldsymbol{k};z)\rangle_S\theta(-\omega)], \quad j = x \text{ or } y; i = z, \end{cases} \tag{6.18}$$

where again $\boldsymbol{k} = k(\omega)\hat{\boldsymbol{k}}$. One obtains now for the various terms in (5.31):

$$-\frac{\mathrm{i}}{2}\sum_{i,j}\hat{S}^-_{ij}(z;\omega)[\partial_j\hat{R}^{i-}_{lm}(z;\omega) - \partial_i\hat{R}^{j-}_{lm}(z;\omega)] = \frac{\omega^2k(\omega)^2}{4c^4\mathcal{N}^4}$$

$$\times\begin{cases} -2\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'[\hat{k}_l\hat{k}_m + \gamma^2\hat{q}_l\hat{q}_m]\rangle_{S,S'}\sin^2(2\pi z/H_0), & l,m = x \text{ or } y \\ \kappa\gamma\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\hat{q}_l\rangle_{S,S'}\sin(4\pi z/H_0), & l = x \text{ or } y; m = z \\ 2\kappa^2\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\rangle_{S,S'}\sin^2(2\pi z/H_0), & l = m = z, \end{cases} \tag{6.19a}$$

$$\sum_{i,j}\hat{S}_{ij}(z;\omega)\bar{T}^{ij+}_{lm}(z;\omega) = \frac{\omega^2k(\omega)^2}{4c^4\mathcal{N}^4}$$

$$\times\begin{cases} 4\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')A_1(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)[\hat{k}_l\hat{k}_m + \gamma^2\hat{q}_l\hat{q}_m]\rangle_{S,S'}, & l,m = x \text{ or } y \\ 2\kappa\gamma\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')[A_2(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{q}_l - \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}'\hat{k}_l]\rangle_{S,S'}\sin(2\pi z/H_0), & l = x \text{ or } y; m = z \\ \kappa^2\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')[1 + (\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')^2 + \gamma^2(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}')^2]\rangle_{S,S'}\sin^2(2\pi z/H_0), & l = m = z, \end{cases} \tag{6.19b}$$

$$\sum_{i,j}\hat{R}^{i+}_{lj}(z;\omega)\hat{R}^{j+}_{mi}(z;\omega) = -\frac{\omega^2k(\omega)^2}{4c^4\mathcal{N}^4}$$

$$\times\begin{cases} 4\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')\Big\{\gamma^2A_3(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{q}_l\hat{q}'_m + A_4(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{k}_l\hat{k}'_m \\ \qquad +A_5(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)[\hat{q}_l\hat{k}'_m - \hat{k}_l\hat{q}'_m]\Big\}\rangle_{S,S'}, & l,m = x \text{ or } y \\ 2\kappa\gamma\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')\{A_6(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{q}_l \\ \qquad +\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}'(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}' - 1)\cos^2(\pi z/H_0)\hat{k}_l\}\rangle_{S,S'}\sin(2\pi z/H_0), & l = x \text{ or } y; m = z \\ \kappa^2\langle\hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')[1 + (\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')^2]\rangle_{S,S'}\sin^2(2\pi z/H_0), & l = m = z, \end{cases} \tag{6.19c}$$

$$\sum_{i,j} \hat{R}^{i-}_{lj}(z;\omega)\hat{R}^{j-}_{mi}(z;\omega) = \frac{\omega^2 k(\omega)^2}{4c^4\mathcal{N}^4}$$

$$\times \begin{cases} 4\langle \hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')\{A_3(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{k}_l\hat{k}'_m + \gamma^2 A_4(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{q}_l\hat{q}'_m \\ \quad +\gamma^2 A_5(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)[\hat{q}_l\hat{k}'_m - \hat{k}_l\hat{q}'_m]\}\rangle_{S,S'}, & l,m = x \text{ or } y \\ 2\kappa\gamma\langle \hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')[A_7(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{k}_l + A_8(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)\hat{q}_l]\rangle_{S,S'} \\ \quad \times \sin(2\pi z/H_0), & l = x \text{ or } y; m = z \\ -\kappa^2\langle \hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')[\gamma^2(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}')^2 + 2\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}']\rangle_{S,S'}\sin^2(2\pi z/H_0), & l = m = z, \end{cases} \tag{6.19d}$$

where matrix elements for $l = z$, $m = x$ or y are obtained from those for $l = x$ or y, $m = z$ simply by substituting m for l, and in which, to condense the notation, we have defined the unit vector $\hat{\boldsymbol{q}} = \hat{\boldsymbol{z}} \times \hat{\boldsymbol{k}}$ orthogonal to $\hat{\boldsymbol{k}}$, and

$$\left.\begin{aligned} A_1(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= [(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')^2 + \gamma^2(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}')^2]\cos^4(\pi z/H_0) + \sin^4(\pi z/H_0), \\ A_2(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= [(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')^2 + \gamma^2(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}')^2]\cos^2(\pi z/H_0) - \sin^2(\pi z/H_0), \\ A_3(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= (\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')^2\cos^4(\pi z/H_0) + \sin^4(\pi z/H_0), \\ A_4(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= \tfrac{1}{2}\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\sin^2(2\pi z/H_0) - \gamma^2(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}')^2\cos^4(\pi z/H_0), \\ A_5(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= \hat{\boldsymbol{k}}'\cdot\boldsymbol{q}[\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\cos^4(\pi z/H_0) + \tfrac{1}{4}\sin^2(2\pi z/H_0)], \\ A_6(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= (\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')^2\cos^2(\pi z/H_0) - \sin^2(\pi z/H_0), \\ A_7(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}'[\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\cos^2(\pi z/H_0) + \sin^2(\pi z/H_0)], \\ A_8(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) &= \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\sin^2(\pi z/H_0) - [\gamma^2(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}')^2 - \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}']\cos^2(\pi z/H_0). \end{aligned}\right\} \tag{6.20}$$

Summing these contributions, one obtains the remarkably compact result

$$D^{(2)}_{2,lm}(z) = \int_{-\infty}^{\infty} \frac{d\omega}{4\pi}\frac{k(\omega)^2}{\omega^2}\hat{S}(\omega)^2\Theta_{lm}(z;\omega) \tag{6.21}$$

with angular factor

$$\Theta_{lm}(z;\omega) = \frac{\langle \hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}')A(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z)[\gamma^2(\hat{\boldsymbol{q}},0)_l(\hat{\boldsymbol{q}}-\hat{\boldsymbol{q}}',0)_m + (\hat{\boldsymbol{k}},0)_l(\hat{\boldsymbol{k}}+\hat{\boldsymbol{k}}',0)_m]\rangle_{S,S'}}{(1+\gamma^2)^2\langle \hat{f}(\boldsymbol{k})\rangle_S^2}, \tag{6.22}$$

with coefficient

$$A(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}',z) = [(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')^2 + \gamma^2(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}')^2]\cos^4(\pi z/H_0) + \sin^4(\pi z/H_0) - \tfrac{1}{2}\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\sin^2(2\pi z/H_0). \tag{6.23}$$

For an isotropic spectrum $\hat{f}(\boldsymbol{k}) = \hat{f}(k)$ the drift velocity $\boldsymbol{u}$ (6.12) vanishes identically and hence so does $D^{(2)}_{1,lm}(z)$. Only $D^{(2)}_{2,lm}(z)$ survives and purely horizontal diffusion

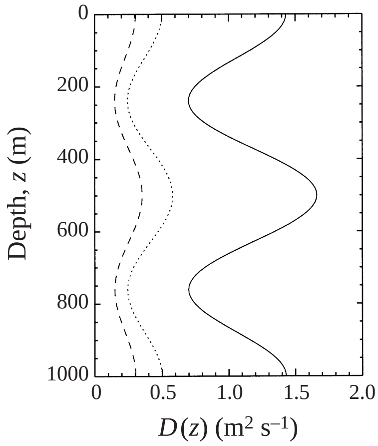

FIGURE 2. Diffusion constant due to BIG waves calculated using the isotropic result (6.24) and the theoretical spectrum (6.26) (suggested in Galzman 1996*b* based on the direct energy cascade due to weakly nonlinear resonant four-wave interactions). Parameters employed are ocean depth $H_0 = 1\,\text{km}$ and Kelvin wave phase speed $c = 3\,\text{m}\,\text{s}^{-1}$. The different curves correspond to different latitudes ϕ: solid curve, $\phi = 20°$, hence $R \simeq 60\,\text{km}$; dotted curve $\phi = 40°$, hence $R \simeq 32\,\text{km}$; dashed curve, $\phi = 60°$, hence $R \simeq 24\,\text{km}$. The Kolmogorov constant $\alpha = 1.6$ and energy dissipation rate $\epsilon = 10^{-11}\,\text{m}^2\,\text{s}^{-3}$ are chosen for illustration purposes to yield reasonable physical values for the r.m.s. horizontal fluid velocity u_0 and r.m.s. sea surface height $\bar{\eta}$. For $\phi = 20°, 40°, 60°$, respectively, one finds $u_0 = 24.4\,\text{cm}\,\text{s}^{-1}$, $22.0\,\text{cm}\,\text{s}^{-1}$, $20.9\,\text{cm}\,\text{s}^{-1}$, and $\bar{\eta} = 4.9\,\text{cm}$, 4.4 cm, 4.2 cm. Notice the interesting oscillation with depth. More realistic (nonlinear) density profiles will lead to quantitatively different diffusion-constant profiles, but the qualitative features should survive.

$D^{(2)}_{lm}(z) = D\delta_{lm}(1-\delta_{lz})(1-\delta_{mz})$ is obtained, with

$$D = \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{16\pi} \frac{k(\omega)^2}{\omega^2} \left[\cos^4(\pi z/H_0) + \frac{2}{1+\gamma^2} \sin^4(\pi z/H_0) - \frac{1-\gamma^2}{2(1+\gamma^2)^2} \sin^2(2\pi z/H_0) \right] \hat{S}(\omega)^2, \tag{6.24}$$

in which the results of Appendix B have been used to compute the remaining angular averages. Recall that $\gamma = f/|\omega|$, and the integrand vanishes for $|\omega| < f$. At the fluid surface, $z = 0$, only the first term survives. From (6.9), the quantity $\hat{S}(\omega) = \hat{S}(z = 0; \omega)$ is the horizontal contribution to the frequency spectrum at the surface, and, with $d = 2$, is the quantity most closely analogous to that defined for the fully translation-invariant result in (5.36). One immediately sees then that the fully translation-invariant result (5.39) for the diffusion coefficient is identical to (6.24) at $z = 0$. This is not surprising as the vertical motion vanishes at the surface and therefore does not contribute to the diffusion coefficient. At the thermocline depth, $z = -H_0/2$, only the second term survives. Figure 2 illustrates the theoretical result (6.24) using a semi-empirical theoretical spectrum of BIG wave turbulence suggested in Glazman (1996*b*) and discussed further in the following subsection (see equation (6.26) below). The resulting $1\,\text{m}^2\,\text{s}^{-1}$ order of magnitude for the diffusion constant corresponds also to the horizontal dashed line in figure 1. The latitudinal dependence of the diffusion coefficient is due to the fact that the lower-frequency cutoff of the fluid velocity spectrum employed in the present calculations is controlled by the Coriolis frequency f which is a strong function of geographic latitude.

Although the contribution to the diffusion from $D^{(2)}_{2,lm}$ is purely horizontal even

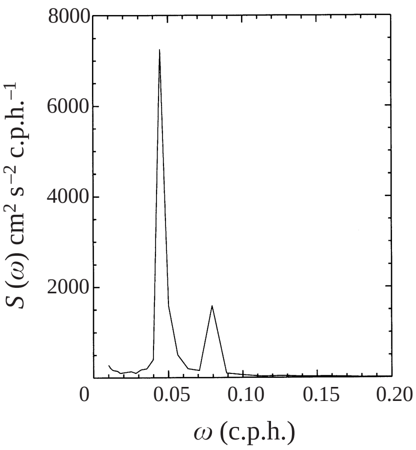

FIGURE 3. Frequency spectrum of ocean currents measured at 34.9° N, 55° W in the North Atlantic at a fixed depth of 600 m. The local ocean depth is 5506 m. Units of frequency are cycles per hour (c.p.h.). The figure has been replotted based on figure 2 of Fu (1981).

for a general anisotropic spectrum, $D^{(2)}_{1,lm}$ contains, in addition to purely horizontal components, *off-diagonal* elements with $l = z$ or $m = z$ (but not $l = m = z$). The existence of these components, and their physical meaning, is sufficiently subtle that a general discussion of them will be deferred to § 8. We comment here only that these components do not lead to vertical transport, but only to a peculiar vertical variation in the horizontal transport. Only a non-zero component $D^{(2)}_{zz}$ can accomplish the former. In § 8 it is argued that $D^{(2)}_{zz}$ should vanish to all orders in u_0/c_0 and that horizontal wave systems of the types studied here therefore cannot produce net vertical transport.

6.4. *Estimates from oceanographic data*

For the purposes of the present paper we will offer only a very preliminary discussion of oceanographic data. A comprehensive discussion must await not only the extraction of relevant experimental spectra from a greater variety of ocean regions, but also a proper theoretical analysis of the BIG wave mode shapes using more realistic nonlinear vertical density profiles. This will be presented in a future publication.

As illustrated in (6.4) above, the dispersion relation in the f-plane approximation (neglecting spatial variation of the Coriolis parameter f) is given by $\omega^2 = f^2 + c^2k^2$, where c is the (constant) wave phase speed in the absence of the Earth's rotation (known as the Kelvin wave speed) which is set by the Brunt–Väisälä frequency and the thermocline depth. Wave solutions with frequency smaller than f therefore do not exist. This turns out to be a slight oversimplification: see below. An experimental frequency spectrum of horizontal velocity fluctuations in the upper ocean layer, based on data reported in Fu (1981), is illustrated in figure 3. Notice that the main spectral peak actually *spans* the Coriolis frequency, f (the second peak is due to the semi-diurnal tide). This means that the modes dominating the spectrum are actually of sufficiently long wavelength that the f-plane approximation is no longer valid: evanescent tails of lower-frequency waves that exist at slightly smaller latitudes actually contribute substantially to the spectrum. This has a very strong effect on estimates of the diffusion constant as the spectral factor $k(\omega)$ in (6.21) vanishes at f, thereby suppressing the integrand in the region near the peak. In future work we will

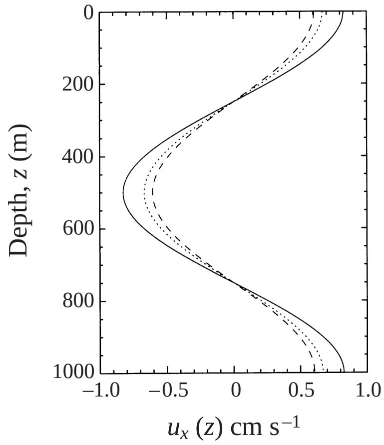

FIGURE 4. Stokes drift vertical profile due to BIG waves. Parameters and spectrum are the same as those used in figure 2. For simplicity, a strongly anisotropic form $\Upsilon(\theta) = 2\pi\delta(\theta)$, corresponding to an effectively one-dimensional wave structure, is chosen for the angular dependence in (6.25). Notice the sign change with depth. More realistic (nonlinear) density profiles will lead to quantitatively different drift profiles, but the qualitative features should survive.

account for this effect properly by using the β-plane approximation (which allows linear variation of $f(y) = f(y_0) + \beta(y - y_0)$ with latitude) to compute the dispersion relation. This will lead to a non-vanishing $k(\omega)$ near the peak, vastly increasing the result for the mean drift and diffusion constant. At present we can rigorously estimate only the contribution from the isotropic short-scale range of the inertia–gravity wave spectrum for which the isotropic dispersion law (6.4) is valid. Only in this range are our main results (6.21) and (6.24) strictly valid. The range of integration is therefore reduced to the interval (ω_0, ∞) with $\omega_0 > f$.

Since there are no data for the full angular dependence of the wavenumber spectrum, estimates for the Stokes drift will be based on the empirical form

$$\hat{f}(\boldsymbol{k}) = \bar{f}(k)\Upsilon(\theta) \tag{6.25}$$

with normalization $\int_{-\pi}^{\pi} \mathrm{d}\theta \Upsilon(\theta) = 2\pi$. The angular averaged spectrum $\bar{f}(k)$ is taken from the semi-empirical theoretical form (6.26) below. In figure 4 we show results based on (6.12) and (6.25) with $\omega_0 = 1.1f$ and, for simplicity, $\Upsilon(\theta) = 2\pi\delta(\theta)$, corresponding to an effectively one-dimensional spectrum of waves travelling toward positive x. Parameters are identical to those used in figure 2 (see also below). The magnitude of the drift u_x originating from this part of the spectrum is of order $1\,\mathrm{cm\,s^{-1}}$. We anticipate that this value will rise significantly (perhaps by a factor of two or more) when the entire spectral peak is taken into account in the β-plane approximation.

Estimates for the diffusion constant will be based for simplicity on the isotropic result (6.24), i.e. $\Upsilon(\theta) \equiv 1$. The experimental spectrum plotted in figure 3 yields $D^{(2)}$ at the ocean surface ranging from $70\,\mathrm{cm^2\,s^{-1}}$ for $\omega_0 = 1.05f$ to $50\,\mathrm{cm^2\,s^{-1}}$ for $\omega_0 = 1.5f$. These values are much smaller than those based on the theoretical spectrum (6.26) below and illustrated in figure 2 for two reasons. First, as mentioned above, we anticipate that the actual $D^{(2)}$ will be much greater when the entire spectral peak is taken into account in the β-plane approximation. A crude estimate of the effect may be obtained by replacing $k(\omega)/\omega$ by $\max\{k(\omega)/\omega, 1/c\}$ in (5.39) so that this coefficient retains the finite value $1/c$ near f. Taking c to be the Kelvin wave speed, and beginning the integration at the local minimum to the left of the spectral peak, we obtain the result $D^{(2)} \simeq 1300\,\mathrm{cm^2\,s^{-1}}$. Second, the linear density profile and the value of the depth H_0 used to compute (6.24) are not really relevant to the case of figure 3: The latter are for a deep '1.5 layer' ocean in which both layers (dramatically different in their thickness – upper layer 500 m and lower layer 4500 m) have uniform densities and hence a relatively constant velocity throughout each layer. Proper theoretical calculations for the transport coefficients in such cases will be presented in future work. The relatively large r.m.s. fluid velocity $u_0 \sim 20\,\mathrm{cm\,s^{-1}}$ employed to obtain figure 2 reflects the fact that in shelf regions long-wave amplitudes are intensified in comparision with those in deep ocean regions. There are at present no experimental measurements of the depth dependence of the diffusion constant with which to compare the theory, even at the qualitative level.

If the tidal peak (which is absent in many ocean regions) is neglected, observed spectra of BIG waves are in good agreement with theoretical spectra suggested in Glazman (1996*b*). In particular, the velocity spectrum of BIG waves at the fluid surface, $z = 0$, is given there in the scaled form (5.40) by

$$s(x) = a_0 \frac{(x^4+7)(x^2+1)}{(x^2)^{1/3}(x^4-1)^{5/3}}, \tag{6.26}$$

valid for $x \equiv \omega/f > 1$, where a_0 is the normalization required by (5.40). This form is actually not normalizable as its integral diverges at $x = 1$, a result also due to the breakdown of the f-plane approximation. As explained earlier, we take $s(x) \equiv 0$ for $x \leqslant 1+\delta$, for various values of δ (see below). Substituting this expression into the last line of (6.9) and using the scaling (5.40) one then obtains for the frequency spectrum at the fluid surface $\hat{S}(\omega) = (u_0^2/f)s(x)$ with

$$\frac{u_0^2}{f} = \frac{\pi\alpha}{6a_0} R^{4/3}\epsilon^{1/3}, \tag{6.27}$$

where ϵ is again the flux of wave energy (or, equivalently, the energy dissipation rate normalized by the fluid mass density), $R = c/f$ is the Rossby radius of deformation, and α is a dimensionless coefficient analogous to the Kolmogorov constant in fluid turbulence. One then obtains $D^{(2)}$ in the form

$$D^{(2)} = B\frac{\alpha^2\epsilon^{2/3}R^{5/3}}{c} \tag{6.28}$$

where the dimensionless coefficient is given by

$$B = \frac{\pi}{288}\int_{1+\delta}^{\infty} \mathrm{d}x \frac{(x^4+7)^2(x^2+1)^2}{x^{10/3}(x^4-1)^{10/3}}. \tag{6.29}$$

The constant B defined here is related to the constant B_d (with dimension $d = 2$) in (5.41) via $B_2 = (6a_0/\pi)^2 B$. Notice that although the spectrum (6.26) decays as a power law at large frequencies, and hence leads to significant small-scale structure, equation (6.29) converges at large x implying that the velocity field is nevertheless sufficiently smooth not to violate the requirements of the Taylor expansion (5.10). For $\delta = 0.1$ numerical integration yields $a_0 \simeq 0.358$, $B \simeq 0.388$ (and hence $B_2 = 0.181$). For purposes of illustration, the curves in figure 2 then correspond to the choice $c = 3\,\mathrm{m\,s^{-1}}$, $\alpha = 1.6$ and $\epsilon = 10^{-11}\,\mathrm{m^2\,s^{-3}}$ with three different latitudes $\phi = 20°$, $40°$, and $60°$, yielding $R \simeq 60$ km, 32 km, 24 km, and $u_0 \simeq 24.4\,\mathrm{cm\,s^{-1}}$, $22.0\,\mathrm{cm\,s^{-1}}$, $20.9\,\mathrm{cm\,s^{-1}}$ respectively. The corresponding r.m.s. wave heights, found using (6.9) and $\bar{\eta}^2 = \int_{-\infty}^{\infty}(\mathrm{d}\omega/2\pi)\hat{S}_\eta(\omega)$, are $\bar{\eta} \simeq 4.9$ cm, 4.4 cm, 4.2 cm, respectively. By way of comparison, using the same estimate $c = 3\,\mathrm{m\,s^{-1}}$, the experimental spectrum figure

3, taken in a mid-latitude region $\phi \simeq 45°$, corresponds to $R = 30\,\mathrm{km}$ and yields the smaller values $u_0 \simeq 7\,\mathrm{cm\,s^{-1}}$ and $\epsilon \sim 10^{-14}\,\mathrm{m^2\,s^{-3}}$.

The fact that the Rossby radius increases as the latitude decreases – exceeding 200 km near the equator – indicates a possible role of BIG waves as a factor in horizontal transport in tropical regions. In contrast to the meso-scale and large-scale eddies, which can propagate only westward, BIG waves can move in both westward and eastward directions and are hence capable of eastwardly transporting the various tracers in these regions.

7. Applications to wind-generated surface gravity waves

For deep-water surface gravity waves the vertical profile in (2.5) (with only a single mode) takes the purely longitudinal form

$$\hat{\boldsymbol{e}}(\boldsymbol{k}, z) = (\hat{\boldsymbol{k}}, -\mathrm{i})\sqrt{k}\mathrm{e}^{kz}, \quad z < 0, \tag{7.1}$$

and the dispersion law is $\omega(\boldsymbol{k}) = \sqrt{gk}$. The wavenumber spectrum, defined in (2.7), then takes the form

$$F_{lm}(\boldsymbol{k}; z, z') = \hat{f}(\boldsymbol{k})(\hat{\boldsymbol{k}}, -\mathrm{i})_l(\hat{\boldsymbol{k}}, \mathrm{i})_m k\mathrm{e}^{k(z+z')}, \tag{7.2}$$

in which $\hat{f}(\boldsymbol{k})$ is defined by (2.6). The surface height field $\eta(\boldsymbol{r}, t)$ is governed by the same set of Fourier amplitudes $a(\boldsymbol{k})$:

$$\eta(\boldsymbol{r}, t) = \frac{1}{\sqrt{g}} \int \frac{\mathrm{d}^2k}{(2\pi)^2} [a(\boldsymbol{k})\,\mathrm{e}^{\mathrm{i}[\boldsymbol{k}\cdot\boldsymbol{r} - \omega(\boldsymbol{k})t]} + \text{c.c.}], \tag{7.3}$$

and therefore has a corresponding spectrum

$$\Phi_\eta(\boldsymbol{k}, \omega) = F_\eta(\boldsymbol{k}) 2\pi\delta(\omega - \sqrt{gk}) + F_\eta(-\boldsymbol{k}) 2\pi\delta(\omega + \sqrt{gk}) \tag{7.4}$$

in which the wavenumber spectrum is given by

$$F_\eta(\boldsymbol{k}) = \frac{1}{g}\hat{f}(\boldsymbol{k}). \tag{7.5}$$

Defining the angular-averaged amplitude spectrum

$$\bar{f}(k) = \langle \hat{f}(\boldsymbol{k}) \rangle_S \equiv \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi} \hat{f}(k, \theta), \tag{7.6}$$

where $\boldsymbol{k} = k[\cos(\theta), \sin(\theta)]$, the associated frequency height spectrum is

$$\hat{S}_\eta(\omega) = \int \frac{\mathrm{d}^2k}{(2\pi)^2} \Phi_\eta(\boldsymbol{k}, \omega) = \frac{2|\omega|^3}{g^3} \bar{f}(\omega^2/g). \tag{7.7}$$

The mean-square height fluctuation is

$$\bar{\eta}^2 = \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \hat{S}_\eta(\omega) = 2 \int \frac{\mathrm{d}^2k}{(2\pi)^2} F_\eta(\boldsymbol{k}) = \int_0^{\infty} \frac{k\mathrm{d}k}{\pi g} \bar{f}(k). \tag{7.8}$$

The tensors $\hat{\boldsymbol{S}}$, $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{T}}$ are then given by

$$\left.\begin{aligned}\hat{S}_{lm}(z;\omega) &= \int \frac{\mathrm{d}^2k}{(2\pi)^2}\varphi(\boldsymbol{k},\omega;z)(\hat{\boldsymbol{k}},-\mathrm{i})_l(\hat{\boldsymbol{k}},\mathrm{i})_m,\\ \hat{R}^i_{lm}(z;\omega) &= \int \frac{\mathrm{d}^2k}{(2\pi)^2}k\varphi(\boldsymbol{k},\omega;z)(\hat{\boldsymbol{k}},-\mathrm{i})_l(\hat{\boldsymbol{k}},\mathrm{i})_m(\hat{\boldsymbol{k}},-\mathrm{i})_i,\\ \bar{T}^{ij}_{lm}(z;\omega) &= \int \frac{\mathrm{d}^2k}{(2\pi)^2}k^2\varphi(\boldsymbol{k},\omega;z)(\hat{\boldsymbol{k}},-\mathrm{i})_l(\hat{\boldsymbol{k}},\mathrm{i})_m(\hat{\boldsymbol{k}},-\mathrm{i})_i(\hat{\boldsymbol{k}},\mathrm{i})_j.\end{aligned}\right\} \tag{7.9}$$

in which we have defined

$$\varphi(\boldsymbol{k},\omega;z) = 2\pi k\mathrm{e}^{2kz}[\hat{f}(\boldsymbol{k})\delta(\omega-\sqrt{gk})+\hat{f}(-\boldsymbol{k})\delta(\omega+\sqrt{gk})]. \tag{7.10}$$

For future reference, it is useful to define the horizontal trace

$$\begin{aligned}\hat{S}(z;\omega) &\equiv \hat{S}_{xx}(z;\omega)+\hat{S}_{yy}(z;\omega) = \hat{S}_{zz}(z;\omega)\\ &= \frac{2|\omega|^5}{g^3}\mathrm{e}^{2\omega^2z/g}\bar{f}(\omega^2/g) = \omega^2\mathrm{e}^{2\omega^2z/g}\hat{S}_\eta(\omega).\end{aligned} \tag{7.11}$$

7.1. *Stokes drift*

The drift velocity (last line of (5.31)) now emerges straightforwardly as

$$\boldsymbol{u}(z) = \int \frac{\mathrm{d}^2k}{(2\pi)^2}\frac{4k^{3/2}}{\sqrt{g}}\mathrm{e}^{2kz}\hat{f}(\boldsymbol{k})(\hat{\boldsymbol{k}},0) = \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{2|\omega|}{g}\hat{S}(z;\omega)\boldsymbol{U}(\omega), \tag{7.12}$$

where $\boldsymbol{U}(\omega)$ is given by the angular average

$$\boldsymbol{U}(\omega) \equiv \frac{\langle \hat{f}(g^{-1}\omega^2\hat{\boldsymbol{k}})(\hat{\boldsymbol{k}},0)\rangle_S}{\bar{f}(\omega^2/g)}, \tag{7.13}$$

which represents the 'dipole moment' of the amplitude spectrum at frequency ω. The result (7.13) agrees precisely with that of HH (whose height spectrum $F(\boldsymbol{k})$ is normalized so that $F_\eta(\boldsymbol{k}) = 2\pi^2F(\boldsymbol{k})$). Not surprisingly, the drift is purely horizontal and decreases exponentially with depth. It is interesting to compare this result to the fully translation-invariant case (5.32) in which the full problem is replaced by an effective two-dimensional problem where the vertical component of the velocity is simply dropped, and the amplitude spectrum $\hat{f}(\boldsymbol{k})$ is assumed unchanged. One obtains in this case

$$\boldsymbol{u}_{2\mathrm{d}} = \tfrac{1}{2}\boldsymbol{u}(z=0), \tag{7.14}$$

so that the vertical motion enhances the purely two-dimensional drift by a factor of two. This occurs presumably because the decrease of the velocity with depth enhances the slight imbalance between the back and forth horizontal motions as the tracer particle bobs up and down.

7.2. *Diffusion tensor*

Consider now the diffusion tensor. As in (6.14), let $D^{(2)}_{1,lm}$ denote the contribution from the terms in (5.31) involving $\rho^{(1)}_l(\boldsymbol{x}) = u_l(z)$. Noting again that only derivatives with respect to z survive, and using the result that $u_z(z) \equiv 0$, one obtains again the form

(6.15). It follows from (7.9) that $\hat{S}_{zm}^{+}$ vanishes unless $m = z$, and one obtains then

$$
\begin{aligned}
D_{1,lm}^{(2)}(z) &= \partial_z u_l(z) \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{4\pi\omega^2} \hat{S}(z;\omega)\delta_{mz} + (l \leftrightarrow m) \\
&= \left\{ \int \frac{\mathrm{d}\omega}{4\pi\omega^2} \hat{S}(z;\omega) \right\} \left\{ \int \frac{\mathrm{d}\omega}{2\pi} \frac{4|\omega|^3}{g^2} \hat{S}(z;\omega)[U_l(\omega)\delta_{mz} + U_m(\omega)\delta_{lz}] \right\}. \qquad (7.15)
\end{aligned}
$$

The contributions from the remaining terms are computed as follows. Performing the summations over i and j one obtains

$$
\begin{aligned}
&-\frac{\mathrm{i}}{2} \sum_{i,j} \hat{S}_{ij}^{-}(z;\omega)[\partial_j \hat{R}_{lm}^{i-}(z;\omega) - \partial_i \hat{R}_{lm}^{j-}(z;\omega)] \\
&= \int \frac{\mathrm{d}^2k}{(2\pi)^2} \int \frac{\mathrm{d}^2k'}{(2\pi)^2} \varphi(\boldsymbol{k},\omega;z)\varphi(\boldsymbol{k}',\omega;z) 2k'^2 \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}' \begin{cases} \hat{k}_l \hat{k}_m, & l,m = x \text{ or } y \\ 1, & l = m = z \\ 0, & \text{otherwise}, \end{cases} \qquad (7.16a)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{i,j} \hat{S}_{ij}^{+}(z;\omega) \bar{T}_{lm}^{ij+}(z;\omega) \\
&= \int \frac{\mathrm{d}^2k}{(2\pi)^2} \int \frac{\mathrm{d}^2k'}{(2\pi)^2} \varphi(\boldsymbol{k},\omega;z)\varphi(\boldsymbol{k}',\omega;z) k'^2 [1 + (\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}')^2] \begin{cases} \hat{k}_l \hat{k}_m, & l,m = x \text{ or } y \\ 1, & l = m = z \\ 0, & \text{otherwise}, \end{cases}
\end{aligned}
$$

(7.16b)

$$
\begin{aligned}
&\sum_{i,j} \hat{R}_{lj}^{i+}(z;\omega) \hat{R}_{mi}^{j+}(z;\omega) \\
&= \int \frac{\mathrm{d}^2k}{(2\pi)^2} \int \frac{\mathrm{d}^2k'}{(2\pi)^2} \varphi(\boldsymbol{k},\omega;z)\varphi(\boldsymbol{k}',\omega;z) kk' \begin{cases} 2\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}' \hat{k}_l \hat{k}'_m, & l,m = x \text{ or } y \\ -[1 + (\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}')^2], & l = m = z \\ 0, & \text{otherwise}, \end{cases}
\end{aligned}
$$

(7.16c)

$$
\begin{aligned}
&\sum_{i,j} \hat{R}_{lj}^{i-}(z;\omega) \hat{R}_{mi}^{j-}(z;\omega) \\
&= \int \frac{\mathrm{d}^2k}{(2\pi)^2} \int \frac{\mathrm{d}^2k'}{(2\pi)^2} \varphi(\boldsymbol{k},\omega;z)\varphi(\boldsymbol{k}',\omega;z) kk' \begin{cases} [1 + (\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}')^2] \hat{k}_l \hat{k}'_m, & l,m = x \text{ or } y \\ -2\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}', & l = m = z \\ 0, & \text{otherwise}. \end{cases}
\end{aligned}
$$

(7.16d)

With the aid of the identity

$$
\int \frac{\mathrm{d}\omega}{4\pi\omega^2} \varphi(\boldsymbol{k},\omega;z)\varphi(\boldsymbol{k}',\omega;z) = \frac{4\pi k^3}{(gk)^{3/2}} \mathrm{e}^{4kz} \hat{f}(\boldsymbol{k}) \hat{f}(\boldsymbol{k}') \delta(k - k'), \qquad (7.17)
$$

summing these four terms and performing the frequency integral, one obtains a contribution

$$D^{(2)}_{2,lm}(z) = \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{4\pi} \frac{\omega^2}{2g^2} \hat{S}(z\,;\omega)^2 \Theta_{lm}(\omega) \tag{7.18}$$

to the diffusion tensor, in which

$$\Theta_{lm}(\omega) = \frac{\langle (1 + \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}')^2 (\hat{\boldsymbol{k}} + \hat{\boldsymbol{k}}', 0)_l (\hat{\boldsymbol{k}} + \hat{\boldsymbol{k}}', 0)_m \hat{f}(\boldsymbol{k}) \hat{f}(\boldsymbol{k}') \rangle_{S,S'}}{\bar{f}(\omega^2/g)^2}. \tag{7.19}$$

Notice that all components in which either $l = z$ or $m = z$ cancel out exactly. Thus $D^{(2)}_{2,lm}(z)$ represents horizontal diffusion only, and contributes components of the diffusion tensor that are entirely complementary to those contributed by $D^{(2)}_{1,lm}(z)$. The result (7.18) with (7.19) exactly matches the result of HH for the full diffusion tensor, who, because they computed only purely horizontal components, missed the contribution $D^{(2)}_{1,lm}(z)$. The total diffusion tensor (6.14) at $O(u_0^4\tau^2/c_0^2)$ contains, as for BIG waves, contributions to all components except for $l = m = z$. As indicated at the end of § 6.3, the physical meaning of the off-diagonal vertical components will be clarified in § 8 below.

In the case of an isotropic spectrum, once again only $D^{(2)}_{2,lm}(z)$ survives. Using the results of Appendix B, one finds that $\Theta_{lm} = \frac{5}{2}$, independent of ω, and the diffusion tensor simplifies to the diagonal form $D^{(2)}_{lm}(z) = D^{(2)}(z)\delta_{lm}(1 - \delta_{lz})(1 - \delta_{mz})$, with

$$D^{(2)}(z) = \int \frac{\mathrm{d}\omega}{4\pi} \frac{5\omega^2}{4g^2} \hat{S}(z\,;\omega)^2. \tag{7.20}$$

Comparing again to the effective two-dimensional result (5.38) (with $k(\omega)^2 = \omega^4/g^2$), we find

$$D^{(2)}_{2\mathrm{d}} = \tfrac{1}{5} D^{(2)}(z = 0). \tag{7.21}$$

Once again, the vertical variations enhance the horizontal transport (this time by a factor of 5) over that predicted by the purely two-dimensional theory.

For an anisotropic spectrum, the off-diagonal components $D^{(2)}_{2,lm}(z)$ of the diffusion tensor are, as for BIG waves, in general non-zero, and a detailed discussion of them is deferred to § 8.

7.3. *Estimates from oceanographic data*

We turn finally to estimates for the transport parameters from oceanographic data. Since Stokes drift due to surface gravity waves has been discussed previously in great detail (see e.g. Phillips 1977), we will discuss here only the diffusion constant. The rather general, empirically based form of the surface height spectrum $S_\eta(\omega)$ is given by

$$S_\eta(\omega) = \beta g^2 (u_w/g)^{4\mu} \omega^{-5+4\mu} \Theta(\omega/\omega_0), \tag{7.22}$$

where u_w is the mean wind speed above the sea surface, g is the acceleration due to gravity, $\omega_0 = g/u_w\xi$ is the spectral peak frequency, ξ is the *wave age* defined as the ratio of the phase speed of the waves at the spectral peak to the wind velocity u_w, $\Theta(\omega/\omega_0)$ is a smoothed step (Heaviside) function which imposes a smooth cutoff at frequencies below the spectral peak, and β is a dimensionless Phillips constant which is a slowly decreasing function of the wave age (Glazman 1994). At a small wave age, the spectrum is dominated by the Phillips saturation range (Phillips 1985) in which $\mu = 0$. At intermediate wave ages, $\mu \approx 1/4$ and (7.22) reduces to the Zakharov–Filonenko spectrum (Zakharov & Filonenko 1966) as controlled by the direct inertial

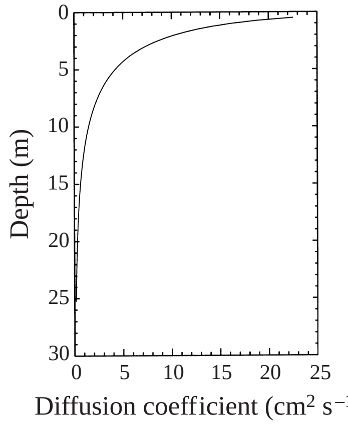

FIGURE 5. Diffusion constant due to surface gravity waves near the ocean surface with wind speed $U = 10\,\mathrm{m\,s^{-1}}$ and wave age $\xi = 2$ using the model spectrum (7.22).

cascade of energy toward smaller scales (analogous to the usual Kolmogorov cascade in isotropic turbulence). At a large wave age, one must use $\mu \approx 1/3$ which corresponds to the *inverse* cascade of wave action (Zakharov & Zaslavskii 1982). In general, in the absence of unambiguous inertial ranges, the exponent $\mu = \mu(\xi)$ may be viewed as an effective exponent that generally increases with wave age (Glazman 1994; Glazman *et al.* 1996). Notice that because the wavelengths of interest are in the range 1–300 m, Coriolis effects are negligible and no latitude dependence occurs in (7.22).

In order to omit effects of smaller-scale ripples influenced by surface tension and other extraneous factors, a high-frequency cutoff can be imposed in (7.22), thus forcing an exponential decay at frequencies above those associated with the 'intrinsic microscale' of the surface gravity range (Glazman & Weichman 1989). However, for a finite depth z, the velocity spectrum (7.11) experiences a sufficiently fast high-frequency roll-off to make the use of an intrinsic microscale unnecessary. Equations (7.11), (7.22) and the gravity wave dispersion law $\omega^2 = gk$ allow one to estimate the diffusion coefficient based on (7.20). For a typical open ocean case where $U = 10\,\mathrm{m\,s^{-1}}$ and $\xi = 2$, one finds the dependence of the diffusion constant on depth shown in figure 5. We conclude from this figure that surface-gravity-wave-induced diffusion is appreciable only within a few metres of the ocean surface. Below 10 m depths the effects of surface gravity waves are negligible.

8. Off-diagonal vertical diffusion

8.1. *Absence of vertical transport*

The existence of non-vanishing off-diagonal vertical components $D_{lz} = D_{zl}$, with $l \neq z$, raises the important issue of vertical transport in the immediate subsurface layer of the ocean. The surface inhibits the growth of three-dimensional eddies, which would otherwise give rise to classical eddy-induced diffusion. The latter is then too small to explain observed magnitudes of vertical transport of various tracers (such as oxygen absorbed at the ocean surface) to significant depths. Vertical transport by waves, if it were found to be of similar magnitude to that of horizontal transport, could provide a simple explanation for the strong vertical diffusion in the subsurface layer. Unfortunately, as we will now discuss, only a non-zero *diagonal* vertical component

D_{zz} would yield net vertical transport, and we will argue that neither SG waves nor BIG waves produce a non-zero value of this coefficient at any order in u_0/c_0.

The fact that in the absence of vertical drift, $u_z = 0$, only D_{zz} leads to net vertical transport is seen by considering the horizontally integrated tracer concentration

$$\bar{\psi}^H(z) = \int \mathrm{d}x\,\mathrm{d}y\,\bar{\psi}(x, y, z), \tag{8.1}$$

which from (4.1) and (4.2) then satisfies the equation

$$\partial_t \bar{\psi}^H = \partial_z[D_{zz}(z)\partial_z \bar{\psi}^H] - \partial_z[E_{zzz}(z)\partial_z^2 \bar{\psi}^H] + \cdots, \tag{8.2}$$

in which it is assumed that all transport coefficients depend on z only. To order $(u_0/c_0)^2$ we have seen that both D_{zz} and E_{zzz} (see (5.23) for the latter) vanish identically, and one concludes that $\bar{\psi}^H(z)$ is *conserved*, and no net change in the mean concentration at a fixed depth can occur, precluding net transport from the surface.

To see why, in general, there should be no net vertical transport, consider the subsequent dynamics of surfaces that begin as horizontal planes at arbitrary fixed depth in a quiescent fluid. These correspond to constant-density surfaces in the case of BIG waves. Conservation of both the fluid density and the tracer concentration implies that a passive tracer particle initially on a particular surface will remain on that same surface for all time. The key point now is that the SG and BIG wave motions do not allow large vertical excursions of these surfaces. This is obvious for the air–sea interface whose vertical fluctuations $\eta(\boldsymbol{r}, t) = O(u_0\tau_0)$ have essentially Gaussian statistics with variance limited by the overall wave amplitude, but is equally true of subsurface fluctuations, which are exponentially smaller with depth for SG waves, and oscillate with depth for BIG waves. Such motions couple to gravity, and large excursions occur only with exponential rarity and, even when they do occur, 'radiate' rapidly away through wave motions in a time comparable to the dominant wave period τ_0 (in fact, sufficiently large localized excursions can lead to wave breaking). We conclude, therefore, that there cannot be vertical diffusion without actual mixing of parcels of fluid at different depths, and subsequent 'entanglement' of the surfaces, as would occur in the presence, for example, of vertical eddies. Neither BIG nor SG wave motions allow for this.

Matching the above general argument with the behaviour of the diffusion equation is somewhat subtle. The z-component $\langle j_\psi^z \rangle = -D_{zx}\partial_x \bar{\psi} - D_{zy}\partial_y \bar{\psi}$ of the mean concentration current is generally locally non-zero even though it vanishes on average, apparently contradicting the above argument which appears to forbid even local vertical transport. However, the conservation of $\bar{\psi}^H(z)$ at any given z means that any increase in concentration at one horizontal position is always counterbalanced instantaneously by a decrease in concentration at other horizontal positions. In particular, concentration can never appear at a height where it did not exist previously at some other horizontal position. The net result is that the off-diagonal terms simply provide a new mechanism for horizontal transport. The existence of locally non-zero $\langle j_\psi^z \rangle$ does not contradict the general argument since the waves do produce (bounded) vertical fluctuations in the tracer particle position. Although we cannot at this point make a rigorous argument to this effect, it is presumably these vertical fluctuations that, upon statistical averaging, produce the locally finite $\langle j_\psi^z \rangle$. Recall that it is also in the nature of waves that for every fluid parcel being transported upward at one horizontal point, there must be another being transported equally downward at some other, perhaps distant, point.

8.2. *Negative diffusion*

The off-diagonal elements of the diffusion tensor point to a very interesting physical effect: they lead to negative diffusion along one principal axis of the diffusion tensor. Although at first sight this result may appear quite alarming, as negative diffusion may be expected to cause finite-time mathematical singularities in the evolution of the concentration field $\bar{\psi}(\boldsymbol{x}, t)$, we will in fact argue (non-rigorously, but we hope persuasively) below that in the present problem, due to (i) a rapid equilibration at short times $t < \tau$, and (ii) the presence of a depth-dependent horizontal drift $\boldsymbol{u}(z)$, the evolution of $\bar{\psi}(\boldsymbol{x}, t)$ is perfectly smooth and well defined.

The principal axes of the diffusion tensor are the eigenvectors of $\boldsymbol{D}$. Since $\boldsymbol{D}$ is symmetric, these eigenvectors are orthogonal. If $\boldsymbol{D}$ is position- (e.g. depth) dependent, these axes, and the corresponding eigenvalues, will also in general be position dependent. The following considerations apply at an arbitrary fixed point in space, which will be suppressed from the notation. In order to simplify the algebra, let us also assume that the initial coordinate system is chosen in such a way that the horizontal 2×2 sub-block of $\boldsymbol{D}$ is diagonal (if it were not, a rotation about the vertical axis would make it so, and such a rotation would leave D_{zz} invariant) with positive diagonal elements. Thus, we consider $\boldsymbol{D}$ of the form

$$\boldsymbol{D} = \begin{pmatrix} D_{xx} & 0 & D_{xz} \\ 0 & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & 0 \end{pmatrix} \tag{8.3}$$

with $D_{xx}, D_{yy} > 0$. The determinant of $\boldsymbol{D}$ is

$$\det(\boldsymbol{D}) = -D_{xx}D_{yz}^2 - D_{yy}D_{xz}^2, \tag{8.4}$$

which is therefore negative, irrespective of the signs of D_{xz} and D_{yz}. Since the determinant is the product of the three eigenvalues of $\boldsymbol{D}$, we conclude that at least one of them is negative. On the other hand, the trace of the matrix is

$$\mathrm{tr}(\boldsymbol{D}) = D_{xx} + D_{yy} > 0. \tag{8.5}$$

Since the trace is the sum of the eigenvalues, we conclude that at least one of them is positive. We conclude then that exactly one of the eigenvalues is negative. An illustrative analytic calculation may be done in the case in which one of the off-diagonal components vanishes, say $D_{xz} = 0$. One then finds eigenvalues

$$D_1 = D_{xx}, \quad D_2 = D_{yy}/2 + \sqrt{D_{yy}^2/4 + D_{yz}^2}, \quad D_3 = D_{yy}/2 - \sqrt{D_{yy}^2/4 + D_{yz}^2}, \tag{8.6}$$

in which D_3 is clearly the negative one. The corresponding (unnormalized) eigenvectors are

$$\boldsymbol{d}_1 = (1, 0, 0), \quad \boldsymbol{d}_2 = (0, 1, D_{yz}/D_2), \quad \boldsymbol{d}_3 = (0, D_3/D_{yz}, 1). \tag{8.7}$$

Suppose that D_{yz} is small compared to D_{xx} and D_{yy}. Then $D_2 \approx D_{yy}[1 + (D_{yz}/D_{yy})^2]$ and $D_3 \approx -D_{yz}^2/D_{yy}$. Then $\boldsymbol{d}_2$ points mainly along $\hat{\boldsymbol{y}}$, with a small vertical tilt, while $\boldsymbol{d}_3$ points mainly vertically, with a small horizontal tilt along $\hat{\boldsymbol{y}}$.

The principal axes of $\boldsymbol{D}$ are the directions along which the three-dimensional diffusion process locally factors into three orthogonal uncorrelated one-dimensional diffusion processes, with diffusion constants D_i, $i = 1, 2, 3$. A negative diffusion constant D_3 indicates a *time-reversed* diffusion process in which the concentration field *contracts* along $\boldsymbol{d}_3$ into the plane defined by $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$. In spite of this contraction, the absence of net vertical transport demonstrated in the previous subsection shows that its z-component is precisely compensated by the z-components of the positive diffusive *spreading* due to the combination of D_1 and D_2.

Positive diffusion has a smoothing effect on an initial concentration field. Negative diffusion *decreases* the smoothness of an initial concentration field, and if the latter is insufficiently smooth, can lead to finite time singularities (i.e. clumping of the concentration) in its subsequent evolution. For example, for a simple homogeneous one-dimensional diffusion process with diffusion constant $D < 0$, the evolution of the spatial Fourier transform $\bar{\psi}(k,t)$ of the concentration field is given by $\bar{\psi}(k,t) = \mathrm{e}^{|D|k^2t}\bar{\psi}(k,0)$. If $\bar{\psi}(k,0) \sim \mathrm{e}^{-r_0^2k^2}$ at large wavenumber k, then a singularity will occur at time $t_s = r_0^2/|D|$. More generally, in the presence of inhomogeneities and drift, diffusion will dominate the evolution at short times, and this estimate will remain locally accurate for sufficiently small r_0.

A related issue is that of conservation of positivity of an initially positive concentration field. The fact that the diffusion equation is a semi-rigorously derived, asymptotic large scale, long time description of the dynamics of a positive concentration field implies that $\bar{\psi}$ ought to remain positive if it is initially so. Under one-dimensional negative diffusion a positive initial condition of the form $\bar{\psi}(x) = 1 - \epsilon\cos(kx)$, with $0 < \epsilon < 1$, evolves according to $\bar{\psi}(x,t) = 1 - \epsilon\,\mathrm{e}^{|D|k^2t}\cos(kx)$ and therefore experiences no finite time singularities, but will violate positivity for $t > t_p \equiv (1/|D|k^2)\ln(1/\epsilon)$.

8.3. *Drift shear, rapid initial spreading, and avoidance of finite time singularities*

Since, physically, one certainly does not expect singularities or negative concentrations to occur in passive scalar transport by waves, one must now understand how they may be avoided within the transport equation we have derived. We shall now describe two sufficient conditions for avoidance of these unphysical phenomena, and how these conditions are physically satisfied. We do not claim to establish these conditions rigorously, but it should become clear that they appeal to all of the correct mathematical and physical elements. The first condition is the existence of drift shear, i.e. a depth-dependent horizontal drift $\boldsymbol{u}(z)$. One sees from (6.15) that off-diagonal vertical diffusion is present *only* when $\partial_z\boldsymbol{u}(z)$ is finite. The presence of drift shear means that tracer particles at different depths will, in the absence of diffusion, drift apart linearly with time as $t\Delta\boldsymbol{u}$, where $\Delta\boldsymbol{u} = \boldsymbol{u}(z_1) - \boldsymbol{u}(z_2)$ is the difference in the drift velocities of the two particles. Since diffusion, whether positive or negative, has a length scale $\sqrt{|D|t}$ varying with the square root of time, drift will always provide the dominant transport mechanism at *large* times such that $t \gg t_u \equiv |D|/|\Delta u|^2$. If the initial mean tracer concentration field is smooth on scale r_0 (in the sense implied by the definition of r_0 at the end of the previous subsection), then, roughly speaking, a finite time singularity can be avoided if $t_s > t_u$ for two particles initially separated by a distance r_0, i.e. if

$$r_0 > \sqrt{|D|/|\partial_z\boldsymbol{u}|}. \tag{8.8}$$

From (6.14), one may estimate the coefficient of $\partial_z\boldsymbol{u}$ to be $O(u_0^2\tau^2)$, and we obtain the requirement $r_0 > u_0\tau \sim d_0 \sim (u_0/c_0)\lambda_0$, where d_0 is the distance travelled by a tracer particle in a dominant wave period. This is the second condition: the initial mean concentration field should be smooth on the scale of the typical tracer fluctuation during a wave period. Given such an initial condition, the drift shear will eliminate any finite time singularities in the mean transport equation that could otherwise arise from negative diffusion.

The identical smoothness condition is also required to ensure positivity. Just as we have argued that the z-dependent drift allows avoidance of finite time singularities by 'stretching' the concentration field at longer times, we expect also that positivity violation will be avoided so long as t_p is sufficiently large, i.e. once again, if the initial condition is sufficiently smooth.

At first sight it would appear that one may choose an arbitrary initial condition with an arbitrarily small value of r_0. In fact, however, one must recall that implicit in the validity of the diffusion equation (4.1) and (4.2) is the assumption that $t > \tau$: as discussed in §4.1, the crucial Markov property that underlies its microscopic derivation requires that the tracer field be allowed to 'equilibrate' for a decorrelation time before the diffusion equation becomes a valid description of the evolution of the mean concentration field. During this time the tracer particle positions indeed undergo semi-periodic back and forth motion of magnitude d_0 on the scale of the dominant wave period. Upon statistical averaging this appears as an essentially linear-in-time, $\sim u_0 t$, spreading of the probability distribution from a delta-function initial condition. This continues up to times of order the wave period, after which the width saturates to an essentially constant value. The final result is that the initial concentration field becomes smoothed over the scale $r_0 \sim d_0$. It is this initially smoothed concentration field, which now satisfies the second condition above, that may be thought of as serving as a proper initial condition for the diffusion equation, and will have a well-defined, smooth evolution for all time.

As an aside, we comment that the equilibration on time scale τ has a physical manifestation in the *spectrum of passive tracer fluctuations* about the mean, leaving an imprint of the wave field fluctuation spectrum on the tracer spectrum on time scales smaller than t_0. This will be discussed in detail in a separate publication (Weichman & Glazman 2000).

8.4. *Formal considerations: positive definiteness*

We now turn to a more formal demonstration of the validity of the above arguments. This demonstration relies on the notion of matrix positive definiteness. A matrix A_{lm} is said to be positive definite if for any real or complex constants c_l not all zero

$$\sum_{l,m} A_{lm} c_l^* c_m > 0. \tag{8.9}$$

This property implies that all eigenvalues of A_{lm} must be positive: if the c_l are chosen to be the components of a normalized eigenvector corresponding to eigenvalue σ, then

$$\sum_m A_{lm} c_m = \sigma c_l \;\Rightarrow\; \sum_{l,m} A_{lm} c_l^* c_m = \sigma \sum_l |c_l|^2 = \sigma > 0, \tag{8.10}$$

in which the normalization implies that $\sum_l |c_l|^2 = 1$.

Now, recall the definition (4.16) of the Lagrangian position correlation matrix for a tracer particle that starts at point $\boldsymbol{x}$ at time s. This matrix is easily seen to be positive definite since

$$\sum_{l,m} \lambda_{lm}^{(2)}(\boldsymbol{x}, t) c_l^* c_m = \left\langle \left| \sum_l c_l (\boldsymbol{Z}_{\boldsymbol{x}s}(t) - \boldsymbol{x})_l \right|^2 \right\rangle \tag{8.11}$$

must be positive so long as all components of $\boldsymbol{Z}_{\boldsymbol{x}s}(t)$ vary non-trivially in time.

Recall now that the matrix $\rho_{lm}^{(2)}$, related to the diffusion tensor through (5.31), was defined via

$$\rho_{lm}^{(2)}(\boldsymbol{x}, t) = \partial_t \lambda_{lm}^{(2)}(\boldsymbol{x}, t). \tag{8.12}$$

The question now is: does the positive definiteness of $\lambda_{lm}^{(2)}$ imply anything about that of $\rho_{lm}^{(2)}$? In general the answer is no, and we have already seen an example where $\rho_{lm}^{(2)}$ has a negative eigenvalue. However, suppose that for sufficiently large $t > t_D$, $\lambda_{lm}^{(2)}$ becomes linear in t:

$$\lambda_{lm}^{(2)}(\boldsymbol{x}, t) \to \rho_{lm}^{(2)}(\boldsymbol{x})t, \quad t > t_D. \tag{8.13}$$

This is the same as the time scale t_D identified near the beginning of § 4.1. Then, for $t > t_D$,

$$\sum_{l,m} \rho_{lm}^{(2)}(\boldsymbol{x}) c_l^* c_m = \frac{1}{t} \sum_{l,m} \lambda_{lm}^{(2)}(\boldsymbol{x}) c_l^* c_m > 0. \tag{8.14}$$

Now, we have already seen explicitly that $\rho_{lm}^{(2)}$ becomes time independent for $t > \tau$, and (8.14) would then seem to require a positive definite result. However, the hidden assumption here is that that one may take $t_D = \tau$. In fact, the constancy of $\rho_{lm}^{(2)}(\boldsymbol{x})$ only shows that

$$\lambda_{lm}^{(2)}(\boldsymbol{x}, t) \to \lambda_{lm}^{(2)}(\boldsymbol{x}) + \rho_{lm}^{(2)}(\boldsymbol{x})t, \quad t > \tau, \tag{8.15}$$

in which $\lambda^{(2)}(\boldsymbol{x})$ is time independent. Now, for sufficiently large t the second term will dominate. If $\lambda_{lm}^{(2)}(\boldsymbol{x})$ were of the same order as $\rho_{lm}^{(2)}(\boldsymbol{x})\tau$, then one could indeed take $t_D = \tau$. The result of the computations in § 5 is that $\rho_{lm}^{(2)}(\boldsymbol{x}) = O[u_0^2 \tau (u_0/c_0)^2]$, requiring that $\lambda_{lm}^{(2)}(\boldsymbol{x}) = O[(u_0\tau)^2(u_0/c_0)^2]$ for validity of the choice $t_D = \tau$. In fact, as will now be demonstrated, even though $\rho_{lm}^{(2)}(\boldsymbol{x})$ only has contributions at $O(u_0^2/c_0^2)$ in perturbation theory, $\lambda_{lm}^{(2)}(\boldsymbol{x})$ has contributions at *zeroth* order.

At zeroth order in u_0/c_0 one may simply replace the Lagrangian velocity by the Eulerian velocity, and one obtains

$$\begin{aligned} \lambda_{lm}^{(2)}(\boldsymbol{x}, t) &= \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \langle v_l(\boldsymbol{x}, s_1) v_m(\boldsymbol{x}, s_2)\rangle \\ &= \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 C_{lm}(\boldsymbol{x}, \boldsymbol{x}, s_1 - s_2) \\ &= \int_{-t}^t \mathrm{d}s \int_{|s|}^{t-|s|} \mathrm{d}u C_{lm}(\boldsymbol{x}, \boldsymbol{x}, s) \\ &= \int_{-t}^t \mathrm{d}s (t - 2|s|) C_{lm}(\boldsymbol{x}, \boldsymbol{x}, s), \end{aligned} \tag{8.16}$$

in which $u = (s_1 + s_2)/2$ and $s = s_1 - s_2$. So long as the Eulerian correlator $C_{lm}(\boldsymbol{x}, \boldsymbol{x}, s)$ vanishes rapidly for $s > \tau$, one obtains for $t > \tau$

$$\begin{aligned} \lambda_{lm}^{(2)}(\boldsymbol{x}, t) &= t \int_{-\infty}^{\infty} \mathrm{d}s C_{lm}(\boldsymbol{x}, \boldsymbol{x}, s) - 2 \int_{-\infty}^{\infty} \mathrm{d}s |s| C_{lm}(\boldsymbol{x}, \boldsymbol{x}, s) \\ &= t \hat{S}_{lm}(\boldsymbol{x}; 0) + \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{\pi\omega^2} [\hat{S}_{lm}^+(\boldsymbol{x}, 0) - \hat{S}_{lm}^+(\boldsymbol{x}, \omega)] \end{aligned} \tag{8.17}$$

in which $\hat{S}_{lm}^+(\boldsymbol{x}, \omega) = [\hat{S}_{lm}(\boldsymbol{x}, \omega) + \hat{S}_{lm}(\boldsymbol{x}, -\omega)]/2$ is again the even part of the frequency spectrum $\hat{S}_{lm}(\boldsymbol{x}, \omega)$. For waves the spectrum vanishes at zero frequency, and one therefore identifies

$$\lambda_{lm}^{(2)}(\boldsymbol{x}) = -2 \int_{-\infty}^{\infty} \mathrm{d}s |s| C_{lm}(\boldsymbol{x}, \boldsymbol{x}, s) = -\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{\pi\omega^2} \hat{S}_{lm}^+(\boldsymbol{x}, \omega). \tag{8.18}$$

Since $C_{lm} = O(u_0^2)$, this term is indeed of $O(u_0^2\tau^2)$.

Substituting the result (8.18) into (8.15), one then immediately sees that one requires $t_D = O[(c_0^2/u_0)^2\tau] \gg \tau$, and *there is no positive definiteness requirement on the computed* $\rho_{lm}^{(2)}(\boldsymbol{x})$ *on time scales less than this.* On the other hand, on time scales of this size, the result (5.31) for $\rho_{lm}^{(2)}(\boldsymbol{x})$ is no longer valid. The derivation of (5.31) was based on the assumption that t was large compared to τ, but small compared to time scales T_0 on which the tracer particle is transported distances comparable to that over which the statistics of the velocity field vary. A crude estimate for this distance is the dominant wavelength λ_0. One then obtains the estimate $T_0|\boldsymbol{u}| \sim \lambda_0$, and hence

$$T_0 \sim \lambda_0/|\boldsymbol{u}| \sim c_0^2\tau/u_0^2 = t_D, \tag{8.19}$$

in which $\lambda_0 \sim c_0\tau$ and $|\boldsymbol{u}| \sim u_0^2/c_0$ have been used. Thus the original calculation breaks down on precisely the same time scale as the positive definiteness requirement (8.14) comes into play.

To summarize, during the initial time of order the decorrelation time τ, there is an initial saturation of the mean-square drift with magnitude of order $(u_0\tau)^2$ (and this sets the scale over which the transition probability $P(\boldsymbol{x}, t|\boldsymbol{x}', s)$ averages for $t = O(\tau)$ in (4.21) and (4.22)). After this initial saturation, diffusion slowly increases the mean-square drift at rate Dt, but where D is of order $u_0^2\tau(u_0/c_0)^2$. Thus it requires a time t_D of order $(c_0/u_0)^2\tau$ for the diffusion to give a drift comparable to the initial saturation and hence, e.g. for a numerical simulation of wave transport to be able to even resolve the existence of diffusion.

The initial saturation effect also solves the negative diffusion problem because it takes a length of time similar to t_D for the negative diffusion coefficient to become noticeable. However, on this time scale the shear in the mean drift will act to overwhelm the negative diffusion and eliminate any unphysical behaviour in the solution to the diffusion equation – negative diffusion is found *only* in the presence of drift shear. The fundamental feature is that the effects of diffusion can exceed the effects of drift only at short times. Since the initial saturation elimates all possible singularities in this time interval, negative diffusion does not violate any fundamental physical requirements. An effective diffusion equation on time scales large compared to t_D would yield a new diffusion tensor, renormalized mainly by shear in the horizontal drift which then pulls nearby tracer particles apart despite the (smaller) contractive effect of the negative diffusion coefficient, and which by (8.14) must be positive definite. One concludes that the off-diagonal vertical components of this 'fully renormalized' tensor must vanish, and only horizontal diffusion and drift can survive.

It should be emphasized that the diffusion equation derived in §§ 4 and 5 remains perfectly valid on all time scales larger than τ, so long as an initial condition is used which already accounts for the initial saturation $\lambda_{lm}^{(2)}(\boldsymbol{x})$. The latter is effectively produced by a rapid spreading on time scales less than τ that then plays no subsequent role. The 'fully renormalized' diffusion tensor then arises from solving this 'partially renormalized' equation on time scales larger than t_D. No further reference to the 'microscopic' wave field dynamics is required.

The negative diffusion discovered here is a new physical effect that seems not to have been encountered previously. In all previous discussions of which we are aware, positive definiteness of the diffusion tensor has always been a fundamental assumption of the theory. As we have seen, on sufficiently large time scales t_D, positive definiteness of the fully renormalized diffusion tensor is assured, but such a description is clearly inferior to the more microscopic, but still diffusive, description valid on time scales $\tau \ll t_D$ where positive definiteness is no longer required.

9. Conclusions

The advection–diffusion equation (2.8) for the mean concentration of a passive tracer is remarkable in several respects. In contrast to the classical theory of eddy-induced turbulent transport, the wave diffusion problem (as demonstrated first by HH) is amenable to rigorous mathematical treatment employing small-perturbation techniques. Hence, the coefficients entering (2.8), and given by (5.31), represent closed-form expressions for the mean (Stokes) drift and diffusion tensor arising from wave-induced fluctuations of the fluid velocity. The underlying wave motion may be of a very general nature. The only requirement is that the degree of the wave nonlinearity be sufficiently small to justify the perturbation expansion of the (Lagrangian) field variables discussed in § 5.1.

As demonstrated in §§ 6 and 7, computation of the diffusion tensor for particular wave systems is a straightforward task given either an experimental or theoretical form for the velocity spectrum. In the special cases of surface gravity and long internal gravity waves in the ocean, these computations yield diffusion constants comparable to those caused by eddy turbulence if the characteristic scales of eddies are under 10 km. As the mesh size in eddy-resolving numerical ocean models is presently at the 10 km scale, the diffusion mechanisms addressed in the present work emerge as the main factors of unresolved 'turbulent' transport.

Among the most important future applications of the theory, we foresee analysis of tracer transport by baroclinic inertia–gravity waves with frequencies near the Coriolis frequency f_0, where the wavelengths become so large that the f-plane approximation breaks down. As seen in figure 3, the spectrum is in fact dominated by these frequencies. The wave motions in this regime originate from the evanescent tails of BIG waves excited at lower latitudes. These waves are then essentially anisotropic and must be described using the β-plane approximation, which allows for a linear gradient in the Coriolis frequency with latitude. Their large contribution to the spectrum shows that they will have an appreciable affect on diffusion and drift that is not correctly captured by the f-plane approximation used in this work. The generality of the present formulation (allowing, in principle, inhomogeneity of a wave field in any subset of directions) allows immediate application of the theory to these kinds of problems.

Although the mean drift arising from BIG waves is negligible compared to that arising from major ocean currents, it nevertheless appears to be of great interest. Attaining a few cm s^{-1}, this drift is of the same magnitude as ocean interior motions away from major current systems. Moreover, unlike localized ocean currents, BIG waves represent a ubiquitous and permanent feature of ocean dynamics. Therefore, even at the magnitude of a few cm s^{-1}, the BIG-wave-induced Stokes transport may play an important role in ocean climate. The fact that the drift switches sign with depth, for example, means that warmer and cooler water are carried in *opposite* directions, providing enhanced heat transport.

While this work is in many ways an extension of that of HH, one particular result, namely the off-diagonal vertical elements in the diffusion tensor, and its accompanying loss of positive definiteness, is new and completely unexpected. Although these elements do not lead to vertical diffusion, they will contribute to horizontal diffusion. Normally, in the presence of a finite drift, diffusion is a subdominant process. However, for BIG waves the drift vanishes at certain depths, yet maintains there a finite vertical *gradient*. As a result, diffusion will be the dominant transport process over some range of depths. In this region, the off-diagonal elements of the diffusion tensor are in principle of the same order as the horizontal elements. The physical, especially geophysical, importance of these terms can be clarified only by solving (2.8) in various physically motivated situations. This will be addressed in future work.

In conclusion, the general formalism presented in this paper lays the basis for further study of wave-induced transport. Future applications of the theory will allow a full evaluation of the importance of this mechanism in various environments where waves exist or coexist with other types of fluid motion.

We thank L. Piterbarg, A. Balk, and R. McLaughlin for useful conversations. Part of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Appendix A. The evaluation of multiple time integrals entering the computation of $\boldsymbol{D}^{(2)}$

In this Appendix we reduce the multiple time integrals appearing in (5.19) and (5.24) to more convenient forms. Let $f(t)$ and $g(t)$ be functions that vanish sufficiently rapidly at infinity such that all of the integrals we consider below converge, and in particular that

$$\left.\begin{aligned} F(t) &\equiv \int_t^\infty \mathrm{d}s f(s), \quad \tilde{F}(t) \equiv \int_{-\infty}^{-t} \mathrm{d}s f(s) = F(-\infty) - F(-t), \\ G(t) &\equiv \int_t^\infty \mathrm{d}s\, g(s), \quad \tilde{G}(t) \equiv \int_{-\infty}^{-t} \mathrm{d}s\, g(s) = G(-\infty) - G(-t) \end{aligned}\right\} \tag{A 1}$$

are well defined. Clearly $F(t), \tilde{F}(t), G(t), \tilde{G}(t)$ vanish at positive infinity.

Consider then the integral

$$\begin{aligned} J_1(t) &\equiv \int_0^t \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 \int_0^{s_2} \mathrm{d}s_3 f(t-s_2) g(s_1-s_3) \\ &= \int_0^t \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 f(t-s_2)[G(s_1-s_2) - G(s_1)] \\ &= -\int_0^t \mathrm{d}s_1 [F(t-s_1) - F(t)] G(s_1) + \int_0^t \mathrm{d}\bar{s}_2 \int_{\bar{s}_2}^t \mathrm{d}s_1 f(t-s_1+\bar{s}_2) G(\bar{s}_2) \\ &= \int_0^t \mathrm{d}s [F(s) - F(t-s)] G(s). \end{aligned} \tag{A 2}$$

In the second term of the third line we have made the change of variable from s_2 to $\bar{s}_2 = s_1 - s_2$ and then interchanged the order of the s_1 and $\bar{s}_2$ integrations. In the last line we have noted that $\int_{\bar{s}_2}^t \mathrm{d}s_1 f(t-s_1+\bar{s}_2) = F(\bar{s}_2) - F(t)$.

Next consider

$$\begin{aligned} J_2(t) &\equiv \int_0^t \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 \int_0^{s_2} \mathrm{d}s_3 f(t-s_1) g(s_2-s_3) \\ &= \int_0^t \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 f(t-s_1)[G(0) - G(s_2)] \\ &= tF(0)G(0) + \int_0^t \mathrm{d}s [F(s)G(t-s) - F(s)G(0) - F(0)G(s)], \end{aligned} \tag{A 3}$$

in which, to obtain the last line, the order of the s_1 and s_2 integrations has been interchanged.

By similar manipulations one obtains

$$\begin{aligned}J_3(t) &\equiv \int_0^t ds_1 \int_0^t ds_2 \int_0^{s_2} ds_3 f(s_1 - s_2) g(t - s_3)\\&= \int_0^t ds_2 [F(-s_2) - F(t - s_2)][G(t - s_2) - G(t)]\\&= -tF(-\infty)G(t) + \int_0^t ds\{[F(s) + \tilde{F}(s)]G(t) + [F(-\infty) - F(s) - \tilde{F}(t - s)]G(s)\},\end{aligned} \tag{A 4}$$

$$\begin{aligned}J_4(t) &\equiv \int_0^t ds_1 \int_0^t ds_2 \int_0^{s_2} ds_3 f(t - s_1) g(s_2 - s_3)\\&= [F(0) - F(t)] \int_0^t ds_2 [G(0) - G(s_2)]\\&= [F(0) - F(t)] \left[tG(0) - \int_0^t ds G(s) \right],\end{aligned} \tag{A 5}$$

$$\begin{aligned}J_5(t) &\equiv \int_0^t ds_1 \int_0^t ds_2 \int_0^{s_1} ds_3 f(s_2 - s_3) g(t - s_1)\\&= \int_0^t ds_1 \int_0^t ds_2 [F(s_2 - s_1) - F(s_2)] g(t - s_1)\\&= -[G(0) - G(t)] \int_0^t ds F(s) + \int_0^t ds_2 \int_{s_2 - t}^{s_2} d\bar{s}_1 F(\bar{s}_1) g(t + \bar{s}_1 - s_2), \quad \bar{s}_1 \equiv s_2 - s_1\\&= tF(-\infty)G(0) - \int_0^t ds\{[F(s) + \tilde{F}(s)]G(0) + [F(-\infty) - F(s) - \tilde{F}(t - s)]G(s)\},\end{aligned} \tag{A 6}$$

$$\begin{aligned}J_6(t) &\equiv \int_0^t ds_1 \int_0^t ds_2 \int_0^{s_1} ds_3 f(s_1 - s_2) g(t - s_3)\\&= \int_0^t ds_1 [F(s_1 - t) - F(s_1)][G(t - s_1) - G(t)]\\&= -tF(-\infty)G(t) + \int_0^t ds\{[F(s) + \tilde{F}(s)]G(t) + [F(-\infty) - \tilde{F}(s) - F(t - s)]G(s)\},\end{aligned} \tag{A 7}$$

$$\begin{aligned}J_7(t) &\equiv \int_0^t ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 f(t - s_3) g(s_1 - s_2)\\&= \int_0^t ds_2 \int_0^{s_1} ds_2 [F(t - s_2) - F(t)] g(s_1 - s_2)\\&= -F(t) \int_0^t ds_1 [G(0) - G(s_1)] + \int_0^t ds_2 \int_{s_2}^t ds_1 F(t - s_2) g(s_1 - s_2)\\&= -tF(t)G(0) + \int_0^t ds\{F(t)G(s) + F(s)[G(0) - G(s)]\},\end{aligned} \tag{A 8}$$

$$
\begin{aligned}
J_8(t) &\equiv \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^{s_2} f(t-s_2)g(s_1-s_3) \\
&= \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 f(t-s_2)[G(s_1-s_2)-G(s_1)] \\
&= [F(t)-F(0)]\int_0^t \mathrm{d}sG(s) + \int_0^t \mathrm{d}\bar{s}_1 \int_0^{t-\bar{s}_1} \mathrm{d}s_2 f(t-s_2)G(\bar{s}_1) \\
&\quad + \int_0^t \mathrm{d}s_1' \int_{s_1'}^t \mathrm{d}s_2 f(t-s_2)G(-s_1') \\
&= tF(0)G(-\infty) + \int_0^t \mathrm{d}s\{F(s)[G(s)-G(-\infty)+\tilde{G}(t-s)] - F(0)[G(s)+\tilde{G}(s)]\},
\end{aligned} \tag{A 9}
$$

where $\bar{s}_1 = s_1 - s_2$ and $s_1' = -\bar{s}_1$, and finally,

$$
\begin{aligned}
J_9(t) &\equiv \int_0^t \mathrm{d}s_1 \int_0^t \mathrm{d}s_2 \int_0^t f(t-s_1)g(s_3-s_2) \\
&= [F(0)-F(t)]\int_0^t \mathrm{d}s_3[G(s_3-t)-G(s_3)] \\
&= t[F(0)-F(t)]G(-\infty) - \int_0^t \mathrm{d}s[F(0)-F(t)][G(s)+\tilde{G}(s)].
\end{aligned} \tag{A 10}
$$

For the diffusion tensor one requires the large-t limits:

$$
\left.
\begin{aligned}
J_1(\infty) &= \int_0^\infty \mathrm{d}sF(s)G(s), \\
J_2(\infty) &= tF(0)G(0) - \int_0^\infty \mathrm{d}s[F(0)G(s)+F(s)G(0)], \\
J_3(\infty) &= \int_0^\infty \mathrm{d}s[F(-\infty)-F(s)]G(s), \\
J_4(\infty) &= tF(0)G(0) - F(0)\int_0^\infty \mathrm{d}sG(s), \\
J_5(\infty) &= tF(-\infty)G(0) - \int_0^\infty \mathrm{d}s\{[F(-\infty)-F(s)]G(s) + [F(s)+\tilde{F}(s)]G(0)\}, \\
J_6(\infty) &= \int_0^\infty \mathrm{d}s[F(-\infty)-\tilde{F}(s)]G(s), \\
J_7(\infty) &= \int_0^\infty \mathrm{d}sF(s)[G(0)-G(s)], \\
J_8(\infty) &= tF(0)G(-\infty) + \int_0^\infty \mathrm{d}s\{F(s)[G(s)-G(-\infty)] - F(0)[G(s)+\tilde{G}(s)]\}, \\
J_9(\infty) &= tF(0)G(-\infty) - \int_0^\infty \mathrm{d}sF(0)[G(s)+\tilde{G}(s)].
\end{aligned}
\right\} \tag{A 11}
$$

The divergent (proportional to t) terms in J_2, J_4, J_5, J_8 and J_9 precisely cancel in the final expression for $\boldsymbol{D}^{(2)}$ in the case of waves. In the more general inhomogeneous case they represent slow variations of the transport parameters with time scale as the diffusion process averages over larger and larger spatial scales. The shorter the length scale of the inhomogeneities in the statistics of the wave field, the more significant these terms become.

A.1. *Conversion to Fourier space*

In this section we convert the time integrals appearing in (A 8) into frequency integrals. Let

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t f(t)\,\mathrm{e}^{\mathrm{i}\omega t}, \quad \hat{g}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\, g(t)\,\mathrm{e}^{\mathrm{i}\omega t} \tag{A 12}$$

be the Fourier transforms of $f(t)$ and $g(t)$. We will assume that $\hat{f}(\omega)$ and $\hat{g}(\omega)$ decay sufficiently rapidly at infinity that all integrals below are well defined. Note that $\hat{f}(0) = F(-\infty)$, where $F(t)$ and $G(t)$ are defined in (A 1).

For completeness, recall first the result

$$\begin{aligned} K_0 \equiv F(0) &= \int_0^{\infty} f(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \hat{f}(\omega) \int_0^{\infty} \mathrm{d}t\,\mathrm{e}^{-\mathrm{i}(\omega-\eta)t} \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi i} \frac{\hat{f}(\omega)}{\omega - i\eta} \\ &= -\mathrm{i}P \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \frac{\hat{f}(\omega)}{\omega} + \tfrac{1}{2}\hat{f}(0), \end{aligned} \tag{A 13}$$

in which $\eta \to 0^+$ is a convergence factor, and in the last line we have used the standard identity

$$\frac{1}{\omega - \mathrm{i}\eta} = P\frac{1}{\omega} + \mathrm{i}\pi\delta(\omega), \tag{A 14}$$

where P denotes principal value. If $\hat{f}$ is odd only the first term survives, while if $\hat{f}$ is even only the second term survives.

Consider next the less-trivial integral

$$\begin{aligned} K_1 &= \int_0^{\infty} \mathrm{d}t F(t) \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \hat{f}(\omega) \int_0^{\infty} \mathrm{d}t \int_t^{\infty} \mathrm{d}s\,\mathrm{e}^{-\mathrm{i}(\omega - i\eta)s} \\ &= -\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \frac{\hat{f}(\omega)}{(\omega - \mathrm{i}\eta)^2} \\ &= -\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \frac{\hat{f}'(\omega)}{\omega - \mathrm{i}\eta} \\ &= -P \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \frac{\hat{f}'(\omega)}{\omega} - \frac{\mathrm{i}}{2}\hat{f}'(0), \end{aligned} \tag{A 15}$$

where in the fourth line an integration by parts has been performed (prime denotes derivative with respect to argument), and in the last line we have used the standard identity (A 14) once more. Since $\hat{f}(\omega)$ may always be decomposed into a sum of an even and an odd function in ω we may consider these two cases separately. If $\hat{f}(\omega)$ is even, we note first that

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \frac{\hat{f}(0)}{(\omega - \mathrm{i}\eta)^2} = 0. \tag{A 16}$$

This may be verified by direct integration, or more formally by noting that the pole at $\omega = \mathrm{i}\eta$ has no residue. Thus, from the third line of (A 13),

$$K_1 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\hat{f}(0) - \hat{f}(\omega)}{\omega^2} \quad (\hat{f} \text{ even}), \tag{A 17}$$

where it is now safe to set $\eta = 0$ so long as $\hat{f}(\omega) - \hat{f}(0)$ vanishes faster than linearly (typically it will vanish as ω^2) at $\omega = 0$. On the other hand if $\hat{f}(\omega)$ is odd, $\hat{f}'(\omega)$ will be even, and the principal value term in the last line of (A 13) vanishes. We then obtain simply

$$K_1 = -\frac{\mathrm{i}}{2}\hat{f}'(0) \quad (\hat{f} \text{ odd}). \tag{A 18}$$

Consider next

$$\begin{aligned} K_2 &= \int_0^{\infty} ds\, F(s)\, G(s) \\ &= \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega_2}{2\pi} \frac{\mathrm{i}\hat{f}(\omega_1)\hat{g}(\omega_2)}{(\omega_1 - \mathrm{i}\eta_1)(\omega_2 - \mathrm{i}\eta_2)(\omega_1 + \omega_2 - \mathrm{i}\eta_1 - \mathrm{i}\eta_2)}, \end{aligned} \tag{A 19}$$

where $\eta_1, \eta_2 \to 0^+$ are convergence factors. Notice now that

$$\int \frac{d\omega_2}{2\pi} \frac{1}{(\omega_2 - \mathrm{i}\eta_2)(\omega_1 + \omega_2 - \mathrm{i}\eta_1 - \mathrm{i}\eta_2)} = 0 = \int \frac{d\omega_1}{2\pi} \frac{1}{(\omega_1 - \mathrm{i}\eta_1)(\omega_1 + \omega_2 - \mathrm{i}\eta_1 - \mathrm{i}\eta_2)}, \tag{A 20}$$

so that

$$\begin{aligned} K_2 &= \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega_2}{2\pi} \frac{\hat{f}(\omega_1) - \hat{f}(0)}{\omega_1} \frac{\hat{g}(\omega_2) - \hat{g}(0)}{\omega_2} \frac{\mathrm{i}}{\omega_1 + \omega_2 - \mathrm{i}\eta} \\ &= \mathrm{i}P \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega_2}{2\pi} \frac{[\hat{f}(\omega_1) - \hat{f}(0)][\hat{g}(\omega_2) - \hat{g}(0)]}{\omega_1\omega_2(\omega_1 + \omega_2)} \\ &\quad + \frac{1}{2} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{[\hat{f}(\omega) - \hat{f}(0)][\hat{g}(-\omega) - \hat{g}(0)]}{\omega^2} \end{aligned} \tag{A 21}$$

in which in the second equality it is now safe to take the limit $\eta_1 = \eta_2 = 0$ in the first two denominators, but $\eta = \eta_1 + \eta_2$ must still be maintained in the third, and it is easily checked that the integral remains absolutely convergent. In the last equality the standard identity (A 15) has again been used, and the principal value controls the only singularity at $\omega_1 = -\omega_2$. We may now consider the four possible combinations of even and odd $\hat{f}(\omega)$ and $\hat{g}(\omega)$. The denominator of the argument of the principal value integral in (A 21) is odd under simultaneous sign reversal of ω_1 and ω_2 This term will therefore vanish if the numerator is even, i.e. if $\hat{f}(\omega)$ and $\hat{g}(\omega)$ are either both odd or both even. Thus

$$K_2 = \int_{-\infty}^{\infty} \frac{d\omega}{4\pi} \frac{[\hat{f}(\omega) - \hat{f}(0)][\hat{g}(-\omega) - \hat{g}(0)]}{\omega^2} \quad (\hat{f}, \hat{g} \text{ both even or both odd}). \tag{A 22}$$

On the other hand the integrand of (A 22) will be odd if exactly one of $\hat{f}(\omega)$ and $\hat{g}(\omega)$ is even and the other is odd. Thus

$$K_2 = \mathrm{i}P \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega_2}{2\pi} \frac{[\hat{f}(\omega_1) - \hat{f}(0)][\hat{g}(\omega_2) - \hat{g}(0)]}{\omega_1\omega_2(\omega_1 + \omega_2)} \quad (\text{one of } \hat{f}, \hat{g} \text{ even, one odd}). \tag{A 23}$$

Obviously, in both (A 22) and (A 23), the zero frequency subtraction vanishes if the function is odd. In the expression (5.29) for the diffusion tensor it is seen that only the form (A 22) appears.

Appendix B. Angular averages for isotropic spectra

In this Appendix it is shown how to compute angular averages of polynomials in the components of $\boldsymbol{k}$ over the unit sphere in d-dimensions. This is accomplished by relating such averages to the usual Gaussian average for which Wick's theorem may be applied. The Gaussian average is

$$\langle k_{i_1} k_{i_2} \ldots k_{i_n} \rangle_G \equiv \frac{\displaystyle\int \mathrm{d}^d k\, \mathrm{e}^{-|\boldsymbol{k}|^2/2} k_{i_1} k_{i_2} \ldots k_{i_n}}{\displaystyle\int \mathrm{d}^d k\, \mathrm{e}^{-|\boldsymbol{k}|^2/2}} \tag{B 1}$$

Wick's theorem reduces this in the usual way to sums of products of pair averages, with the latter given by $\langle k_i k_j \rangle = \delta_{ij}$. On the other hand, the unit sphere average is defined as

$$\langle \hat{k}_{i_1} \hat{k}_{i_2} \ldots \hat{k}_{i_n} \rangle_S \equiv \frac{\displaystyle\int \mathrm{d}^{d-1} \Omega \hat{k}_{i_1} \hat{k}_{i_2} \ldots \hat{k}_{i_n}}{\displaystyle\int \mathrm{d}^{d-1} \Omega}, \tag{B 2}$$

in which $\hat{k}_i$ are the components of the unit vector $\hat{\boldsymbol{k}}$, and $\mathrm{d}^{d-1}\Omega$ is the angular volume element.

Now, the full d-dimensional volume element may be written in the form $\mathrm{d}^d k = k^{d-1}\,\mathrm{d}k \mathrm{d}^{d-1}\Omega$, and therefore

$$\langle k_{i_1} k_{i_2} \ldots k_{i_n} \rangle_G = \frac{\displaystyle\int_0^\infty \mathrm{d}k\, k^{n+d-1} \mathrm{e}^{-k^2/2}}{\displaystyle\int_0^\infty k^{d-1} \mathrm{e}^{-k^2/2}} \langle \hat{k}_{i_1} \hat{k}_{i_2} \ldots \hat{k}_{i_n} \rangle_S, \tag{B 3}$$

and one then obtains immediately

$$\langle \hat{k}_{i_1} \hat{k}_{i_2} \ldots \hat{k}_{i_n} \rangle_S = \frac{\Gamma(\frac{1}{2}\mathrm{d})}{2^{n/2} \Gamma((n+d)/2)}) \langle k_{i_1} k_{i_2} \ldots k_{i_n} \rangle_G. \tag{B 4}$$

This leads then, for example, to

$$\left. \begin{aligned} \langle \hat{k}_i \hat{k}_j \rangle_S &= \frac{1}{d} \delta_{ij} \\ \langle \hat{k}_i \hat{k}_j \hat{k}_l \hat{k}_m \rangle_S &= \frac{1}{d(d+2)} [\delta_{ij}\delta_{lm} + \delta_{il}\delta_{jm} + \delta_{im}\delta_{jl}], \end{aligned} \right\} \tag{B 5}$$

and so on.

REFERENCES

Balk, A. M. & McLaughlin, R. M. 1999 Passive scalar in a random wave field: the weak turbulence approach. *Phys. Lett.* A **256**, 299–306.

Batchelor, G. K. 1952 Diffusion in a field of homogeneous turbulence. II. The relative motion of particles. *Proc. Camb. Phil. Soc.* **48**, 345–363.

Chertkov, M. & G. Falkovich, G. 1996 Anomalous scaling exponents of a white-advected passive scalar. *Phys. Rev. Lett.* **76**, 2706–2709.

Chertkov, M., Falkovich, G. & Lebedev, V. 1996 Nonuniversality of the scaling exponents of a passive scalar convected by a random flow. *Phys. Rev. Lett.* **76**, 3707–3710.

Fabrikant, A. L. 1991 Resonant wave-vortex interaction. In *Nonlinear Dynamics of Structures* (ed. R. Z. Sagdeev, U. Frisch, F. Hussain, S. S. Moiseev & N. S. Erokhin), pp. 359–368. World Scientific.

Falkovich, G. 1992 Inverse cascade and wave condensate in mesoscale and atmospheric turbulence. *Phys. Rev. Lett.* **69**, 3173–3176.

Falkovich, G. & Medvedev, S. B. 1992 Kolmogorov-like spectrum for turbulence of intertia-gravity waves. *Europhys. Lett.* **19**, 279–284.

Ford, R. 1994 Gravity wave radiation from vortex trains in rotating shallow water. *J. Fluid Mech.* **281**, 81–118.

Fu, L. L. 1981 Observations and models of inertial waves in the deep ocean. *Rev. Geophys. Space Phys.* **19**, 141–170.

Gawedzki, H. & Kupiainen, A. 1995 Anomalous scaling of the passive scalar. *Phys. Rev. Lett.* **75**, 3834–3837.

Gill, A. E. 1982 *Atmosphere-Ocean Dynamics.* Academic.

Glazman, R. E. 1994 Surface gravity waves at equilibrium with a steady wind. *J. Geophys. Res.* **99** (C3), 5249–5262.

Glazman, R. E. 1996*a* Scale-dependent ocean wave turbulence. In *Stochastic Models in Geosystems* (ed. S. A. Molchanov & W. A. Wojczinski), pp. 97–114. Springer.

Glazman, R. E. 1996*b* Spectra of baroclinic inertia-gravity wave turbulence. *J. Phys. Oceanogr.* **26**, 1256–1265.

Glazman, R. E., Fabrikant, A. & Srokosz, M. A. 1996 Numerical analysis of the sea state bias for satellite altimetry. *J. Geophys. Res.* **101** (C2), 3789–3799.

Glazman, R. E. & Weichman, P. B. 1989 Statistical geometry of a small surface patch in a developed sea. *J. Geophys. Res.* **94** (C4), 4998–5010.

Herterich, K. & Hasselmann, K. 1982 The horizontal diffusion of tracers by surface waves. *J. Phys. Oceanogr.* **12**, 704–711 (referred to herein as HH).

Kraichnan, R. H. 1970 Diffusion by a random velocity field. *Phys. Fluids* **13**, 22–31.

Kubo, R. 1963 Stochastic Liouville equations. *J. Math. Phys.* **4**, 174–183.

LeBlond, P. H. & Mysak, L. A. 1978 *Waves in the Ocean.* Elsevier.

Monin, A. S. & Ozmidov, R. V. 1981 *Ocean Turbulence* (in Russian). Hydrometeoizdat, Leningrad.

Monin, A. S. & Yaglom, A. M. 1971 *Statistical Fluid Mechanics*, Vol. 1. MIT Press.

Phillips, O. M. 1977 *The Dynamics of the Upper Ocean.* Cambridge University Press.

Phillips, O. M. 1985 Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *J. Fluid Mech.* **156**, 505–531.

Piterbarg, L. I. 1997 Short-correlation approximation in models of turbulent diffusion. In *Stochastic Models in Geosystems* (ed. S. A. Molchanov & W. A. Wojczinski), pp. 313–352. Springer.

Richardson, L. F. 1926 Atmospheric diffusion shown on a distance-neighbour graph. *Proc. R. Soc. Lond.* A 110 **756**, 709–737.

Richardson, L. F. & Stommel, H. 1948 Note on eddy diffusion in the sea. *J. Met.* **5**, 238–240.

Stepanyants, Yu. A. & Fabrikant, A. L. 1989 Progagation of waves in hydrodynamic shear flows. *Sov. Phys. Usp.* **32**, 783–805.

Stommel, H. 1949 Horizontal diffusion due to oceanic turbulence. *J. Mar. Res.* **8**, 199–225.

Taylor, G. I. 1921 Diffusion by continuous movements. *Proc. Lond. Math. Soc.* **20A**, 196–211.

Troitskaya, Yu. A. & Fabrikant, A. L. 1989 Resonant amplification of internal gravity waves in a shear flow. *Radio Phys. Quant. Electron.* **32**, 1221–1227.

Weichman, P. B. & Glazman, R. E. 2000 Spatial variations of a passive tracer in a random wave field. Submitted to *J. Phys. Oceanogr.*

Zakharov, V. E. & Filonenko, N. N. 1966 The energy spectrum stochastic oscillation of a fluid's surface. *Dokl. Acad. Nauk SSSR* **170**, 1292–1295 (in Russian).

Zakharov, V. E. & Zaslavskii, M. M. 1982 The kinetic equation and Kolmogorov spectra in the weak turbulence theory of wind waves. *Izv. Atmosp. Ocean Phys.* **18**, 747–753 (English translation).