J. Fluid Mech. (1994), *vol.* 271, *pp.* 267–284

Horizontal wave propagation in the equatorial waveguide

By P. RIPA

Centro de Investigación Científica y de Educación Superior de Ensenada, Km. 107 Carr. Tijuana-Ensenada, (22860) Ensenada B.C., México

(Received 30 June 1993 and in revised form 30 November 1993)

The dispersive (i.e. non-Kelvin) linear wave field on the equatorial β-plane, in a single vertical mode, is fully described by a single potential φ. Long Rossby waves, which are weakly dispersive, are represented in this field. This description is free from the problem of the 'spurious solution' encountered when working with an evolution equation for the meridional velocity; addition of this unwanted solution represents a gauge transformation that leaves the physical fields unaltered.

The general solution of the ray equations is found, including trajectories, and the amplitudes and phase fields. This solution is asymptotically valid for either high or low frequencies. The ray paths are identical in both limits, but the phase field is not, reflecting the isotropy of Poincaré waves, in one case, and the zonal anisotropy of Rossby waves, in the other.

Two examples are studied by ray theory: meridional normal modes and wave radiation from a point source in the equator. In the first case, the exact dispersion relation is obtained. In the second one, northern and southern caustics bend towards the equator, meeting there at focal points. The full solution is the superposition of many leaves and has a structure that would be hard to find in a normal modes expansion.

1. Introduction

Over two centuries ago Laplace published the definitive model of the tides in a rotating planet (Laplace 1778), introducing what is now known as the Coriolis parameter f. Many researchers have found particular solutions to Laplace's model and recognized the peculiarities of the equator (namely, the line where $f = 0$), e.g. to be the axis of a waveguide. A definitive paper was the result of the doctoral thesis of Taroh Matsuno (1966), who found and classified all *normal modes*, for one- or two-layer models in the equatorial β-plane.

Matsuno (1966) worked with the equation satisfied by the meridional velocity component (say, $\mathscr{L}v = 0$) and showed that in addition to a set of Poincaré and Rossby waves (labelled by a positive integer n) and both Yanai waves ($n = 0$), there was a spurious solution, corresponding also to $n = 0$. Matsuno also proved that the set of equatorial modes was complete, adding to the former solutions the Kelvin wave, which is not a solution of $\mathscr{L}v = 0$.

The same year that Matsuno's paper appeared, Blandford (1966) published independently the normal modes solution and, furthermore, used ray theory (RT) for equatorial Poincaré and Rossby waves. Jacobs (1967) also used RT in the equatorial β-plane, solving initial value problems instead of the monochromatic solutions used by all other authors quoted here.

More recently, Schopf, Anderson & Smith (1981) and Kessler (1990) used the low-frequency limit of equatorial RT in order to study the fate of a family of rays that leave due west from an eastern boundary. These authors find a focusing point at the equator, where the energy is concentrated. Chang & Philander (1988) added the effect of a mean current to this problem; thus, in addition to the refraction produced by the variations of f, there is a local Doppler shift and, more important, another contribution to the refraction due to the height changes associated with the current.

Except for Wunsch (1977) and McCreary (1984), who considered the vertical propagation of equatorial signals, all of the authors mentioned above restricted their attention to horizontal ray-wave propagation. This will also be the framework of the present paper, leaving vertical effects aside.

It should be mentioned that most of the work on RT (e.g. the papers quoted here, except that of Jacobs 1967), is restricted to the rays' trajectories; phases and amplitudes are not analysed. This may have been an incorrect inheritance from the developments of RT for isotropic waves, for which ray paths and phase surfaces are orthogonal. On the contrary, I will show here that for anisotropic waves the structure of the phase field is rich in information which is unobtainable from the rays' paths. For instance, in the equatorial-waves case, the trajectories form a pattern which is *independent* of the frequency, aside from a scale factor, whereas the phase field is very different for high or low frequencies (e.g. east–west symmetric in the first case versus a strong asymmetry in the second one).

The scope of this paper is two fold: first, to propose a way to avoid the problem of 'contamination' by the spurious mode and second to present a more complete solution of the ray equations. The horizontal domain for this study is the unbounded β-plane, although occasional references to the problem with one or two zonal walls will also be made. The rest of this paper is thus organized as follows. In §2 the model equation and conservation laws are reviewed. A potential field φ is introduced in §3 as a way to avoid the spurious mode problem. Ray theory is developed in §4, and the general solution to the ray equations is found in §5. The theory is applied, in §6, to two different problems: the meridional normal modes and an equatorial point source. Concluding remarks are presented in §7 and some mathematical details are left for an Appendix.

2. Model equations

Consider infinitesimal deviations from an ocean at rest in the equatorial β-plane (Moore & Philander 1977). The density stratification is characterized by the Brunt–Väisälä – or static stability – vertical profile $N^2(z)$, which may be everywhere bounded and non-zero (continuous stratification) or equal to a collection of Dirac's delta functions (layered model). The linearized equations for the velocity components (u, v, w), the deviation of kinematic pressure p, and the isopycnals vertical displacement ζ are

$$\partial_t u = -\partial_x p + fv, \tag{2.1a}$$

$$\partial_t v = -\partial_y p - fu, \tag{2.1b}$$

$$0 = -\partial_z p - N^2 \zeta, \tag{2.1c}$$

$$\partial_t \zeta = w, \tag{2.1d}$$

$$\partial_z w = -\partial_x u - \partial_y v, \tag{2.1e}$$

where $f (= \beta y)$ is the Coriolis parameter. Note that ζ enters in the buoyancy term in (2.1 c) and the density conservation law (2.1 d).

Let us assume a common vertical structure for the pressure and the zonal and meridional velocity components, in the form

$$\partial_z[N^{-2}\partial_z(p,u,v)] = -c^{-2}(p,u,v), \tag{2.2}$$

where c^2 is the *separation constant*. This implies $\partial_z \zeta = p/c^2$ from (2.1 *c*) and thus ζ is eliminated as a prognostic variable. No further mention of the (common) vertical structure will be made, save for the value of c, which must be in the internal-waves range (order of metres per second) in order for the unbounded equatorial β-plane to be a good approximation of the spherical geometry.

The evolution equations then reduce to

$$\partial_t p = -c^2\partial_x u - c^2\partial_y v, \tag{2.3a}$$

$$\partial_t u = -\partial_x p + fv, \tag{2.3b}$$

$$\partial_t v = -\partial_y p - fu. \tag{2.3c}$$

This system has several conservation laws, which are the linearized version of truly nonlinear ones (Ripa 1982). The linearized potential vorticity perturbation, given by

$$\xi := \partial_x v - \partial_y u - fp/c^2, \tag{2.4a}$$

satisfies

$$\partial_t \xi + \beta v = 0, \tag{2.4b}$$

which is the linearized version of $\mathrm{d}q/\mathrm{d}t = 0$; energy and pseudomomentum, have (linearized) densities given by

$$E := \tfrac{1}{2}(u^2+v^2+p^2/c^2) \tag{2.5a}$$

and

$$P := up/c^2 - \xi^2/2\beta, \tag{2.5b}$$

and are conserved, in the sense of

$$\partial_t E + \nabla\cdot(p\boldsymbol{u}) = 0 \tag{2.6a}$$

and

$$\partial_t P + \partial_x(u^2 - v^2 + p^2/c^2)/2 + \partial_y(uv) = 0. \tag{2.6b}$$

Note that if the horizontal domain has rigid boundaries where the normal component of the velocity vanishes, then $\iint E$ is a constant of motion. In order for $\iint P$ to be an integral of motion, though, the rigid boundaries must be zonal.

3. Kelvin solution and the potential field

I will now find a particular solution, the Kelvin wave, and reduce the evolution problem of the non-Kelvin (i.e. dispersive) part of the wave field to the solution of a single equation. From (2.3 *a, b*) one easily finds

$$(\partial_t \pm c\,\partial_x)(p \pm cu) = -c^2\partial_y v \pm cfv, \tag{3.1}$$

which shows that $p \pm cu$ (and thus p and u) can almost be obtained from v, integrating along the characteristics $x = x_0 \pm ct$. 'Almost' means in the sense of up to the addition of the general solution of the homogeneous equations (right-hand side equal to zero) which, using (2.3 *c*), is found to be given by

$$\left.\begin{aligned} p &= K(t-sx)\exp(-\tfrac{1}{2}s\beta y^2), \\ u &= sp, \quad v \equiv 0, \end{aligned}\right\} \tag{3.2a}$$

where $K(\,)$ is arbitrary and $sc = \pm 1$; clearly the only acceptable solution (in the unbounded β-plane, $-\infty < y < \infty$) corresponds to

$$sc = 1, \tag{3.2b}$$

which is the *Kelvin wave*. (If one writes a 'dispersion' relation for this hyperbolic component, then s coincides with the slowness defined below.)

In order to describe the non-Kelvin part of the solution to (2.3), a potential $\varphi(\boldsymbol{x}, t)$ is introduced, such that

$$p = -c^2(\partial_{ty} + f\partial_x)\varphi, \tag{3.3a}$$

$$u = (c^2\partial_{xy} + f\partial_t)\varphi, \tag{3.3b}$$

$$v = (\partial_{tt} - c^2\partial_{xx})\varphi; \tag{3.3c}$$

these expressions are suggested by operating with $\partial_t \mp c\partial_x$ on (3.1) and then eliminating u or p (note that the streamfunction in the quasi-geostrophic limit is given by $-c^2\partial_x\varphi$). Given the physical fields (p, u, v), φ is undetermined to within a solution of the homogeneous equations, viz.

$$\varphi_S := A(x + ct)\exp(-\beta y^2/2c), \tag{3.4}$$

with arbitrary $A()$; i.e. $c^2(\partial_{ty} + f\partial_x)\varphi_S = (c^2\partial_{xy} + f\partial_t)\varphi_S = (\partial_{tt} - c^2\partial_{xx})\varphi_S = 0$. Thus $\varphi \to \varphi + \varphi_S$ is a *gauge* transformation, which leaves the dynamical fields unaltered. (Notice that φ_S is not an 'westward-Kelvin' solution, i.e. (3.4) is not equal to (3.2*a*) with $sc = -1$; in fact, the latter diverges away from the equator, whereas (3.4) does not.)

Substitution of (3.3) in (2.3*c*) gives the following evolution equation:

$$\mathscr{L}\varphi := (\partial_{tt} + f^2 - c^2\nabla^2)\partial_t\varphi + c^2\boldsymbol{\beta}\cdot\nabla\varphi = 0, \tag{3.5a}$$

where

$$\boldsymbol{\beta}(\boldsymbol{x}) := \hat{\boldsymbol{z}} \times \nabla f = (-\beta, 0, 0) \tag{3.5b}$$

is the β-vector field; a similar equation, without the triple time-derivative term, can be found for topographic Rossby waves (Ripa & Carrasco 1993).

Let me consider now the conservation laws associated with this evolution equation. *A priori*, they might not be the same as (2.4) and (2.6) because φ describes only the non-Kelvin part of the wave field. However, it will be shown at the end of this section that the potential vorticity, energy and pseudomomentum conservation laws are indeed satisfied exactly by the subsystem modelled by (3.5).

In the first place, $\partial_x\mathscr{L}\varphi = 0$ may be seen to be equal to the potential vorticity equation (2.4*b*), where

$$\xi := (\partial_{tt} + f^2 - c^2\nabla^2)\partial_x\varphi - \beta\partial_t\varphi. \tag{3.6}$$

Consider now the quadratic integrals of motion. The expression $\varphi\mathscr{L}\varphi = 0$ is equivalent to the following conservation law

$$\partial_t I + \nabla\cdot\boldsymbol{F} = 0: \begin{cases} I = (\partial_t\varphi)^2 - 2\varphi\,\partial_{tt}\varphi - f^2\varphi^2 - c^2(\nabla\varphi)^2 \\ \boldsymbol{F} = c^2(2\varphi\nabla\partial_t\varphi - \boldsymbol{\beta}\varphi^2). \end{cases} \tag{3.7}$$

(Notice the appearance of high time derivatives both in (3.5) and (3.7) on account of having transformed from three fields (p, u and v) into only one (φ).) The variable I corresponds to neither the energy density nor the pseudomomentum one, defined in (2.5). In fact, conservation laws (2.6) may be independently derived from (3.5) in the following way:

$$\left.\begin{aligned} v\mathscr{L}\varphi = 0 &\Rightarrow \partial_t E + \nabla\cdot\boldsymbol{F}_E = 0, \\ \xi\partial_x\mathscr{L}\varphi = 0 &\Rightarrow \partial_t P + \nabla\cdot\boldsymbol{F}_P = 0, \end{aligned}\right\} \tag{3.8}$$

where $v[\varphi]$ and $\xi[\varphi]$ are given by (3.3*c*) and (3.6), respectively. The expressions for $E[\varphi]$, $\boldsymbol{F}_E[\varphi]$, $P[\varphi]$ and $\boldsymbol{F}_P[\varphi]$ are complicated (e.g. see (A 3) and (A 5) in the Appendix); the fluxes are not necessarily those given in (2.6*a*, *b*), i.e. they may differ from them by a trivially non-divergent vector field.

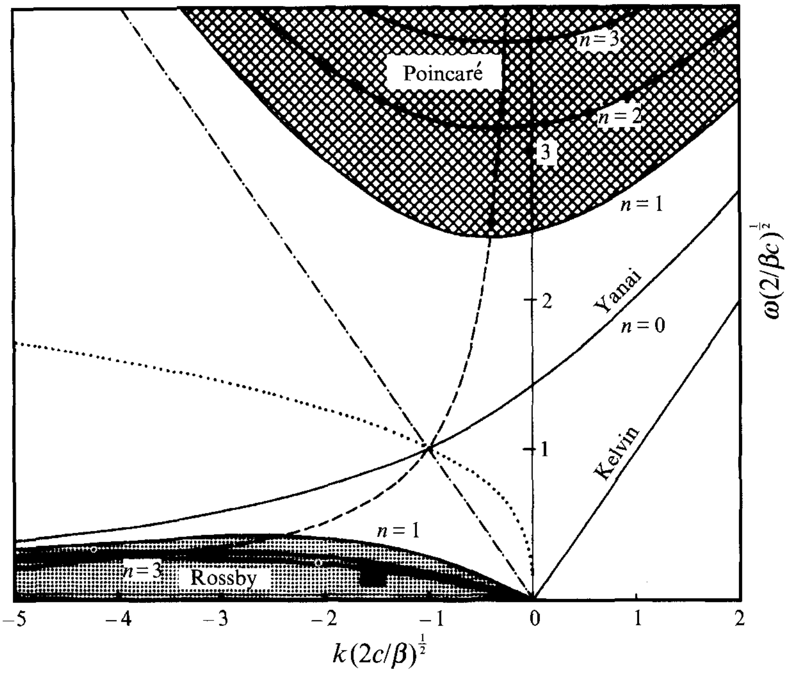

FIGURE 1. Dispersion relation between the frequency ω and the zonal wavenumber k of equatorial waves, with a vertical structure characterized by the *separation constant* c; β is the meridional gradient of the Coriolis parameter. For the *normal modes*, n is a non-negative integer which characterizes the meridional structure; the dot-dashed line represents a spurious $n = 0$ solution. On the other hand, for *ray theory*, n is a continuum variable which binds the variations of the meridional position y and wavenumber $l(y)$ by $l^2c^2+\beta^2y^2 = (2n+1)\beta c (= \omega^2-k^2c^2-\beta k/\omega)$. The dashed and dotted lines represent the zero and infinity values of the zonal group velocity $\partial\omega/\partial k$.

The important result is that the potential vorticity, energy and pseudomomentum conservation laws are exactly satisfied after the filtering of the Kelvin part of the wave field represented by (3.3); the reason for this is explained below. These three conservation laws correspond to similar ones in the fully nonlinear problem, whereas (3.7) is probably a product of the linearization. However, it will be shown that in the RT limit, the phase averages of I, E and P are related.

Normal modes

Using $\varphi = \psi_n((\beta/c)^{\frac{1}{2}}y)\exp[\mathrm{i}(kx-\omega t)]$, it is found that (3.5) is satisfied if the $\psi_n(\,)$ are Hermite functions (for $-\infty < y < \infty$, n is a non-negative integer), and the following dispersion relations are satisfied:

$$\omega(\omega-kc) = \beta c \quad (n=0), \tag{3.9a}$$

$$\frac{\omega^2}{c^2}-\beta\frac{k}{\omega} = k^2+(2n+1)\frac{\beta}{c} \quad (n = 1,2,\ldots). \tag{3.9b}$$

For each value of k, (3.9*a*) has two solutions, the Yanai waves, and (3.9*b*) has three, two Poincaré waves and one Rossby wave (see figure 1).

It is easier to write down the solutions of (3.9) in parametric or implicit form, using the *slowness*

$$s := k/\omega; \tag{3.10a}$$

namely

$$\left.\begin{array}{ll} n=0: & \omega^2 = \dfrac{\beta c}{1-sc}, \quad -\infty < sc < 1: \quad \text{Yanai} \\ n>0: & \omega^2 = \beta c \dfrac{2n+1+sc}{1-s^2c^2} \quad \begin{cases} -1 < sc < 1: & \text{Poincaré} \\ -\infty < sc \leqslant 2n+1: & \text{Rossby} \end{cases} \end{array}\right\} \quad (3.10b)$$

From these expressions it is natural to choose the point corresponding to $sc = -1$ as the boundary between the Poincaré-like and Rossby-like parts of the Yanai waves. The corresponding wavenumber and frequency define appropriate scales for the equatorial β-plane, in the form

$$\tilde{k} := (\beta/2c)^{\frac{1}{2}}, \quad \tilde{\omega} := (\tfrac{1}{2}\beta c)^{\frac{1}{2}}; \quad (3.11)$$

these scales are used in figure 1 (see the quadruple crossing at $k = -\tilde{k}$ and $\omega = \tilde{\omega}$). Notice that $\tilde{\omega}$ is both a Poincaré and a Rossby frequency scale, in the sense that $\tilde{\omega} = \tilde{k}c = \beta/2\tilde{k}$. The inverse of $\tilde{k}$ can be rightly called the equatorial *deformation radius*. (Other common definitions are a factor of $\surd 2$ larger or smaller than this one.)

The roots of (3.9*b*) for $n = 0$ are both solutions of (3.9*a*) and a third one, which is precisely φ_S; since in the present formalism addition of φ_S is but a gauge transformation, its presence is innocuous. On the other hand, the equation for v mentioned in the Introduction is but $\mathscr{L}v = 0$, where the operator $\mathscr{L}$ is defined in (3.5*a*). As a solution of this equation, $v \propto \varphi_S$ is spurious because it implies $p, u \sim \infty$ as $|y| \to \infty$. In a Galerkin expansion using the normal modes, this spurious solution is not a major problem, because it is simply not included in the series. However, with other methods (for instance, RT) it may not be possible to eliminate the spurious solution in v *a priori*. This is the advantage of working with the equation $\mathscr{L}\varphi = 0$, introduced here, rather than with $\mathscr{L}v = 0$.

Moore (1994) found independently the scalar representation of the linearized perturbations and addressed the interesting – and usually eluded – problem of the Coriolis force due to the horizontal component of Earth's rotation vector. This author works with the three-dimensional representation, which can be obtained using (2.2) in (3.3), i.e. replacing $c^2\varphi$ by ϕ and $\varphi(=c^{-2}\phi)$ by $-\partial_z(N^{-2}\partial_z\phi)$. Moore (1994) points out that the Kelvin mode is derivable from a certain potential $\varphi_{K+}(x-ct, y)$, which when introduced in (3.3) gives (3.2). However, this function satisfies $\partial_y^2\varphi_{K+} = (f^2/c^2 + \beta/c)\varphi_{K+}$ and therefore diverges exponentially in the *unbounded* β-plane, which is the domain used in this paper.

With one or two zonal walls (Philander 1977; Cane & Sarachik 1979) the dispersion relation of the normal modes is still given by (3.9*b*) but with the factor $(2n+1)$ replaced by a certain constant μ_n. The three roots of (3.9*b*) must be included for all $n \geqslant 0$, i.e. there is no longer a spurious solution (with $n = 0$). A very important result is that $\mu_n > 2n+1$, which implies $\omega^2 \neq k^2c^2$ for all dispersion curves. The structure of the normal modes is also modified because φ must vanish at a zonal coast, as implied by (3.3*c*) together with $\omega^2 \neq k^2c^2$. In sum, with one or two zonal walls, the 'spurious' solution becomes a physically acceptable one, and it is no longer equal to the gauge mode φ_S given by (3.4).

With two zonal walls, the westward Kelvin mode, (3.2*a*) with $sc = -1$, is also physically acceptable and can be derived from a certain potential $\varphi_{K-}(x+ct, y)$. Notice, however, that the potentials $\varphi_{K\pm}$ do not satisfy the evolution equation (3.5) but rather $(\partial_t \pm c\,\partial_x)\varphi_{K\pm} = 0$; consequently, Kelvin rays differ from those of the φ-field (developed in the following sections) because of the former have no meridional propagation.

The normal modes provide an orthogonal and complete basis for the expansion of the fields (p, u, v), in which the spurious solution of the unbounded case, $v \propto \varphi_S$, does not appear (Matsuno 1966). This expansion is useful to explain why the potential vorticity, energy and pseudomomentum conservation laws are exactly satisfied by the state space described by φ, which does not include the Kelvin mode. Firstly, the Kelvin mode has $\xi \equiv v \equiv 0$ and thus (2.4) is trivially satisfied. Secondly, both the energy and pseudomomentum have a diagonal representation in the normal modes basis, namely

$$E = \sum_a E_a \quad \text{and} \quad P = \sum_a s_a E_a, \tag{3.12}$$

where s_a is the slowness of mode a (Ripa 1982). Consequently, excluding the Kelvin mode does not break the laws of energy and pseudomomentum conservation. Note that the Kelvin mode has the maximum pseudomomentum per unit energy.

4. Ray theory

I will now develop the RT approximation to the evolution equation (3.5) and the conservation laws (3.7) and (3.8). Proposing the ansatz

$$\varphi(\boldsymbol{x}, t) = A(\boldsymbol{x}, t)\, \mathrm{e}^{\mathrm{i}\vartheta(\boldsymbol{x}, t)}, \tag{4.1}$$

the local wavenumber vector and the local frequency are defined by

$$\boldsymbol{k}(\boldsymbol{x}, t) := \boldsymbol{\nabla}\vartheta(\boldsymbol{x}, t), \quad \omega(\boldsymbol{x}, t) := -\partial_t \vartheta(\boldsymbol{x}, t). \tag{4.2}$$

Ray theory is based on the assumption that neither these fields nor the amplitude $A(\boldsymbol{x}, t)$ vary much across a wavelength or during one period; this will be called the *RT limit*, and it will be specified below. In this limit, the physics of the problem is reflected in that $\boldsymbol{k}$ and ω are not independent but, rather, are related through a dispersion relation $\omega = \Omega(\boldsymbol{k}, \boldsymbol{x})$, i.e.

$$\frac{\partial\vartheta}{\partial t} + \Omega\left(\frac{\partial\vartheta}{\partial \boldsymbol{x}}, \boldsymbol{x}\right) = 0. \tag{4.3}$$

Solutions of this highly nonlinear equation are found by the method of characteristics (Lighthill 1978), i.e. by solving the *ray equations*

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \frac{\partial\Omega}{\partial \boldsymbol{k}}, \quad \frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t} = -\frac{\partial\Omega}{\partial \boldsymbol{x}}. \tag{4.4}$$

Finally, from (4.4) and (4.2) we obtain an equation for the phase of the form

$$\frac{\mathrm{d}\vartheta}{\mathrm{d}t} = \boldsymbol{k}\cdot\boldsymbol{C} - \omega, \tag{4.5}$$

where $\boldsymbol{C} := \partial\Omega/\partial\boldsymbol{k}$ is the group velocity. This equation is neglected by most authors, i.e. by those who limit themselves to the evaluation of ray trajectories.

Substitution of the ansatz (4.1) into the evolution equation (3.5) gives two coupled nonlinear equations for A and ϑ, namely the real and imaginary parts of equation (A 2) in the Appendix. Multiplying (A 2) by $-A$, the real part gives

$$\partial_t\langle I\rangle + \boldsymbol{\nabla}\cdot\langle \boldsymbol{F}\rangle = 0, \tag{4.6}$$

where $\langle\ \rangle$ denotes an average in one cycle (say, one period or one wavelength), and $\langle I\rangle$ and $\langle \boldsymbol{F}\rangle$ are fully given by (A 1). It is interesting to notice that, so far, no approximation has been made: the conservation law (3.7) is exactly satisfied after the averaging.

In the RT limit described above, (at least) the right-hand side of (A 2) is neglected; the imaginary part then gives the dispersion relation

$$\frac{f^2-\omega^2}{c^2}+\beta\frac{k}{\omega}+k^2+l^2=0, \tag{4.7}$$

suitable for the solution of the ray equations (4.4) and (4.5). Consistent with this limit, the expressions in (A 1) for (4.6) are replaced by

$$\langle I\rangle=\tfrac{1}{2}(2\omega^2+\beta c^2k/\omega)A^2, \tag{4.8a}$$

$$\langle \boldsymbol{F}\rangle=\tfrac{1}{2}(2\boldsymbol{k}\omega-\boldsymbol{\beta})A^2\equiv\langle I\rangle \boldsymbol{C}, \tag{4.8b}$$

where

$$\boldsymbol{C}=\frac{2\boldsymbol{k}\omega-\boldsymbol{\beta}}{2\omega^2/c^2+\beta k/\omega} \tag{4.9}$$

is the group velocity corresponding to the dispersion relation (4.7). Equations (4.8), and (A 4) and (A 6) in the Appendix show that

$$\langle E\rangle=\tfrac{1}{2}(\omega^2-k^2c^2)\langle I\rangle, \tag{4.10a}$$

$$\langle P\rangle=\frac{k}{\omega}\langle E\rangle. \tag{4.10b}$$

Notice that for a ray, $\langle P\rangle/\langle E\rangle$ is equal to the slowness k/ω, i.e. the same relation found for the normal modes in (3.11). Since neither x nor t appears explicitly in the dispersion relation (4.7), it follows that $(\partial_t+\boldsymbol{C}\cdot\boldsymbol{\nabla})\boldsymbol{k}=0$ and $(\partial_t+\boldsymbol{C}\cdot\boldsymbol{\nabla})\omega=0$. Consequently, in this limit (4.6) is equivalent to the energy and pseudomomentum conservation laws

$$\partial_t\langle E\rangle+\boldsymbol{\nabla}\cdot(\langle E\rangle \boldsymbol{C})=0, \tag{4.11a}$$

$$\partial_t\langle P\rangle+\boldsymbol{\nabla}\cdot(\langle P\rangle \boldsymbol{C})=0. \tag{4.11b}$$

The amplitude A is calculated in the following form. Equations (4.6) or (4.11) are equivalent to $\mathrm{d}A^2/\mathrm{d}t+A^2\boldsymbol{\nabla}\cdot\boldsymbol{C}=0$. Now, assume that a solution of (4.4) and (4.5) is written in the implicit form $\boldsymbol{x}=\boldsymbol{x}(a,b,t)$, $\boldsymbol{k}=\boldsymbol{k}(a,b,t)$, $\vartheta=\vartheta(a,b,t)$, where (a,b) is a pair of parameters that define each point in the ray. From the chain rule and (4.4) it follows that the Jacobian

$$J=\frac{\partial(x,y)}{\partial(a,b)}, \tag{4.12}$$

satisfies $\mathrm{d}J/\mathrm{d}t=J\boldsymbol{\nabla}\cdot\boldsymbol{C}$, and therefore the amplitude is simply of the form

$$A(\boldsymbol{x},t)=A_0(a,b)(J_0/J)^{\frac{1}{2}}, \tag{4.13}$$

where J_0 is the value of J at some reference point in the ray path, where the amplitude is set to be equal to A_0. At a caustic the Jacobian J goes through a zero, and a $\frac{1}{2}\pi$ phase jump must be added to ϑ, on account of the square root in (4.13).

The relationship between the amplitude of φ and the energy and pseudomomentum densities, given by (4.8) and (4.10), shows that: (i) the amplitude A diverges as $\omega^2\to k^2c^2$, i.e. when approaching the dispersion relation curves of the Kelvin mode and the gauge mode φ_S (see the two straight lines in figure 1); (ii) the amplitude A vanishes as $\omega^3\to\frac{1}{2}\beta c^2k$, i.e. when approaching the infinite-group-velocity line (see (4.9) and the dotted curve in figure 1).

5. Solution of the ray equations

The general solution to the ray equations, (4.4) and (4.5), for the dispersion relation (4.7) is presented next. At the end of the section, the corresponding RT limits are discussed and are shown to be related to the parameter γ, given by

$$\gamma^2 = \frac{\omega^2}{\beta c} + \frac{\beta c}{4\omega^2} \equiv \frac{1}{2}\left(\frac{\omega^2}{\tilde{\omega}^2} + \frac{\tilde{\omega}^2}{\omega^2}\right); \tag{5.1}$$

consequently $\gamma \gg 1$ means either $\omega \gg \tilde{\omega}$ (the *Poincaré limit*) or $\omega \ll \tilde{\omega}$ (the *Rossby limit*); a sufficient condition for $\gamma \gg 1$ is $f^2 + l^2c^2 \gg \beta c$. In order to present the results, it is convenient to define a *proper time*

$$\mu := (t - t_0)/\tau, \tag{5.2a}$$

where

$$\tau := \omega/\beta c + kc/2\omega^2. \tag{5.2b}$$

Any ray trajectory, a solution of (4.4) for the dispersion relation (4.7), can be written in the form

$$[x - x_0, y] = \gamma(c/\beta)^{\frac{1}{2}}[\mu\cos\alpha, \sin\mu\sin\alpha], \tag{5.3a}$$

where α is a parameter of the ray, equal to the slope of the equatorial crossings (which happen at $x = x_0 \operatorname{mod}(\gamma(c/\beta)^{\frac{1}{2}}\pi\cos\alpha)$ and $t = t_0 \operatorname{mod}(\pi\tau)$). The corresponding wavenumber is given by

$$[k + \beta/2\omega, l] = \gamma(\beta/c)^{\frac{1}{2}}[\cos\alpha, \cos\mu\sin\alpha]. \tag{5.3b}$$

Finally, the phase, obtained by integrating (4.5) along the ray, is

$$\vartheta = \vartheta_0 + \tfrac{1}{2}\gamma^2\sin^2\alpha(\mu + \sin\mu\cos\mu) + k(x - x_0) - \omega t, \tag{5.3c}$$

where $\vartheta_0 - \omega t_0$ is the phase at the equatorial crossing.

Each ray moves along a sinusoid, between the turning latitudes $\pm\gamma(c/\beta)^{\frac{1}{2}}\sin\alpha(\mu = \frac{1}{2}\pi \operatorname{mod}(\pi))$, and conserves its values of ω, k and $\partial\omega/\partial k$. The approximation of ray theory might break down for $|\sin\alpha| \ll 1/\gamma$, namely for rays so shallow that the turning latitude is a fraction of the deformation radius. The second term of the expression for ϑ equals the path integral $\int l(y)\,\mathrm{d}y$.

Poincaré and Rossby limits

In the Poincaré limit, $\gamma^2 \sim \omega^2/\beta c \to \infty$, the dispersion relation (4.7) is replaced by

$$\omega^2 = f^2 + (k^2 + l^2)c^2, \tag{5.4}$$

and the solution, (5.3*a–c*), to the ray equations is approximated by

$$[x - x_0, y] = (\omega/\beta)[\mu\cos\alpha, \sin\mu\sin\alpha], \tag{5.5a}$$

$$[k, l] = (\omega/c)[\cos\alpha, \cos\mu\sin\alpha], \tag{5.5b}$$

$$\vartheta = \vartheta_0 + (\omega^2/4\beta c)[\sin 2\mu\sin^2\alpha + 2\mu(1 + \cos^2\alpha)] - \omega t, \tag{5.5c}$$

where $\mu = (t - t_0)\beta c/\omega$, i.e. $\tau = \omega/\beta c$.

On the other hand, in the Rossby limit, $\gamma^2 \sim \beta c/4\omega^2 \to \infty$, the dispersion relation is

$$\omega = \frac{-\beta k}{k^2 + l^2 + f^2/c^2} \tag{5.6}$$

instead of (4.7), and the solution to the ray equations is

$$[x-x_0, y] = -(c/2\omega)[\mu\cos\alpha, \sin\mu\sin\alpha], \tag{5.7a}$$

$$[k, l] = -(\beta/2\omega)[1+\cos\alpha, \cos\mu\sin\alpha], \tag{5.7b}$$

$$\vartheta = \vartheta_0 + (\beta c/16\omega^2)[\sin 2\mu \sin^2\alpha + 2\mu(1+\cos\alpha)^2] - \omega t, \tag{5.7c}$$

where $\mu = (t-t_0)2\omega^2/kc$, i.e. $\tau = kc/2\omega^2 \equiv -(1+\cos\alpha)\beta c/4\omega^3$. Notice that for $\omega > 0$, μ decreases with time ($\tau < 0$), and consequently γ was chosen to be negative.

Construction of the solution

These expressions for $\boldsymbol{x}$, $\boldsymbol{k}$ and ϑ are sufficient to obtain the RT approximation (4.1) for the solution of the evolution equation (3.5). This is done by specifying the phase ϑ along some closed curve Γ, as a function of time. For instance, Schopf *et al.* (1981) choose $\vartheta = -\omega t$ at an eastern meridional coast, $x = L$, whereas Kessler (1990) did likewise for an 'inclined' coast, $x = L - \alpha y$. The definitions (4.2) are then used to calculate the wavenumber parallel to Γ and the instantaneous frequency (which need not be constant); the other component of wavenumber is calculated using the dispersion relation. Thus at each point of Γ there is, at all times, a well-defined value of $\boldsymbol{k}$ and of ϑ with which it is possible to integrate (4.4) and (4.5) along the trajectory. Notice that the ray paths may be time dependent, namely when ω is not uniform.

Each ray then has two parameters (for instance, a label in Γ and the value of t when it left that curve), which can be related to x_0 and t_0 of the general solution presented above. Choosing the amplitude at Γ, say A_0, then (4.12) and (4.13) are used to calculate A along the path, and the physical fields are obtained from the RT limit of (3.3), which gives

$$p = c^2(l\omega + \mathrm{i}fk)A\,\mathrm{e}^{\mathrm{i}\vartheta}, \tag{5.8a}$$

$$u = -(c^2kl + \mathrm{i}f\omega)A\,\mathrm{e}^{\mathrm{i}\vartheta}, \tag{5.8b}$$

$$v = (\omega^2 - k^2c^2)A\,\mathrm{e}^{\mathrm{i}\vartheta}. \tag{5.8c}$$

Recall that, in general, $\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{x}, t)$ and $\omega = \omega(\boldsymbol{x}, t)$.

This is the right place to assess the degree of accuracy of the solution, in both RT limits. Assume that the frequencies excited at Γ are such that $\gamma = O(\varepsilon^{-1})$ with $\varepsilon \to 0$. The phase field is formally of the form

$$\vartheta(\boldsymbol{x}, t) = \varepsilon^{-2}\Theta(\varepsilon\boldsymbol{x}, \varepsilon^{2\mp1}t), \tag{5.9}$$

where the upper (lower) sign corresponds to the Poincaré (Rossby) limit. Therefore the accuracy is $\boldsymbol{x} = O(\varepsilon^{-1})$, $\boldsymbol{\nabla} = O(\varepsilon)$, $\boldsymbol{k} = O(\varepsilon^{-1})$, $t = O(\varepsilon^{-2\pm1})$, $\partial_t = O(\varepsilon^{2\mp1})$, and $\omega = O(\varepsilon^{\mp1})$, where ε is (for instance) the ratio of the equatorial deformation radius to the horizontal lengthscale (Jacobs used the Earth's radius instead of the latter).

Let me analyse first the Poincaré limit. Using the above estimates in the imaginary part of (A 2), it is found that the three terms in the dispersion relation (5.4) are of the same order (recall that $f^2 = O(\varepsilon^{-2})$); relative to them, the β-term in (4.7) is $O(\varepsilon^2)$, and the leading term neglected in the right-hand side is $O(\varepsilon^4)$. Exactly the same scaling is obtained for the β-term and the leading neglected term in the amplitude equation (4.6). The β-term can then be included, to improve the accuracy from $O(\varepsilon^2)$ to $O(\varepsilon^4)$; this term is responsible for the shift of the frequency minima to the left of the $k = 0$ axis in figure 1.

On the Rossby limit, on the other hand, the three terms in the dispersion relation (5.6) are of the same order and, relative to them, the ω^2-term in (4.7) is $O(\varepsilon^4)$. This is the same order as that of the leading term neglected in the right-hand side of (4.7). An identical result is found for the amplitude equation. Therefore the Rossby-limit solution, (5.6) and (5.7), is already accurate to $O(\varepsilon^4)$, and use of the general solution, (5.3 *a–c*), does not represent a consistent improvement (unlike the case of the Poincaré limit).

In sum, the general solution, (5.1)–(5.3 *c*), provides an $O(\varepsilon^4)$-accurate solution in both limits, $\omega = O(\varepsilon^{\mp 1})$, but for low frequencies, (5.6) and (5.7) has the same degree of accuracy.

6. Applications

In the examples presented here, ω is chosen to be constant; μ may be taken as a parameter along the ray path, unrelated to time. The phase field in (5.3 *c*) has the simple form

$$\vartheta(\boldsymbol{x}, t) = \gamma^2 S((\beta/c)^{\frac{1}{2}}\boldsymbol{x}/\gamma) - \omega t, \tag{6.1}$$

where S and its arguments are order unity. In this particular case, both the wavenumber and amplitude fields, $\boldsymbol{k}(\boldsymbol{x})$ and $A(\boldsymbol{x})$, are time independent.

Normal modes

A quantitative relation can be found between the refraction experienced by the rays, which prevents them from getting away from the equatorial zone, and the existence of normal modes. The dispersion relation obtained is exact, and amplitudes and phases correspond to the WKBJ approximation within the turning latitudes.

Consider a set of rays originated along the equator (the curve Γ mentioned above) with phase $kx - \omega t$, with constant ω and k. This corresponds to a fixed value of α and of t_0 in (5.3 *a–c*), and variable x_0, with $\vartheta_0 = kx_0$. The Jacobian (4.12) equals

$$J = \frac{\partial(x, y)}{\partial(x_0, \mu)} \propto \cos\mu \sin\alpha. \tag{6.2}$$

It is apparent in figure 2 that the solution is the superposition of many parts: one made of the rays going from the southern caustic to the northern one (solid lines), the return version (dashed lines), and so on. But, for a given value of ω not all angles α are possible; there must be a perfect phase matching between the 'original' rays and those that have made two turns, one at each caustic, otherwise incoherent superposition destroys the signal.

Consequently, the solution is composed of only two different parts, the sum of which results in a standing pattern in the meridional direction; the only apparent phase propagation is along the zonal coordinate. Each part is of the form (4.1) with $A = 1/(\cos\mu)^{\frac{1}{2}}$ and $\vartheta = \frac{1}{2}\gamma^2 \sin^2\alpha(\mu + \sin\mu\cos\mu) + kx - \omega t$. The amplitude diverges at the caustics, which correspond to $\mu = \frac{1}{2}(2m+1)\pi$ with integer m; this divergence is a complication of ray theory which can be resolved (see for instance Ludwig 1966; Peregrine & Smith 1979; Klauder 1988). Because of the square root, the phase jumps by $-\frac{1}{2}\pi$ at each caustic.

Consider then any ray that after leaving the equator at $(x_0, y_0 = 0, t_0)$ returns to that line, with the same group velocity, at $(x_1 = x_0 + 2\pi\gamma(c/\beta)^{\frac{1}{2}}\cos\alpha,\ y_1 = 0, t_1)$, i.e. after turning in each of both caustics. (See for instance, the path that goes from one extreme to the other in figure 2.) Since μ varies from 0 to 2π, the total phase change experienced by that ray equals $[\gamma^2\pi\sin^2\alpha - \frac{1}{2}\pi - \frac{1}{2}\pi] + k(x_1 - x_0) - \omega(t_1 - t_0)$. But since the phase of φ is imposed to be $kx - \omega t$ along the equator, the expression between square brackets must be equal to $2n\pi$, with n integer, in order to avoid destructive interference. Therefore, $\gamma^2\sin^2\alpha$ equals $2n+1$, which implies

$$l^2c^2 + f^2 = (2n+1)\beta c \quad (n = 0, 1, \ldots), \tag{6.3}$$

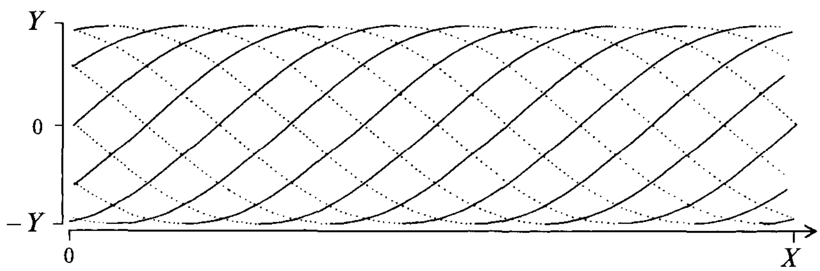

FIGURE 2. Set of rays that make up a normal mode. The equator, $y=0$, is halfway between the two caustics (the northern and southern envelope of the rays) situated at $y=\pm\gamma(c/\beta)^{\frac{1}{2}}\sin\alpha\,(=\pm Y)$, where $\gamma^2=\omega^2/\beta c+\beta c/4\omega^2$ and α is the slope of the equatorial crossings. The solid lines show the rays that go from the southern caustic to the northern one, whereas dashed lines show the return version; the net displacement of each ray is $X=\gamma(c/\beta)^{\frac{1}{2}}\cos\alpha$. Tuning the parameters so that there is no destructive interference after making several turns, results in $\gamma\sin\alpha=(2n+1)^{\frac{1}{2}}$ which the exact dispersion relation of equatorial modes.

see (5.3*a*, *b*); on replacing (6.3) in (4.7) one gets *exactly* the dispersion relation of the equatorial trapped modes (3.9), as promised. (The periodicity of the rays' trajectory should not be confused with that of the normal mode, viz. neither x_1-x_0 is equal to the wavelength nor t_1-t_0 is equal to the period of the wave.) The ray path equation (5.3*a*) shows that vanishing zonal group velocity corresponds to $\cos\alpha=0$, i.e. to $\gamma^2=2n+1$; Poincaré and Rossby normal modes are restricted to $\gamma^2\geqslant 2n+1$, while Yanai modes reach the minimum value $\gamma^2=1$ at $2\omega^2/\beta c=1$ ($sc=-1$), i.e. at the values defined in (3.11).

Adding the contribution of both families of rays it is found that

$$\varphi\sim\frac{1}{|\cos\mu|^{\frac{1}{2}}}\cos\left(\tfrac{1}{2}(2n+1)(\mu+\sin\mu\cos\mu)-\tfrac{1}{4}\pi\right)\mathrm{e}^{\mathrm{i}(kx-\omega t)},\tag{6.4a}$$

$$y=((2n+1)\,c/\beta)^{\frac{1}{2}}\sin\mu,\tag{6.4b}$$

which is the WKBJ approximation to the Hermite function $\psi_n((\beta/c)^{\frac{1}{2}}y)$ (Ripa 1983).

A similar argument in the Poincaré (high-frequency) limit, i.e. using (5.5) instead of (5.3*c*) for ϑ, yields $(\omega^2/\beta c)\sin^2\alpha=2n+1$. This represents the normal mode dispersion relation $\omega^2=k^2c^2+(2n+1)\beta c$, which is a good approximation to two roots of the exact one, (3.9*b*), corresponding to making $|sc|\ll 2n+1$ in (3.10*b*).

On the other hand, in the Rossby (low-frequency) limit, i.e. using (5.7) instead of (5.3*c*) for ϑ, $(\beta c/4\omega^2)\sin^2\alpha=2n+1$ results. This, in turn, implies the normal mode dispersion relation $\omega=-\beta k/[k^2+(2n+1)\beta/c]$, which is a very good approximation to the remaining root of the exact one, (3.9*b*), which corresponds to making $|sc|\gg 1$ in (3.10*b*).

These two approximations break down for $n=0$ or $\gamma\sin\alpha=1$, at $\omega^2\approx\frac{1}{2}\beta c$, which implies $\gamma\approx 1$ and $sc\approx -1$, i.e. the region where the Yanai wave makes the transition from the Poincaré to the Rossby family. From the point of view of ray theory, this corresponds to $\sin\alpha\approx 1$, i.e. rays that propagate meridionally.

Finally, the normal modes for the system with one or two walls, the problem studied by Philander (1977) and Cane & Sarachik (1979), can also be estimated using RT. The procedure is the same as developed above, except that at the reflection on a rigid wall the phase does not change by $\frac{1}{2}\pi$, as in a caustic, but by π (Ripa & Carrasco 1993). If at least one of the walls is within the turning latitudes corresponding to the frequency of the mode, i.e. there is reflection on a rigid coast, then the eigenvalue $\gamma\sin\alpha\,(=\mu_n^{\frac{1}{2}}$ in

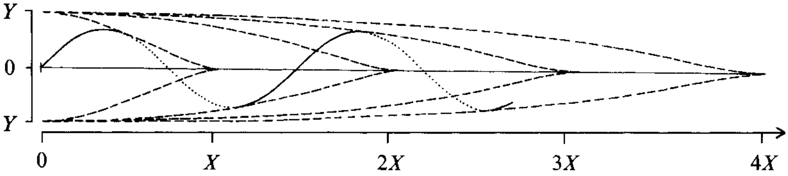

FIGURE 3. First four caustics (dashed lines) for a point source located at the far left of the figure in the equator (solid line). Only the eastern part is shown, because the western part is symmetric. The ray that departs with a 45° azimuth is also drawn. In this and the following figures, extreme latitudes (north and south of the source) are at $y = \pm\gamma(c/\beta)^{\frac{1}{2}}(=\pm Y)$. In the Poincaré limit (high-frequency) the meridional bounds are $\pm\omega/\beta$; in the Rossby limit (low-frequency) they are $\pm c/2\omega$. The caustics have cusps in the equator, at a distance $m\pi Y(=mX)$ from the source (with integer m), which are nearly focusing points, where the energy is concentrated.

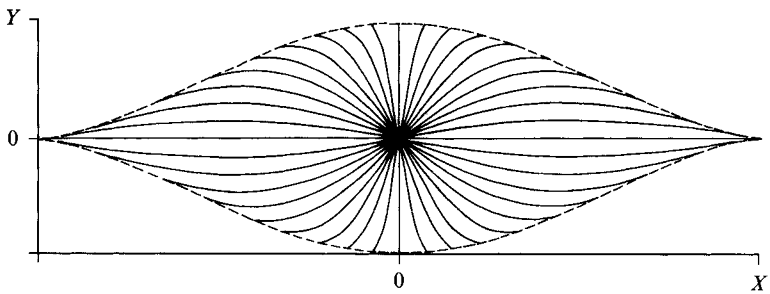

FIGURE 4. Trajectories of 40 rays that go from the source, where they leave 9° apart, to the first caustic. Zonal and meridional scales (X and Y) are defined in the previous figure. The energy density decreases, and so does the amplitude, where the rays diverge. There is both east–west and north–south symmetry.

§3) is the solution of a transcendental equation; this question will not be pursued any further here.

Point source

Consider now a point source somewhere in the equator; this point is a limiting case of the curve Γ mentioned above. The ray trajectories are given by (5.3*a*) with x_0 fixed (say, $x_0 = 0$), $\vartheta_0 = 0$, and α varying from 0 to 2π. Near the origin the rays are radial and uniformly distributed in the azimuth α. The Jacobian (4.12), equals

$$J = \frac{\partial(x,y)}{\partial(\alpha,\mu)} \propto \cos^2\alpha\sin\mu + \sin^2\alpha\mu\cos\mu, \tag{6.5}$$

and vanishes at the caustics, given by

$$\left.\begin{aligned} x &= \pm\gamma(\beta/c)^{\frac{1}{2}}\mu/(1-\tan\mu/\mu)^{\frac{1}{2}} \\ Y &= \pm\gamma(\beta/c)^{\frac{1}{2}}\sin\mu/(1-\mu/\tan\mu)^{\frac{1}{2}} \end{aligned}\right\} \quad (m-\tfrac{1}{2})\pi \leqslant \mu \leqslant m\pi, \tag{6.6}$$

where $m = 1, 2, \ldots,$ and both $\pm$ signs make up for the four quadrants; the eastern half of the first four caustics is shown in figure 3. The caustic does not correspond to the turning latitude of the rays, unlike in the normal modes case.

The caustics have cusps in the equator at $x = m\pi\gamma(\beta/c)^{\frac{1}{2}}$, with m the non-zero integers ($\mu = m\pi$), similar to the focusing point found by Schopf *et al.* (1981); Jacobs (1967), who dealt with the initial value problem, found foci in space–time. Figure 4 shows 40 rays going from the origin, where they leave 9° apart, to the first caustic. The paths are symmetric under reflections in x and y; on the other hand, the phase will be

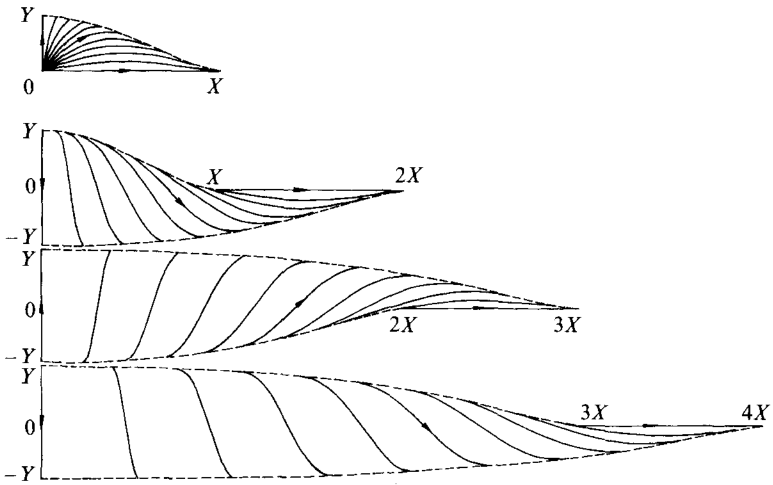

FIGURE 5. Trajectories of the 11 rays of figure 4 that start in the first quadrant (top) and then go from one caustic to the next one (below). Other parts are obtained by east–west and north–south reflection of these drawings. Notice that higher contributions have small amplitude except near the focus of the target caustic.

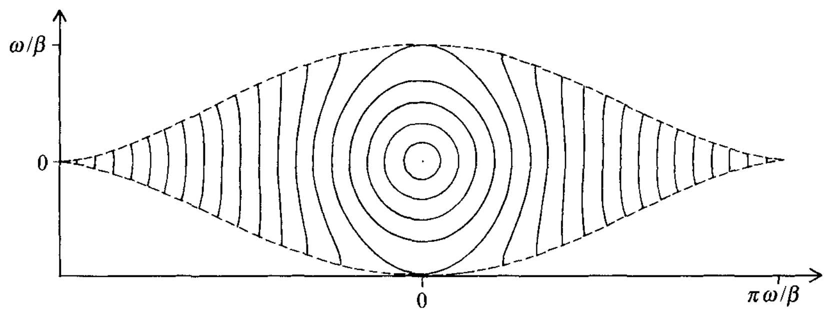

FIGURE 6. Phase function, $S(\beta x/\omega, \beta y/\omega)$, in the Poincaré limit and for the rays that travel from the source to the first caustic. The complete phase is $\vartheta = (\omega^2/\beta c)\,S - \omega t$. S increases away from the origin and the contour interval is $\Delta S = \pi/20$. There is both east–west and north–south symmetry. Phase propagation is first radial and later zonal; far from the source it is $\vartheta \approx \omega(|x|/c - t)$.

shown to possess only meridional symmetry. The full solution is the superposition of the contribution of the rays that go from the origin to the first caustic, those that go from the first to second, and so on. Trajectories are shown in figure 5 for the 11 rays of figure 4 found originally in the first quadrant (the other three sets of rays are omitted for simplicity and can be imaged by east–west and north–south reflections). Notice that the rays are tangent to the caustic, at the point of touching. The rays fill the band between the two extreme turning latitudes $y = \pm\gamma(\beta/c)^{\frac{1}{2}}$, and diverge with time, with a resulting decrease in amplitude and energy density, except in the neighbourhood of the foci (where the caustics touch the equator).

The phase field has a marked frequency dependence; next are presented the cases corresponding to the Poincaré (high-frequency) and Rossby (low-frequency) limits.

In the Poincaré limit, (5.4) and (5.5), the successive caustics are still given by (6.6), replacing γ by $\omega/(\beta c)^{\frac{1}{2}}$. Contours of the function S, defined in (6.1), are presented in figure 6 for the rays that go from the source to the first caustic, and in figure 7 for the rays that travel from the first to the second caustic. In this limit, unlike in the Rossby one or for the case of finite ω, there is perfect east–west symmetry. At the beginning,

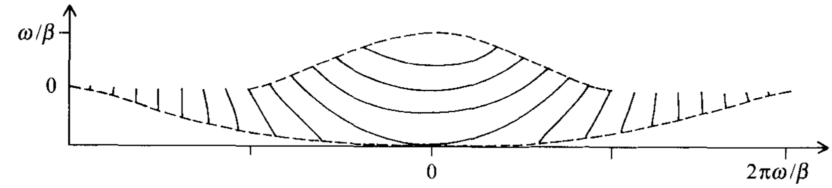

FIGURE 7. As in figure 6, but for the rays that go from the first to the second caustic. Contour interval is now $\Delta S = \pi/8$. Far from the source, phase propagation is as in figure 6. The southward phase propagation in the centre corresponds to a relatively small amplitude (see figure 5) and, furthermore, will be partially compensated by a northward phase propagation due to the component which is a north–south reflection of this one.

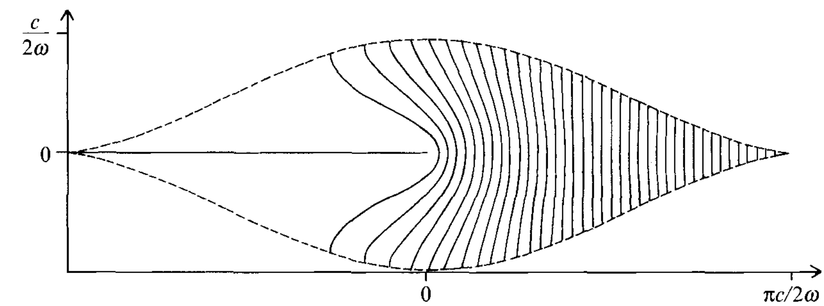

FIGURE 8. Phase function, $S(2x\omega/c, 2y\omega/c)$, in the Rossby limit and for the rays that travel from the source to the first caustic. The complete phase is $\vartheta = (\beta c/\omega^2)S - \omega t$. S increases to the west and the contour interval is $\Delta S = \pi/20$. There is north–south symmetry but not east–west. To the east of the source, $x > 0$, the phase propagation becomes quickly zonal; more precisely, $\vartheta \approx -(\beta x/\omega + \omega t)$. To the west, phase propagation is extremely rapid.

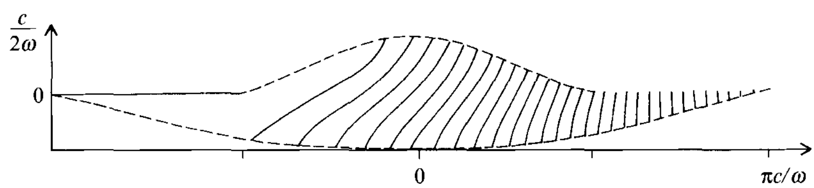

FIGURE 9. As in figure 8, but for the rays that go from the first to the second caustic. S increases to the west, the contour interval is now $\Delta S = \pi/8$. Far from the source, phase propagation is as described in figure 8.

the phase contours are nearly circular because the Coriolis effect is unimportant. Further away from the source, phase propagation becomes nearly zonal. For large values of μ, (6.5) shows that significant amplitudes (small J) correspond to $\alpha \approx 0$ or $\alpha \approx \pi$. By using this in (5.5) one obtains $\vartheta \approx (\omega^2/\beta c)\mu - \omega t \approx \omega(|x|/c - t)$, for $|x| \gg \omega/\beta$, which corresponds to short Poincaré waves.

In the Rossby limit, (5.6) and (5.7), the successive caustics are still given by (6.6), replacing γ by $-(\beta c)^{\frac{1}{2}}/2\omega$. Contours of the function S, defined in (6.1), are presented in figure 8 for the rays that go from the source to the first caustic, and in figure 9 for the rays that travel from the first to the second caustic. Unlike for the Poincaré case, there is a clear distinction between east and west in this limit.

East of the source, phase propagation is nearly zonal and westward (i.e. zonal phase speed is negative). More precisely, making $\alpha \approx 0$ in (5.7), as before, one gets $\vartheta \approx (\beta c/2\omega^2)\mu - \omega t \approx -(\beta x/\omega + \omega t)$, for $x \gg c/|\omega|$, which corresponds to short Rossby waves.

West of the source, on the other hand, phase propagation is extremely rapid. If $\alpha = \pi + \nu$, with $|\nu| \ll 1$, in (5.7) it follows that $k = O(\nu^2)$ but $l, y = O(\nu)$ and $x = O(\nu^0)$; therefore $\omega \approx -\beta k/(l^2+\beta^2 y^2/c^2)$, which corresponds to very long (non-dispersive) Rossby waves. In particular, for $x < 0$ and $y = 0$, it is $S \equiv 0$ (instantaneous phase propagation!). This represents the limit $\nu \rightarrow 0$, for which the zonal group velocity, $-4\omega^2/\beta\nu$, eventually diverges (see figure 1 and the last paragraph in §4), no matter how small $2\omega^2/\beta c$ may be. Clearly the validity of ray theory must break down in the neighbourhood of the negative x-axis, as a consequence of the asymptotic lack of dispersion of Rossby waves in the limit $k \rightarrow 0$.

7. Concluding remarks

Linearized free dynamics in the equatorial β-plane is studied, considering a single vertical mode, characterized by the *separation constant* c; an appropriate frequency scale is given by $\tilde{\omega} = (\frac{1}{2}\beta c)^{\frac{1}{2}}$. The fields of the pressure and both velocity components are written – for linear dynamics – in terms of a single potential φ and its first and second time derivatives; a single equation controls its evolution. Introduction of this potential eliminates the problem of the spurious solution, which is encountered when working with a single equation for the meridional velocity component. In the new formalism, addition of the spurious solution is but a gauge transformation of the potential, which leaves the dynamical fields unaltered. The new formalism filters out the Kelvin waves and retains the correct conservation laws of potential vorticity, energy and pseudomomentum.

The single evolution equation is suitable for the development of the ray theory (RT) approximation. The general solution of the ray equations is derived; this includes the ray paths and amplitude, and also the often-neglected phase field. This solution is found to be asymptotically correct in either the Poincaré ($\omega \gg \tilde{\omega}$) or Rossby ($\omega \ll \tilde{\omega}$) limits. More precisely, it represents an $O(\gamma^{-4})$ approximation, where $2\gamma^2 = (\omega/\tilde{\omega})^2 + (\tilde{\omega}/\omega)^2$ is a large parameter.

Two monochromatic examples are presented to illustrate the method: the meridional normal modes and the solution corresponding to a point source on the equator.

In the first case, the exact dispersion relation is obtained, by requiring a phase match between a set of rays that leave the equator with phase proportional to the longitude, and those that have turned at both caustics (turning latitudes).

In the second example, the caustics bend towards the equator, which they reach, in the form a cusp, at a focal point. Similar foci were found by Schopf *et al.* (1981), in another monochromatic problem, and by Jacobs (1967), for an initial value problem.

A typical frequency scale $\tilde{\omega}$ for the first baroclinic mode in the tropical Pacific (c in the range 2.5–3 m s^{-1}) corresponds to a period of about 13 days, which implies a deformation radius $c/\tilde{\omega}$ of 490 km (4.4°). For a point source with a one-month period ($\gamma = 1.66$, $\gamma^{-4} = 0.13$) the meridional half-width $\gamma(c/\beta)^{\frac{1}{2}}$ equals 580 km (5.2°) and the distance between foci $\pi\gamma(c/\beta)^{\frac{1}{2}}$ equals 1.8 Mm (16°). Moreover, the wavenumber scale $\gamma(\beta/c)^{\frac{1}{2}}$ corresponds to a wavelength of 1.3 Mm, twice that of short Rossby waves. Identical scales are obtained for the other frequency which gives the same value of γ, i.e. for a period of 5.6 days ($5.6 = 13^2/30$). The order of magnitude of these numbers indicates that phenomena predicted by ray theory, like focusing, may be useful in the understanding of equatorial dynamics.

The two main results of this paper are the description of the non-Kelvin wave field in terms of a single scalar potential and its use in ray theory, stressing the importance of the amplitude and phase fields, in addition to the ray trajectories. An interesting extension would be to perform the ray theory calculations in three dimensions, e.g. using the potential developed by Moore (1993). In order to apply these results to realistic oceanic or atmospheric phenomena one should include the effects of mean currents (see for instance Chang & Philander, 1988). These important extensions are beyond the scope of the present paper.

I want to express my gratitude to Drs Julio Sheinbaum and José Ochoa for comments and corrections, and to the CONACyT (México) for its support under grant 1282-T9204.

Appendix

Taking successive time derivatives (denoted by a subscript t) of $\varphi = A\,\mathrm{e}^{\mathrm{i}\vartheta}$, it is obtained that

$$\varphi_t = (A_t - \mathrm{i}\omega A)\,\mathrm{e}^{\mathrm{i}\vartheta},$$
$$\varphi_{tt} = (A_{tt} - \omega^2 A - \mathrm{i}\omega_t A - 2\mathrm{i}\omega A_t)\,\mathrm{e}^{\mathrm{i}\vartheta},$$

etc. Similarly, it is also found that

$$\nabla\varphi = (\nabla A + \mathrm{i}\boldsymbol{k}A)\,\mathrm{e}^{\mathrm{i}\vartheta},$$
$$\nabla\varphi_t = (\nabla A_t + \omega\boldsymbol{k}A - \mathrm{i}\omega\nabla A + \mathrm{i}\boldsymbol{k}A_t - \mathrm{i}A\nabla\omega)\,\mathrm{e}^{\mathrm{i}\vartheta}.$$

(A trivial consequence of the definition (4.4) of $\boldsymbol{k}$ and ω is $\boldsymbol{k}_t = -\nabla\omega$, which was used in the derivation of last equation.) These two equations are used to evaluate the average, in one phase cycle, of both ingredients of the conservation law (3.7), viz.

$$\left.\begin{aligned}\langle I\rangle &= \tfrac{1}{2}(3\omega^2/c^2 - f^2/c^2 - \boldsymbol{k}^2)\,A^2 + (A_t^2 - 2AA_{tt})/2c^2 - \tfrac{1}{2}(\nabla A)^2,\\ \langle \boldsymbol{F}\rangle &= (\boldsymbol{k}\omega - \tfrac{1}{2}\boldsymbol{\beta})\,A^2 + A\nabla A_t.\end{aligned}\right\} \qquad (\mathrm{A}\ 1)$$

In order to obtain the representation of the evolution equation (3.5), the following derivatives of (4.1) are also needed:

$$\nabla^2\varphi = (\nabla^2 A - \boldsymbol{k}^2 A + \mathrm{i}2\boldsymbol{k}\cdot\nabla A + \mathrm{i}A\nabla\cdot\boldsymbol{k})\,\mathrm{e}^{\mathrm{i}\vartheta},$$
$$\varphi_{ttt} = (A_{ttt} - 3\omega\omega_t A - 3\omega^2 A_t + \mathrm{i}\omega^3 A - \mathrm{i}\omega_{tt} A - 3\mathrm{i}\omega_t A_t - 3\mathrm{i}\omega A_{tt})\,\mathrm{e}^{\mathrm{i}\vartheta},$$
$$\nabla^2\varphi_t = (\nabla^2 A - \boldsymbol{k}^2 A)_t + \omega(2\boldsymbol{k}\cdot\nabla A + A\nabla\cdot\boldsymbol{k}) - \mathrm{i}\omega(\nabla^2 A - \boldsymbol{k}^2 A) + \mathrm{i}(2\boldsymbol{k}\cdot\nabla A + A\nabla\cdot\boldsymbol{k})_t.$$

Substitution in (3.5) and multiplication by $\mathrm{e}^{-\mathrm{i}\vartheta}$ gives the following horrendous equation:

$$\frac{-3\omega\omega_t A - 3\omega^2 A_t + f^2 A_t}{c^2} + (\boldsymbol{k}^2 A)_t - \omega(2\boldsymbol{k}\cdot\nabla A + A\nabla\cdot\boldsymbol{k})$$
$$+\boldsymbol{\beta}\cdot\nabla A + \mathrm{i}\left[\omega\left(\frac{\omega^2 - f^2}{c^2} - \boldsymbol{k}^2\right) + \boldsymbol{\beta}\cdot\boldsymbol{k}\right]A = -\frac{A_{ttt}}{c^2} + \nabla^2 A_t + \mathrm{i}(2\boldsymbol{k}\cdot\nabla A + A\nabla\cdot\boldsymbol{k})_t$$
$$+\frac{\mathrm{i}(\omega_{tt} A + 3\omega_t A_t + 3\omega A_{tt})}{c^2} - \mathrm{i}\omega\nabla^2 A, \qquad (\mathrm{A}\ 2)$$

which is used to derive the conservation law (4.6) and the dispersion relation (4.7).

In order to get the representation of the energy density E, the expressions (3.3) (for p, u and v as a function of φ) are substituted in (2.5a); this gives

$$2E = c^2\varphi_{yt}^2 + c^2 f^2 \varphi_x^2 + c^4\varphi_{yx}^2 + f^2\varphi_t^2 + \varphi_{tt}^2 + c^4\varphi_{xx}^2$$
$$-2c^2\varphi_{tt}\varphi_{xx} + 2c^2 f\varphi_{yt}\varphi_x + 2c^2 f\varphi_{xy}\varphi_t. \qquad (\mathrm{A}\ 3)$$

The last two terms can be replaced by $2c^2\,\partial_y(f\varphi_t\varphi_x)-2c^2\beta\varphi_t\varphi_x$. In correspondence with the RT limit, when calculating the phase average of E, only derivatives of the phase are retained and not those of $\boldsymbol{k}$, ω and A. Thus, since $\langle 2c^2\,\partial_y(f\varphi_t\varphi_x)\rangle = c^2\,\partial_y(\omega kA^2)$ this term is not included. Consequently,

$$4\langle E\rangle/A^2 = (f^2+l^2c^2)(\omega^2+k^2c^2)+(\omega^2-k^2c^2)^2+2\beta c^2\omega k$$
$$\equiv (\omega^2-k^2c^2)(2\omega^2+\beta c^2k/\omega), \quad \text{(A 4)}$$

where the dispersion relation (4.7) has been used to simplify the expression.

For the pseudomomentum density (2.5*b*), using (3.3) and (3.6) and after subtracting some exact divergences it is similarly found that

$$2P = 2\varphi_t(\varphi_{tt}-c^2\varphi_{xx})_x+\beta c^2\varphi_x^2-(\varphi_{ttx}+f^2\varphi_x-c^2\nabla^2\varphi_x)^2/\beta, \quad \text{(A 5)}$$

which gives the following phase average in the RT limit:

$$4\langle P\rangle/A^2 = 2\omega k(\omega^2-k^2c^2)+\beta c^2k^2-(\omega^2-f^2-\boldsymbol{k}^2c^2)^2/\beta$$
$$\equiv (k/\omega)(\omega^2-k^2c^2)(2\omega^2+\beta c^2k/\omega). \quad \text{(A 6)}$$

REFERENCES

Blandford, R. 1966 Mixed gravity-Rossby waves in the oceans. *Deep-Sea Res.* **13**, 941–961.

Cane, M. A. & Sarachik, E. S. 1979 Forced baroclinic ocean motion, III: The linear equatorial basin case. *J. Mar. Res.* **37**, 355–398.

Chang, P. & Philander, S. G. H. 1988 Rossby wave packets in baroclinic mean currents. *Deep-Sea Res.* **36**, 17–37.

Jacobs, S. J. 1967 An asymptotic solution of the tidal equations. *J. Fluid Mech.* **30**, 417–438.

Kessler, W. S. 1990 Observations of long Rossby waves in the Northern Tropical Pacific. *J. Geophys. Res.* **95**, 5183–5217.

Klauder, J. R. 1988 New asymptotics for old wave equations. *Science* **239**, 760–762.

Laplace, P. S. de 1778 Recherches sur plusieurs points du système du monde. *Mém. Acad. R. Sci. Paris* 75–182.

Lighthill, J. 1978 *Waves in Fluids*. Cambridge University Press.

Ludwig, D. 1966 Uniform asymptotic expansions at a caustic. *Commun. Pure Appl. Maths* **19**, 215–250.

Matsuno, T. 1966 Quasi-geostrophic motions in the equatorial area. *J. Met. Soc. Japan* **48**, 25–43.

McCreary, J. P. 1984 Equatorial beams. *J. Mar. Res.* **42**, 395–430.

Moore, D. 1994 Equatorial waves and the effect of the horizontal component of the Earth's rotation. *Geophys. Res. Lett.* Submitted.

Moore, D. W. & Philander, S. G. H. 1977 Modelling of the tropical oceanic circulation. In *The Sea*, vol. 6 (ed. E. D. Goldberg, I. N. McCave, J. J. O'Brien & J. H. Steele), pp. 319–361. John Wiley.

Peregrine, D. H. & Smith, R. 1979 Nonlinear effects upon waves near caustics. *Phil. Trans. R. Soc. Lond.* A **268**, 341–370.

Philander, S. G. H. 1977 The effect of coastal geometry on equatorial waves (Forced waves in the Gulf of Guinea). *J. Mar. Res.* **35**, 509–523.

Ripa, P. 1982 Nonlinear wave–wave interactions in a one-layer reduced gravity model on the equatorial beta-plane. *J. Phys. Oceanogr.* **12**, 97–111.

Ripa, P. 1983 Definite Integral of the product of Hermite functions, with applications to the theory of nonlinear interactions among equatorial waves. *J. Geophys. Res.* **88**, 9741–9744.

Ripa, P. & Carrasco, A. C. 1993 Ray theory and the scattering of Topographic Rossby Waves by an abrupt change of shelf width and coastline. *J. Geophys. Res.* **98** C, 22693–22705.

Schopf, P., Anderson, D. L. T. & Smith, R. 1981 Beta-dispersion of low-frequency Rossby waves. *Dyn. Atmos. Oceans* **5**, 187–214.

Wunsch, C. 1977 Response of an equatorial ocean to a periodic monsoon. *J. Phys. Oceanogr.* **7**, 497–511.