Equatorial Solitary Waves. Part 2: Envelope Solitons

JOHN P. BOYD

Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor, MI 48109

(Manuscript received 23 December 1981, in final form 17 September 1982)

ABSTRACT

Via the method of multiple scales, it is shown that the time and space evolution of the envelope of wave packets of weakly nonlinear, strongly dispersive equatorial waves is governed by the Nonlinear Schrödinger equation. The diverse phenomena of this equation—envelope solitons, sideband instability, FPU recurrence and more—is briefly reviewed. Which of the alternatives occurs is determined largely by the relative signs of the coefficients of the dispersive and nonlinear terms in the Nonlinear Schrödinger equation which, for a given latitudinal mode, are functions of a single nondimensional parameter, the zonal wavenumber k. Gravity waves propagating toward the east form solitary waves and are subject to sideband instability only for large k. The mixed Rossby-gravity wave has solitons only in a range of intermediate k.

For waves with group velocities toward the west, the physics is much more complicated because these waves have an *infinite* number of second harmonic resonances and long wave/short wave resonances tucked into a finite interval of intermediate wavenumber. These two species of resonance form the topic of the two companion papers Boyd (1983a,b). One finds that for westward-propagating gravity waves, the resonances are very weak, and solitary waves occur more or less continuously within an intermediate range of wavenumber. For Rossby waves one can also state that solitons are forbidden for both large and small k, but the resonances completely dominate the intermediate range of wavenumbers so that the situation is very confused and complicated.

Together with earlier papers, this present work completes the description of the weakly nonlinear evolution of equatorial waves in a shallow water wave model.

1. Introduction

In two previous papers, the author has discussed nonlinear effects on the non-dispersive equatorial Kelvin wave (Boyd, 1980a) and on weakly dispersive long Rossby waves (Boyd, 1980b). This third paper and its two companions explore nonlinearity for strongly dispersive equatorial waves including short Rossby waves, the mixed Rossby-gravity wave, and all eastward and westward travelling gravity waves. The most striking of these nonlinear effects is the creation of solitary waves. Although this present work is the third in the series, it is only the second (hence "Part 2" in its title) to deal with solitary waves because the Kelvin waves of Boyd (1980a), being nondispersive, cannot form solitary waves.

As in these two earlier papers, the method of multiple scales is applied to the nonlinear shallow water wave equations to create an analytic singular perturbation theory with the wave amplitude as the expansion parameter. Since one physical definition of "strong dispersion" is that a single isolated crest will break up into several crests and troughs on a time scale of order unity, it follows that a balance between nonlinearity and dispersion for such a single crest is impossible within the framework of the theory except for the long—and weakly dispersive—Rossby waves of Boyd (1980b). If one considers an initial packet whose Fourier transform is sharply peaked about some central packet wavenumber k or equivalently, a disturbance of the form

$$u(x, y, t) = A(x, t) \exp[ikx - iw(k)t]\phi_n(y) + \text{c.c.}, \quad (1.1)$$

where c.c. is an abbreviation for "complex conjugate," $\phi_n(y)$ describes the latitudinal modal structure of the packet, and $A(x, t)$ is a complex function called the "envelope" which varies slowly with x on a length scale $L \approx O(\epsilon^{-1})$ where $\epsilon \ll 1$, then solitary waves are still possible. Because the range of wavenumbers in the packet has a width of only $O(\epsilon)$, the characteristic time scale for the spreading of the envelope is the slow scale $T \approx O(\epsilon^{-1})$. Weak nonlinearity—the only kind for which the perturbation theory is valid—can balance the slow dispersive spreading of the envelope to create a steadily moving nonlinear wave of permanent form: an "envelope" soliton. The reason for the name is, although the envelope $A(x, t)$ has but a single peak (for a single soliton), $u(x, y, t)$ has many individual crests and troughs [$O(\epsilon^{-1})$ in number] as shown schematically in Fig. 1.

The latitudinal structure function $\phi_n(y)$ and the frequency $w(k)$ are precisely those of linear equatorial wave theory, but the envelope $A(x, t)$ is governed by the so-called Nonlinear Schrödinger equation or NLS for short, i.e.,

$$iA_t + \tfrac{1}{2}w''A_{xx} + \nu|A|^2A = 0, \quad (1.2)$$

where w'' is the second derivative of the dispersion relation $w(k)$ with respect to k and where $\nu(k;\, n)$ is the so-called Landau constant, independent of x but varying with both the wavenumber k and the latitudinal mode number n. The central goal of the perturbation theory is to derive (1.2) and calculate the Landau constant. The next section will present an overview of the theory and Section 3 the actual derivation. Unfortunately, Eq. (1.2) is not the whole story: resonances cause the Landau constant to have a pole at certain values of k and n, and in the neighborhood of these resonances, one must abandon (1.2) in favor of coupled pairs of equations. The resonances are discussed in Section 4; a full treatment of the long wave/short wave resonance equations and of the resonant dyad equations is given in the companion papers (Boyd, 1983a,b).

Section 5 discusses the variations of the Landau constant with k and n. The graphs for Rossby waves and for westward-travelling gravity waves are very confusing because of all the resonances.

Section 6 explores the different solutions of the Nonlinear Schrödinger equation: solitons, "breathers," FPU recurrence and polycnoidal waves.

The final section is a summary and prospectus, while the three appendices discuss (A) numerical methods for computing the Landau constant, (B) error theorems for approximate equatorial wave dispersion formulas, and (C) the derivation of the Nonlinear Schrödinger equation from the Korteweg–deVries equation.

2. Strong versus weak dispersion: An overview of the derivation of the Nonlinear Schrödinger Equation

The Korteweg–deVries equation of Boyd (1980b) is (in canonical form)

$$A_t + AA_x + A_{xxx} = 0, \quad (2.1)$$

which is *quadratically* nonlinear whereas the NLS equation (1.2) is *cubically* nonlinear. In this section, we will briefly explore the reasons for this difference and present an overview of the derivation of (1.2) whose details are postponed until the next section.

Resonance lies at the heart of the whole perturbation theory. Both the Korteweg–deVries and Nonlinear Schrödinger equations describe the resonant interaction of a single wave mode with itself. These two differ in form simply because the strength of the dispersion has a decisive role in determining how this resonant self-interaction takes place and at what order in the perturbation theory. The resonant triad equations of Domaracki and Loesch (1977), Loesch and Deininger (1979) and Ripa (1983a,b) and the long wave/short resonance equations and resonant dyad equations of Boyd (1983a,b) describe pairs or triplets of wave modes whose mutual interaction is stronger than their self-interaction. In general, however, these mutual resonances occur only for certain discrete values of the wavenumbers and mode numbers. A given mode, however, is *always* in resonance with itself at appropriately high perturbation order. Consequently, the NLS equation describes the nonlinear dynamics of strongly dispersive waves everywhere in parameter space (that the dispersion is strong) except for certain deleted neighborhoods where one of these mutual resonances is more important.

The Korteweg–deVries equation itself represents one of these deleted neighborhoods. The dispersion relation for linear Rossby waves is approximately

$$c \approx \frac{-1}{2n + 1 + k^2}, \quad (2.2)$$

where n is the latitudinal mode number. For long Rossby waves for which

$$k \approx O(\epsilon^{1/2}), \quad (2.3)$$

where

$$\epsilon \ll 1, \quad (2.4)$$

it follows that

$$c(k) - c(0) \approx O(\epsilon). \quad (2.5)$$

In words, waves whose wavenumbers differ by O(1) will have phase speeds which differ only by a small amount. Eq. (2.5) can be taken as the mathematical definition of "weak" dispersion. It implies that even if a packet of long Rossby waves is not sharply peaked in wavenumber space about some central packet wavenumber k, the packet nonetheless will spread very slowly, on a time scale $T \approx O(\epsilon^{-1})$, because its constituent waves are all traveling at approximately the same speed. If the amplitude of the packet is also $O(\epsilon)$, then one can balance the nonlinearity against the dispersion to create the unimontane, singly-humped solitons shown in Fig. 1a.

From a resonance viewpoint, Eq. (2.5) implies that there is approximate resonance between *all* long Rossby waves of the same meridional mode. Consequently, the Korteweg–deVries equation is only quadratically nonlinear: a long Rossby wave of wavenumber k is simultaneously in approximate resonance [to $O(\epsilon)$] with both its second harmonic (wavenumber $2k$) and the mean flow ($k = 0$).

For strongly dispersive waves, however,

$$c(k) - c(0) \approx O(1) \quad (2.6)$$

almost everywhere in parameter space. The self-interaction of a wave packet of the form of (1.1) will give rise to forced waves which depend on x and t as either

$$A^2(x, t)e^{2ik[x-c(k)t]} \text{ [second harmonic]} \quad (2.7)$$

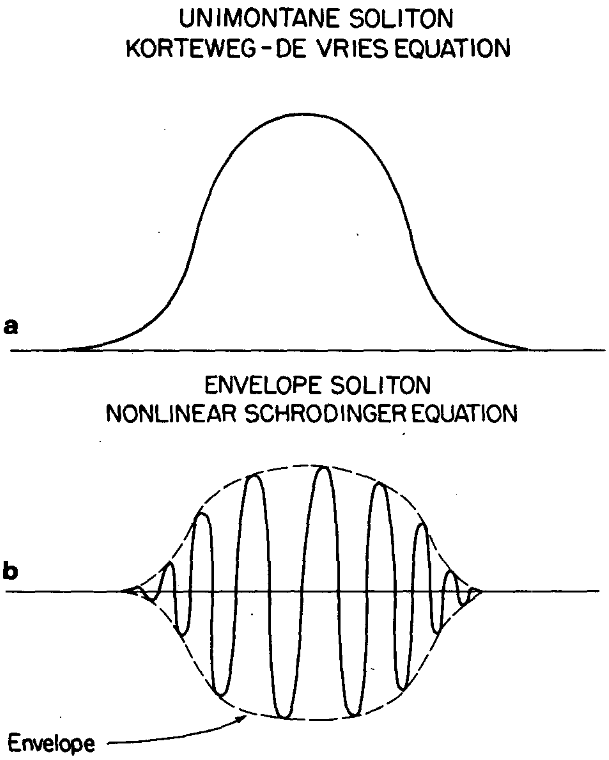

FIG. 1. (a) A soliton of the Korteweg–deVries equation or the long-wave component of a soliton of the long-wave/short-wave resonance equations. (b) An envelope soliton of the Nonlinear Schrödinger equation or the short-wave component of the long-wave/short-wave resonance equations.

and its complex conjugate or as

$$|A(x, t)|^2 \text{ [long wave or "mean"]}. \tag{2.8}$$

If the packet were a simple plane wave, then (2.8) would be the "mean," i.e., the zonally-averaged flow, but when $A(x, t)$ varies with x, Eq. (2.8) is dynamically a long Rossby wave, so the other description is more accurate. Because of (2.6), however, both (2.7) and (2.8) [multiplied by the appropriate latitudinal modal functions] are forced waves, travelling at speeds sufficiently different from those of the free oscillations of the same wavenumber and mode number so that resonance generally does not occur. [When it does, one has one of the mutual resonances discussed further in Section 4 and in Boyd (1983a,b).]

In the absence of such resonances, however, these forced waves are $O(\epsilon)$ in comparison to the fundamental wave packet (1.1)—and not very important. At the next perturbation order, the nonlinear interaction between the forced components and the fundamental will give rise to terms proportional to

$$|A|^2 A e^{ik(x-c(k)t)}, \tag{2.9}$$

which *are* resonant with the fundamental. Because of the resonance, Eq. (2.9), although of higher order than (2.7) and (2.8), can create O(1) changes in the structure and evolution of the wave packet and dominate its nonlinear dynamics.

The tree diagram (Fig. 2) summarizes the situation. The dynamics of the lowest order solution are completely determined, except for the factor $A(x, t)$, by linear equatorial wave theory. The forced wave components (2.7) and (2.8) are only small corrections, but ironically, the bulk of the labor in computing the Landau constant is to calculate these forced components because the value of $\nu(k; n)$ is determined solely by the interaction of these forced waves with the fundamental.

3. Derivation of the Nonlinear Schrödinger equation: Mathematical detail

The nonlinear shallow water wave equations are given in nondimensional form by

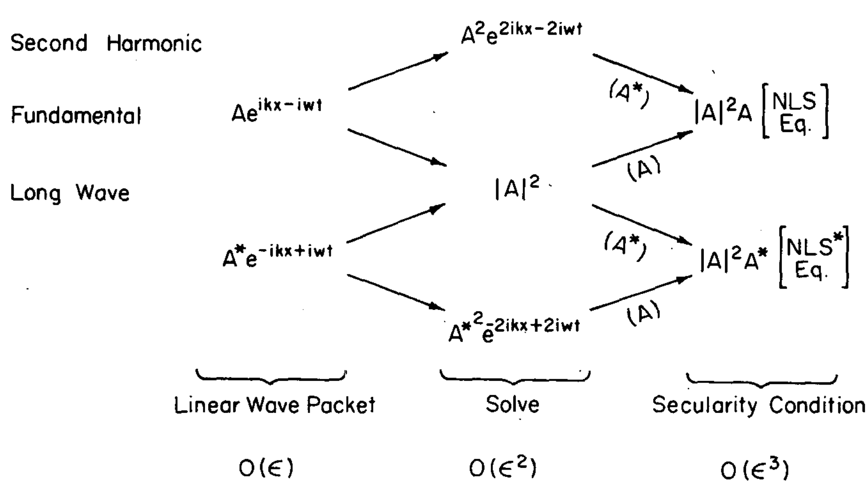

FIG. 2. "Tree" diagram schematically showing how higher harmonics are generated by multiplication at successive orders. The third harmonic [$O(\epsilon^3)$] and the y-dependence of the fundamental and its harmonics have been omitted.

$$u_t + uu_x + vu_y - yv + \phi_x = 0, \quad (3.1)$$

$$v_t + uv_x + vv_y + yu + \phi_y = 0, \quad (3.2)$$

$$\phi_t + (\phi u)_x + (\phi v)_y + u_x + v_y = 0, \quad (3.3)$$

where ϕ is the height. As in Boyd (1980a,b) and most other recent papers on equatorial waves in the ocean, the nondimensional scales are

$$L = \tilde{E}^{-1/4} a \text{ [length scale]}, \quad (3.4)$$

$$T = \tilde{E}^{1/4}(2\Omega)^{-1} \text{ [time scale]}, \quad (3.5)$$

where

$$\tilde{E} = \frac{4\Omega^2 a^2}{gH} \quad (3.6)$$

is often referred to as "Lamb's parameter" with Ω the angular frequency of the earth's rotation ($2\pi/86\,400$ s), a the radius of the earth, and H the "equivalent depth." For the first baroclinic mode, the equivalent depth H is 40–60 cm so that $L \approx 300$ km and $T \approx 1.7$ days.

To proceed further, we let ϵ be the amplitude of the wave packet where $\epsilon \ll 1$ and apply the method of multiple scales. We expand u as

$$u \approx \epsilon \sum_{r=0}^{\infty} \epsilon^r \sum_{s=0}^{r+1} A_{rs}(\zeta, \tau)\theta_{rs}(y)E^s + \text{c.c.} \quad (3.7)$$

and similarly for v and ϕ, where

$$E \equiv e^{ikx - iw(k)t}, \quad (3.8)$$

where k is the central wavenumber of the packet (i.e., the location of the peak of the Fourier transform of the wavepacket) with $w(k)$ the linear dispersion relation and

$$\tau = \epsilon^2 t, \quad (3.9)$$

$$\zeta = \epsilon[x - c_g(k)t], \quad (3.10)$$

are the "slow" time and space scales with $c_g(k) = dw/dk$ the linear group velocity of the packet.

The scales in (3.9) and (3.10) are chosen so that dispersive and nonlinear effects will be the same order of magnitude, making envelope solitons possible. The group velocity appears explicitly in (3.10) because this is approximately the speed at which the envelope of the wave packet travels. Since (3.7)–(3.10) have appeared unchanged in innumerable previous derivations of the NLS equation–Johnson (1976) is a readable example—no further attempt to justify them will be made beyond the statement that these equations are the mathematical embodiment of the ideas sketched in Section 2. The differences between one NLS derivation and another lie solely in the linear dispersion $w(k)$ and in latitudinal structure functions $\theta_{nm}(y)$, so it is upon these matters that we shall mainly concentrate.

The order-by-order perturbation equations are obtained by substituting (3.7) (and the similar expansions for v and ϕ) into (3.1)–(3.3) and using the chain rules

$$\frac{\partial}{\partial t} = -iw - \epsilon c_g \frac{\partial}{\partial \zeta} + \epsilon^2 \frac{\partial}{\partial \tau}, \quad (3.11)$$

$$\frac{\partial}{\partial x} = ik + \epsilon \frac{\partial}{\partial \zeta}. \quad (3.12)$$

The O(ϵ) perturbation equations are identical with those of ordinary linear theory for an equatorial wave of zonal wavenumber k.

O($\epsilon^1 E^1$):

$$-iwu^{01} - yv^{01} + ik\phi^{01} = 0, \quad (3.13a)$$

$$-iwv^{01} + yu^{01} + \phi_y^{01} = 0, \quad (3.13b)$$

$$-iw\phi^{01} + iku^{01} + v_y^{01} = 0, \quad (3.13c)$$

with the solution

$$u^{01} = A_{01}Q^{-1}(\zeta, \tau)E[S_1 H_{n+1}(y) + S_3 H_{n-1}(y)]e^{-(1/2)y^2} + \text{c.c.}, \quad (3.14a)$$

$$v^{01} = A_{01}Q^{-1}(\zeta, \tau)E[iS_2 H_n(y)]e^{-(1/2)y^2} + \text{c.c.}, \quad (3.14b)$$

$$\phi^{01} = A_{01}Q^{-1}(\zeta, \tau)E[S_1 H_{n+1}(y) - S_3 H_{n-1}(y)]e^{-(1/2)y^2} + \text{c.c.}, \quad (3.14c)$$

with

$$S_1 = \tfrac{1}{2}(w + k), \quad (3.15a)$$

$$S_2 = -(w^2 - k^2), \quad (3.15b)$$

$$S_3 = n(w - k), \quad (3.15c)$$

where Q is the normalization factor

$$Q^2 \equiv \int_{-\infty}^{\infty} dy\, e^{-y^2}[2S_1^2 H_{n+1}(y)^2 + S_2^2 H_n(y)^2 + 2S_3^2 H_{n-1}(y)^2]. \quad (3.16)$$

and $H_n(y)$ is the nth (unnormalized) Hermite polynomial. The normalization factor is proportional to the latitudinally-averaged total energy (kinetic plus potential), so this choice for (3.16) may be referred to as an "energy normalization" for the primary wave packet. [See also the Appendix of Boyd (1982b).]

The dispersion relation $w(k)$ is determined by

$$w^3 - w(2n + 1 + k^2) - k = 0. \quad (3.17)$$

For $n \geq 1$, this has three roots, physically corresponding to a westward-traveling Rossby wave, an eastward-traveling gravity wave, and a westward-traveling gravity wave. For $n = 0$, one of roots ($w = -k$) leads to a spurious solution of (3.13); the two other roots give a westward traveling mixed Rossby–gravity wave and an eastward traveling gravity wave. The Kelvin wave for which $w(k) \equiv k$ is assigned $n = -1$, but because it is non-dispersive, it is irrelevant to this paper.

At $O(\epsilon^2)$, we have three contributions to solve for, proportional to E^0 ("long wave"), E^1 ("fundamental") and E^2 ("second harmonic"), respectively. The equations for the $O(\epsilon^2)$ fundamental are easily solved because their forcing terms are linear in the lowest order fields.

$O(\epsilon^2 E^1)$:

$$-iwu^{11} - yv^{11} + ik\phi^{11} = c_g u_\zeta^{01} - \phi_\zeta^{01}, \quad (3.18a)$$

$$-iwv^{11} + yu^{11} + \phi_y^{11} = c_g v_\zeta^{01}, \quad (3.18b)$$

$$-iw\phi^{11} + iku^{11} + v_y^{11} = c_g \phi_\zeta^{01} - u_\zeta^{01}, \quad (3.18c)$$

with the solution

$$u^{11} = -i\frac{\partial A_{01}}{\partial \zeta} E\left(\frac{dS_1}{dk} H_{n+1} + \frac{dS_3}{dk} H_{n-1}\right) \times e^{-(1/2)y^2} + \text{c.c.}, \quad (3.19a)$$

$$v^{11} = \frac{\partial A_{01}}{\partial \zeta} E\left(\frac{dS_2}{dk} H_n\right) e^{-(1/2)y^2} + \text{c.c.}, \quad (3.19b)$$

$$\phi^{11} = -i\frac{\partial A_{01}}{\partial \zeta} E\left(\frac{dS_1}{dk} H_{n+1} - \frac{dS_3}{dk} H_{n-1}\right) \times e^{-(1/2)y^2} + \text{c.c.}, \quad (3.19c)$$

which can be generated from the linear lowest order solution (3.14) by differentiating $A_{01}(\zeta, \tau)$ with respect to ζ, the S_i's with respect to k, and multiplying by $-i$.

Strictly speaking, Eq. (3.18) has no solution unless a non-secularity or solvability condition is satisfied. However, we shall not go into the details at this order because the condition merely yields $c_g = dw/dk$, which we knew already from the general linear theory of dispersive waves. Eqs. (3.18)–(3.19) merely perturbatively correct for the fact that the lowest order solution is not a sinusoidal wave of wavenumber k but is rather a wave packet.

Unfortunately, the equations for the $O(\epsilon^2)$ long waves and second harmonics that are forced by the nonlinear self-interaction of the wave packet are complicated; most of the labor in calculating the coefficients of the NLS equation is to solve them. The equations are

$O(\epsilon^2 E^0)$ [long wave]:

$$-c_g u^{10} - y\tilde{v}^{20} + \phi^{10} = F_1, \quad (3.20a)$$

$$yu^{10} + \phi_y^{10} = F_2, \quad (3.20b)$$

$$-c_g \phi^{10} + \tilde{v}_y^{20} + u^{10} = F_3. \quad (3.20c)$$

Eqs. (3.20a) and (3.20b) have been integrated with respect to ζ, and $\tilde{v}^{20}$ is the indefinite ζ integral of v^{20}. The reason that v^{20}, rather than v^{10}, appears in (3.20a) and (3.20c) with u^{10} and ϕ^{10}, is that one can show that for long waves, $v \approx O(\epsilon)u$ where the zonal length scale is $O(\epsilon^{-1})$. Put another way, the solutions of (3.20) are forced ultra-long Rossby waves, and the smallness of the north–south current in comparison to the zonal velocity is one of the characteristics of such waves. The forcing functions are

$$F_{1\zeta} = -u^{01}u_\zeta^{01} - v^{01}u_y^{11} - v^{11}u_y^{01}, \quad (3.21a)$$

$$F_2 = -u^{01}v_x^{01} - v^{01}v_y^{01}, \quad (3.21b)$$

$$F_{3\zeta} = -(u^{01}\phi^{01})_\zeta - (v^{11}\phi^{01})_y - (v^{01}\phi^{11})_y. \quad (3.21c)$$

An implicit convention is used above: an expression like $u^{01}u_\zeta^{01}$ means that part of $u^{01}u_\zeta^{01}$ which is proportional to $(|A_{01}|^2)_\zeta$, excluding those other terms in $u^{01}u_\zeta^{01}$ which are second harmonics. Similarly, in writing down the forcing terms for the second harmonics below, we shall implicitly exclude the long-wave terms included in F_1–F_3 above.

The reader can easily verify by replacing u^{01}, u^{11}, etc., by the solutions given above that all terms in (3.21a) and (3.21c) are proportional to $(|A_{01}|^2)_\zeta$ so that the integration with respect to ζ is trivial. (Note that u^{11}, v^{11} and ϕ^{11} are proportional to the ζ-derivative of A_{01}.)

The appearance of the $O(\epsilon^2)$ quantities u^{11}, v^{11} and ϕ^{11} in equations which determine other $O(\epsilon^3)$ quantities (u^{10} and ϕ^{10}) is at first surprising. However, the long-wave quantities are independent of the "fast" length scale so that $u_x^{10} = \epsilon u_\zeta^{10}$ + higher order $\approx O(\epsilon^2)$, even though u^{10} is itself of $O(\epsilon^2)$. Thus, all the terms in (3.21a) and (3.21c) are $O(\epsilon^3)$ (before the ζ-integration) and $O(\epsilon^3)$ contributions from terms like $v^{01}u_y^{11}$ must be included. $O(\epsilon^2)$ terms like $v^{01}u_y^{01}$ do not appear because v^{01} is 90° out of phase with u^{01} and ϕ^{01} as mathematically expressed by the factor of $i = \sqrt{-1}$ in (3.14b). (This is discussed further in Appendix A.) This "deferred determination of the long waves or mean flow" is quite common in multiple-scale perturbation theories of hydrodynamic waves, whether stable or unstable.

To solve (3.20), we use Galerkin's method with the Hermite functions as the basis set. If applied in a brute force way, this would necessitate solving a (truncated) infinite number of linear algebraic equations in an infinite number of unknowns. By using two techniques previously employed in Boyd (1980a,b), however, this infinite set of equations can be split into uncoupled triplets of equations, each of which can be solved independently of the others.

The first procedure is to replace u^{10} and ϕ^{10} by the "sum" and "difference" variables

$$S^{10} \equiv \phi^{10} + u^{10}, \quad (3.22a)$$

$$D^{10} \equiv \phi^{10} - u^{10}. \quad (3.22b)$$

Using L and R to denote the linear operators

$$L \equiv \frac{\partial}{\partial y} + y, \quad (3.23a)$$

$$R \equiv \frac{\partial}{\partial y} - y, \tag{3.23b}$$

$$(1 - c_g)S^{10} + R\tilde{v}^{20} = F_1 + F_3, \tag{3.24a}$$

$$\tfrac{1}{2}LS^{10} + \tfrac{1}{2}RD^{10} = F_2, \tag{3.24b}$$

$$(-1 - c_g)D^{10} + L\tilde{v}^{20} = F_3 - F_1. \tag{3.24c}$$

The second trick is to recognize that R and L, which contain all the y derivatives in (3.24) are the "raising" and "lowering" operators for the Hermite functions, i.e.,

$$R[H_n(y)e^{-(1/2)y^2}] = -H_{n+1}(y)e^{-(1/2)y^2} \text{ [raising operator]}, \tag{3.25a}$$

$$L[H_n(y)e^{-(1/2)y^2}] = 2nH_{n-1}(y)e^{-(1/2)y^2} \text{ [lowering operator]}. \tag{3.25b}$$

If we now expand

$$F_1 = e^{-(1/2)y^2} \sum_{m=0}^{\infty} (F_1)_m H_m(y), \tag{3.26a}$$

$$S^{10} = e^{-(1/2)y^2} \sum_{m=0}^{\infty} S_m^{10} H_m(y), \tag{3.26b}$$

with similar expansions for the other forcings and unknowns, then substituting (3.26) into (3.24) and using (3.25) gives the single equation

$$S_0^{10} = [(F_1)_0 + (F_3)_0]/(1 - c_g), \tag{3.27}$$

which gives the forced Kelvin wave plus the triplets which in matrix form are

$$\begin{vmatrix} (1 - c_g) & -1 & 0 \\ (m + 1) & 0 & -\tfrac{1}{2} \\ 0 & 2m & (-1 - c_g) \end{vmatrix} \begin{vmatrix} S_{m+1}^{10} \\ \tilde{v}_m^{20} \\ D_{m-1}^{10} \end{vmatrix} = \begin{vmatrix} (F_1)_{m+1} + (F_3)_{m+1} \\ (F_2)_m \\ (F_3)_{m-1} - (F_1)_{m-1} \end{vmatrix}. \tag{3.28}$$

There is one such triplet for all integral m, $m \geq 1$. Technically, there is also a doublet of equations for $m = 0$, but one can show that, regardless of the symmetry of the primary wave packet, the forcing terms for (3.24a, c) are symmetric about $y = 0$ while F_2 is always antisymmetric. This implies that (3.28) has a non-zero solution only for *odd m*. Elementary linear algebra gives

$$S_{m+1}^{10} = (1/\Delta)\{-m[(F_3)_{m+1} + (F_1)_{m+1}] + (c_g + 1) \times (F_2)_m + \tfrac{1}{2}[(F_1)_{m-1} - (F_3)_{m-1}]\}, \tag{3.29a}$$

$$D_{m-1}^{10} = (1/\Delta)\{-2m(m + 1)[(F_1)_{m+1} + (F_3)_{m+1}] + 2m(1 - c_g)[F_2]_m + (m + 1) \times [(F_1)_{m-1} - (F_3)_{m-1}]\}, \tag{3.29b}$$

where

$$\Delta \equiv 1 + (2m + 1)c_g. \tag{3.30}$$

The term $\tilde{v}^{20}$, being of higher order in ϵ, is irrelevant to the calculation of the NLS equation coefficients, so its solution is omitted.

As discussed in Appendix A, it is trivial to calculate the Hermite expansion coefficients of F_1, F_2 and F_3 via Gauss-Hermite quadrature; the resulting series for S^{10} and D^{10} are easily evaluated for any value of y by exploiting the three-term recurrence relationship that can be used to calculate the Hermite functions; one can obtain u^{10} and ϕ^{10} from (3.22). If N is the number of points in the quadrature scheme, then the operation count for calculating the Landau constant (including the steps below) is linearly proportional to N; the error decreases exponentially with N and $N = 32$ is sufficient to give four or five decimal places. Consequently, the Gauss-Hermite quadrature combined with the techniques used with the sum/difference variables and raising/lowering operators is an extremely accurate and efficient way of calculating the Landau constant.

Before turning to the second harmonics, however, we must note that the solution for a particular long wave is infinite whenever $\Delta = 0$, i.e.,

$$c_g = \frac{-1}{2m + 1}. \tag{3.31}$$

The right-hand side of (3.31) is the phase (and group) velocity of the ultra-long Rossby wave of meridional mode number m. Physically, Eq. (3.31) is thus a condition for resonance between the primary wave packet and one of the free, ultra-long Rossby waves which are the homogeneous solutions of (3.20). The conditions on k and the packet mode number that must be satisfied so that (3.31) is satisfied are discussed in the next section and also in Boyd (1983a).

The treatment of the second harmonics is very similar in spirit and content to that of the long waves. The perturbation equations are

$$-2ikc_p u^{12} - yv^{12} + 2ik\phi^{12} = \bar{F}_1, \tag{3.32a}$$

$$-2ikc_p v^{12} + yu^{12} + \phi_y^{12} = \bar{F}_2, \tag{3.32b}$$

$$-2ikc_p \phi^{12} + v_y^{12} + 2iku^{12} = \bar{F}_3, \tag{3.32c}$$

where the forcings are

$$\bar{F}_1 = -iku^{01}u^{01} - v^{01}u_y^{01}, \tag{3.33a}$$

$$\bar{F}_2 = -iku^{01}v^{01} - v^{01}v_y^{01}, \quad (3.33b)$$

$$\bar{F}_3 = -2iku^{01}\phi^{01} - (v^{01}\phi^{01})_y. \quad (3.33c)$$

Here, the convention is used that expressions like $v^{01}v_y^{01}$ denote that portion of $v^{01}v_y^{01}$ which is proportional to A_{01}^2, as opposed to $|A_{01}|^2$ or A_{01}^{*2} where the asterisk denotes the complex conjugate. Defining "sum" and "difference" variables as before

$$S^{12} = \phi^{12} + u^{12}, \quad (3.34a)$$

$$D^{12} = \phi^{12} - u^{12}, \quad (3.34b)$$

substituting Hermite expansions for $\bar{F}_1$, $\bar{F}_2$ and $\bar{F}_3$ and for S^{12}, v^{12} and D^{12} into (3.32) and exploiting the properties of the "raising" and "lowering" operators, one obtains

$$S_0^{12} = [(\bar{F}_1)_0 + (\bar{F}_3)_0]/[2k(1 - c_p)] \quad (3.35)$$

plus, for $n \geq 1$, the uncoupled triplets

$$\begin{vmatrix} -2ik(c_p - 1) & -1 & 0 \\ (m+1) & -2ikc_p & -\tfrac{1}{2} \\ 0 & 2m & -2ik(c_p+1) \end{vmatrix} \begin{vmatrix} S_{m+1}^{12} \\ v_m^{12} \\ D_{m-1}^{12} \end{vmatrix} = \begin{vmatrix} (\bar{F}_1)_{m+1} + (\bar{F}_3)_{m+1} \\ (\bar{F}_2)_m \\ (\bar{F}_3)_{m-1} - (\bar{F}_1)_{m-1} \end{vmatrix}, \quad (3.36)$$

with the solution

$$S_{m+1}^{12} = 1/[4k\Delta])\{i[8k^2c_p^2 + 8k^2c_p - 2m] \times [(\bar{F}_1)_{m+1} + (\bar{F}_3)_{m+1}] - 4k(c_p + 1) \times (\bar{F}_2)_m + i[(\bar{F}_1)_{m-1} - (\bar{F}_3)_{m-1}]\}, \quad (3.37a)$$

$$v_m^{12} = (\tfrac{1}{2}\Delta)\{(2m + 2)(c_p + 1)[(\bar{F}_1)_{m+1} + (\bar{F}_3)_{m+1}] + 4ik(c_p^2 - 1)(\bar{F}_2)_m + (c_p - 1) \times [(\bar{F}_1)_{m-1} - (\bar{F}_3)_{m-1}]\}, \quad (3.37b)$$

$$D_{m-1}^{12} = (1/[2k\Delta])\{-2mi(m + 1)[(\bar{F}_1)_{m+1} + (\bar{F}_3)_{m+1}] + 4km(c_p - 1)(\bar{F}_2) - i[4k^2c_p^2 - 4k^2c_p - m - 1][(\bar{F}_1)_{m-1} - (\bar{F}_3)_{m-1}]\}, \quad (3.37c)$$

where

$$\Delta \equiv 4k^2c_p^3 - 4k^2c_p - 2mc_p - c_p - 1. \quad (3.38)$$

Again, resonance—in this case, resonance with a second harmonic—occurs whenever $\Delta = 0$. Recalling that $kc_p = w(k; n)$, one finds that introducing

$$W = 2w(k; n) \quad (3.39)$$

and

$$K = 2k \quad (3.40)$$

gives

$$\Delta = [W^3 - W([2m + 1] + K^2) - K]/2k, \quad (3.41)$$

which is merely the dispersion relation for equatorial waves. In words, Eq. (3.41) states that second harmonic resonance occurs whenever one of the forced second harmonics, which for all n have twice the frequency and wavenumber of the fundamental, is also a free oscillation of the linear shallow water wave equations. The conditions for this to occur and its implications are discussed in the next section and in Boyd (1983b).

Now that the solution is fully determined to $O(\epsilon^2)$, the worst is over. The coefficients of the Nonlinear Schrödinger equation are determined at $O(\epsilon^3 E)$, but as is usual in the method of multiple scales, it is not necessary to explicitly solve this set; the solvability condition is sufficient. The equations are

$O(\epsilon^3 E^1)$:

$$-iwu^{21} - yv^{21} + ik\phi^{21} = \tilde{F}_1, \quad (3.42a)$$

$$-iwv^{21} + yu^{21} + \phi_y^{21} = \tilde{F}_2, \quad (3.42b)$$

$$-iw\phi^{21} + iku^{21} + v_y^{21} = \tilde{F}_3, \quad (3.42c)$$

where

$$\tilde{F}_1 = [-u_\tau^{01} + c_g u_\xi^{11} - \phi_\xi^{11}] - iku^{10}u^{01} - iku^{12}u^{01} - v^{01}u_y^{10} - v^{12}u_y^{01} - v^{01}u^{12}y, \quad (3.43a)$$

$$\tilde{F}_2 = [-v_\tau^{01} + c_g v_\xi^{11}] - iku^{10}v^{01} - iku^{12}v^{01} - 2iku^{01}v^{12} - v^{01}v_y^{12} - v^{12}v_y^{01}, \quad (3.43b)$$

$$\tilde{F}_3 = [-\phi_\tau^{01} + c_g\phi_\xi^{11} - u_\xi^{11}] - iku^{01}\phi^{10} - iku^{10}\phi^{01} - iku^{01}\phi^{12} - iku^{12}\phi^{01} - (v^{01}\phi^{10})_y - (v^{01}\phi^{12})_y - (v^{12}\phi^{01})_y. \quad (3.43c)$$

The solvability or non-secularity condition is that the column vector of forcing functions should be orthogonal to the homogeneous solution of (3.42), which is simply the lowest order solution (u^{01}, v^{01}, ϕ^{01}). Mathematically, the vector inner product which must vanish is

$$\int_{-\infty}^{\infty} dy(u^{01} * \tilde{F}_1 + v^{01} * \tilde{F}_2 + \phi^{01} * \tilde{F}_3) = 0, \quad (3.44)$$

where the asterisk denotes the complex conjugate. After performing the integrations in y via Gauss–Hermite quadrature, using Hermite function series to evaluate u^{01}, v^{01}, ϕ^{01} and the terms in $\tilde{F}_1$, $\tilde{F}_2$, and $\tilde{F}_3$, one obtains

$$i(A_{01})_\tau + \tfrac{1}{2}w''(A_{01})_{\xi\xi} + \nu|A_{01}|^2A_{01} = 0, \quad (3.45)$$

where $w'' \equiv d^2w/dk^2$ and $\nu(k)$ is the "Landau constant." The linear terms in (3.45) come directly from the linear terms in the square brackets in (3.43).

Domaracki and Loesch (1977) give an explicit formula for triad resonance coefficients which is extremely complicated (129 additive terms!). Here we must go to third order rather than second order and must explicitly solve the second-order equations in the form of infinite series, so an expression similar to theirs for our $\nu(k)$ from first principles would be many times more difficult. It follows that we must be content with computing the "Landau constant" numerically although polynomial curve-fitting to the numerical results for each mode is always possible.

Because of the resonances, unfortunately, Eq. (3.45) is not a valid model everywhere in parameter space even when $\epsilon \ll 1$, so these resonances form the topic of the next section. In later portions of the paper, we shall present graphs of $w''(k)$ and $\nu(k)$ for various modes and discuss the properties of solutions of the Nonlinear Schrödinger equation.

4. Resonances and poles of the Landau constant

When either a long-wave mode or a second-harmonic mode is resonant with the fundamental, that mode becomes infinite in amplitude and the Landau constant $\nu(k)$ has a simple pole at that value of k for which the resonance occurs. Mathematically, such infinities destroy the rationale for the perturbation theory, which requires that all the waves forced by the self-interaction of the fundamental should remain small in comparison to the fundamental. Physically, such infinities imply that the NLS equation is no longer an adequate model because the mutual interaction of the two resonant waves, which may permit the long wave or second harmonic to feed on the energy of the fundamental and grow until it is the same order of magnitude as the fundamental, is now more important than the self-interaction of a single mode with itself which the NLS equation describes. To cope with this, it is necessary to replace the one-equation NLS model with a *coupled set of two equations* in the neighborhood of these resonant values of k.

For Rossby waves, however, a third type of resonance is possible as $k \rightarrow 0$. In this limit, Rossby waves are non-dispersive. This implies that for small k, both the long-wave and second-harmonic Rossby waves of mode number $m = n$, where n is that of the fundamental, will have approximately the same phase and group velocities as the fundamental. However, the forced second harmonic and long waves are always symmetric with respect to the equator (m is always an *odd* integer), so this simultaneous long-wave-and-second-harmonic resonance can only occur for symmetric, ultra-long Rossby waves. Again, it is necessary to replace the NLS equation by something else. In the limit $k \rightarrow 0$, however, it is no longer necessary to make a distinction between the "fast" length scale $1/k$ and the "slow" length scale of $A_{01}(\zeta, \tau)$ or even between the long wave, the fundamental and the second harmonic. Instead, all the zonal dependence of the resonant wave components (which all have the same mode number for this type of resonance, remember) can be lumped into a single slowly-varying amplitude function, and the proper replacement equation is the Korteweg–deVries equation of Boyd (1980b).

Thus, a discussion of wave packets involves not merely the Nonlinear Schrödinger equation but rather a total of four equations or sets of equations, the other three being

$$A_\tau + A_{\zeta\zeta\zeta} + AA_\zeta = 0 \qquad \text{[Korteweg–deVries equation]}, \quad (4.1)$$

$$\left.\begin{aligned} iA_\tau + \tfrac{1}{2}w''A_{\zeta\zeta} + \nu_L AB &= 0 \\ B_\tau &= \mu(|A|^2)_\zeta \end{aligned}\right\} \qquad \begin{bmatrix}\text{long wave/short wave} \\ \text{resonance equations}\end{bmatrix}, \quad (4.2)$$

where $A(\zeta, \tau)$ is the envelope of the short wave packet and $B(\zeta, \tau)$ is the amplitude of the long Rossby wave with ν_L and μ as constants and $w''(k) \equiv d^2w/dk^2$ and

$$\left.\begin{aligned} A_\tau &= IAH \\ H_\tau + (c_g^H - c_g)H_\zeta &= -IA^2 \end{aligned}\right\} \qquad \begin{bmatrix}\text{resonant dyad equations} \\ \text{second-harmonic resonance}\end{bmatrix}, \quad (4.3)$$

where H is the envelope of the second harmonic wave packet, I the interaction coefficient, a constant, and c_g^H the group velocity of the second harmonic. All three equation sets are written in a frame of reference moving with the linear group velocity of the fundamental and each is discussed in detail in Boyd (1980b), Boyd (1983a) and Boyd (1983b), respectively (see also Ripa (1983a,b).

Since the calculations of the coefficients in (4.1)–(4.3) involve exactly the same sort of operations on exactly the same nonlinear forcing terms, it is hardly surprising that there exists simple relationships between the coefficients of (4.1)–(4.3) and the poles of the Landau constant of the NLS equation, to wit

$$\nu(k) \approx \frac{a^2}{6bk}\begin{bmatrix}\text{symmetric Rossby} \\ \text{waves, } k \ll 1\end{bmatrix}, \quad (4.4)$$

$$\nu(k) \approx \frac{-\nu_L\mu}{w''(k - k_{\text{res}})} \text{ [long wave resonance]}, \quad (4.5)$$

$$\nu(k) \approx \frac{I^2}{2[c_g^H(2k; m) - c_g(k; n)](k - k_{\text{res}})} \qquad \text{[second harmonic resonance]}, \quad (4.6)$$

where k_{res} is the resonant value of k. Eqs. (4.5) and

(4.6) follow from careful comparison of the NLS equation derivation given in Section 3 with the corresponding derivations of (4.2) and (4.3) in Boyd (1983a,b). Since one can estimate the residues of the poles simply by calculating $\nu(k)$ near the resonances, Eqs. (4.5) and (4.6) are listed only for completeness and their derivation will be omitted. Eq. (4.4), however, can be directly obtained from the corresponding Korteweg–deVries equation. Since this illustrates the steps of Section 3 while bypassing the Hermite function tricks that make Section 3 so lengthy, the proof of (4.4) is given in Appendix C.

For reference, Table 1 gives the coefficients of the Korteweg–deVries (and modified Korteweg–deVries) equations for the first 20 Rossby waves. This is a much larger number of modes than in Boyd (1980b), and for conformity with this latter work, the coefficients are given in terms of the "energy" normalization explained in Section 3. The interaction coefficients and resonant values of k for the other two types of resonances are given in Boyd (1983a,b).

These resonances have a multiple significance. First, they chop up parameter space into "forbidden zones" where the NLS equation is useless which are surrounded by regions where it is an accurate model. Second, the resonances separate regions in which $\nu(k)$ is of opposite sign as is obvious from (4.5) and (4.6). This is crucial because dispersion may either oppose linear dispersion, creating modulational or sideband instability which leads ultimately to envelope solitons, or augment dispersion to create "superlinear dispersion" where all wave packets break up quickly and neither sideband instability nor solitary waves is possible—all depending on the sign of $\nu(k)$. Thus, the resonances also separate regions of instability and solitons from regions of "superlinear dispersion" and sideband stability. Third, the resonant sets have both instabilities and solitary waves of their own, notably the paired envelope soliton/unimontane soliton of the Long Wave/Short Wave Resonance Equations. The overall effect is to make the nonlinear theory of dispersive equatorial waves rather complicated.

Fortunately, there are some limits on the chaos. Formulas for resonances are derived and fully discussed in Boyd (1983a,b); here, it will suffice to give only a broad outline of what is shown there.

First, there are no resonances of any kind for eastward travelling gravity waves. Thus, the graphs of $\nu(k)$ that will be presented in the next section are very smooth for these modes. The mixed Rossby–gravity wave has but a single second harmonic resonance at $k = (1/2)\sqrt{1/3}$, so its Landau constant, too, has simple behavior.

Alas, for the other westward travelling waves, the situation is confusing. The second harmonic resonances are finite in number—one for each gravity wave and $[(n - 1)/2]$ for each Rossby mode where n is the mode number and the brackets denote the "integral part of"—but the long-wave resonances are infinite in number although the resonant values of k are confined within a finite interval of moderate wavenumber. It can be shown, however, that the interaction coefficients for resonance with a long Rossby wave of mode number m decrease exponentially fast with m. Thus, if one graphs $\nu(k)$ with a finite grid spacing in k, only a handful of resonances will be detectable on the plot even with rather fine spacing although more will appear as the graphical resolution is increased. In spite of this compensating decrease of the residues of the poles associated with high mode resonances and the restriction of the resonances to a modest, fixed interval of wavenumber, the plots of $\nu(k)$ for all modes with westward group velocities are quite complicated.

TABLE 1. Coefficients of the Korteweg–deVries (odd mode number) or modified Korteweg–de Vries equation (even mode number) for the lowest 20 Rossby modes. The nonlinear coefficients given here differ from the more limited results given in Boyd (1980b) because the modes are "energy-normalized" here as explained in Section 3. To facilitate comparisons with the unnormalized results, S_2, the normalized multiplier of the north–south current v, as defined by Eq. (3.15b), is also tabulated. The Korteweg–deVries equation is $u_t + auu_x + bu_{xxx} = 0$ while the modified Korteweg–deVries equation is $u_t + au^2u_x + bu_{xxx} = 0$. The nondimensional length and time scales are those defined by (3.4) and (3.5). The notation "$E-n$" denotes scientific notation, i.e., multiplication by 10^{-n}.

n	S_2	a Nonlinear	b Dispersion
1	0.289	−0.444	−0.0988
3	0.0406	−0.118	−0.0200
5	0.00364	−0.0607	−0.00820
7	0.000241	−0.0387	−0.00442
9	0.126E−4	−0.0275	−0.00276
11	0.547E−6	−0.0209	−0.00189
13	0.202E−7	−0.0166	−0.00137
15	0.651E−9	−0.0136	−0.00104
17	0.186E−10	−0.0115	−0.000816
19	0.476E−12	−0.00982	−0.000657
	(Modified KDV equation below)		
2	0.116	0.144	−0.0384
4	0.0127	0.0925	−0.0122
6	0.000968	0.0698	−0.00588
8	0.566E−4	0.0486	−0.00345
10	0.269E−5	0.0543	−0.00226
12	0.107E−6	0.0488	−0.00160
14	0.369E−8	0.0442	−0.00119
16	0.112E−9	0.0384	−0.000917
18	0.301E−11	0.0352	−0.000730
20	0.734E−13	0.0367	−0.000595

5. The Landau constant

Fig. 3 shows the dispersion coefficient $\frac{1}{2}w''$ and the Landau constant ν as a function of k for the $n = 1$ westerly (that is, propagating toward the east) gravity wave. Since the eastward-travelling gravity waves

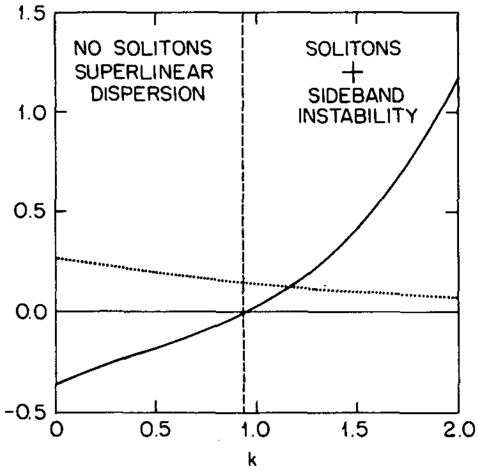

FIG. 3. The Landau constant $\nu(k)$ (solid curve) and dispersion coefficient $[(1/2)w''(k)]$ (dotted curve) are plotted versus wavenumber for the $n = 1$ gravity wave propagating toward the east ($w > 0$).

have *no resonances* (including $n = 0$), all these modes have graphs qualitatively like Fig. 3. For small k, there is "superlinear" dispersion because the dispersion coefficient and Landau constant have opposite signs. Envelope solitons with central packet wavenumbers that are small are not possible, but plane waves with such wavenumbers are stable against sideband instability. For large k, the dispersion coefficient and Landau constant have the same sign, so eastward-travelling gravity wave envelope solitons can be excited, and sideband instability will disrupt plane waves. The Landau constant increases without bound for large k [roughly as $O(k^5)$ from empirical curve-fitting], but the dispersion coefficient is asymptotic to zero from above. This implies that very short gravity waves will be much more strongly affected by nonlinearity than dispersion unless their amplitude is extremely small. The borderline between non-soliton/soliton regions is given by $k = 0.065$ ($n = 0$), 0.942 ($n = 1$), 1.53 ($n = 2$), 2.05 ($n = 3$), and so on for the higher modes.

The graphs of the Landau constant and dispersion coefficient for the $n = 0$ mixed Rossby–gravity wave shown in Fig. 4 are nearly as simple as those for the eastward-travelling gravity waves. Although the mixed Rossby–gravity propagates toward the west, its *group* velocity is always toward the east so that long-wave/short-wave resonance with long Rossby waves, whose group velocity must always be in the opposite direction, is impossible. The mixed Rossby–gravity wave has only a single resonance, a second harmonic resonance at $k = 0.288$, but the Landau constant goes through a nonsingular change of sign at $k = 0.624$ so that parameter space is chopped up into three regions with solitons and sideband instability only in the intermediate region $0.288 < k < 0.624$. For large k, the Landau constant is again positive and increasing without bound (but with a linear slope): $\nu(k) \approx -0.57 + 0.81k$ for $k \gg 1$ and the dispersion coefficient is again asymptotic to zero, but the mixed Rossby-gravity dispersion coefficient is negative in contrast to that for eastward-travelling gravity waves. A more striking difference is that except in the neighborhood of the resonance, the magnitude of the Landau constant is very small for $k < 1$; for $k = 1$, $\nu = 0.07$, which is an order of magnitude smaller than for the $n = 0$ gravity wave at this same value of k.

The behavior of gravity waves propagating toward the west is unfortunately considerably more complicated, at least in theory. The dispersion coefficient, shown in Fig. 5 for the $n = 1$ mode (dotted line), is one-signed and negative for all n and is asymptotic to zero from below, just as for the mixed Rossby-gravity wave, which is of course the $n = 0$ mode travelling in the same direction. The Landau constant, however, is very complicated because the gravity waves have an infinite number of long-wave resonances in addition to a single second-harmonic resonance. The general form of the Landau constant is therefore

$$\nu(k) = \frac{R_{SH}}{k - k_{SH}} + \sum_{j=1}^{\infty} \frac{R_j}{k - k_j} + g(k), \quad (5.1)$$

where R_{SH} and k_{SH} are the residue and location of the pole due to the second harmonic resonance, R_j and k_j are the corresponding residue and pole for the jth long-wave resonance, and $g(k)$ is a function analytic everywhere on the real axis. [We shall leave $g(k)$ otherwise unspecified since it can only be computed numerically.] Using the tables of Boyd (1983a,b), one finds that for the $n = 1$ mode, for example,

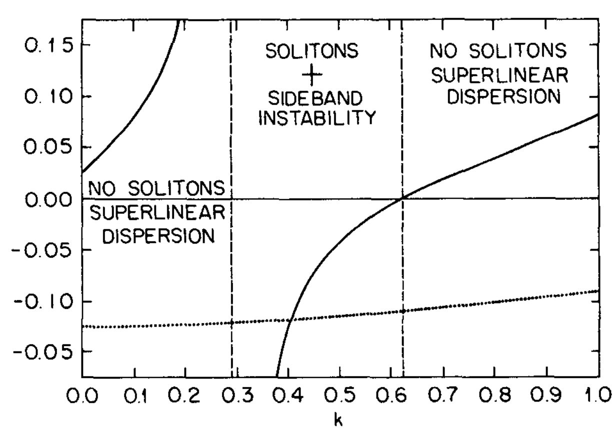

FIG. 4. Landau constant $\nu(k)$ (solid curve) and dispersion coefficient $[(1/2)w'']$ (dotted curve) are plotted versus wavenumber for the $n = 0$ mixed Rossby–gravity wave. The dashed curves mark the boundaries between regions where solitons are possible and where they are not. For larger k (off the graph), $\nu(k)$ is approximately linear while the dispersion coefficient is asymptotic to zero from below.

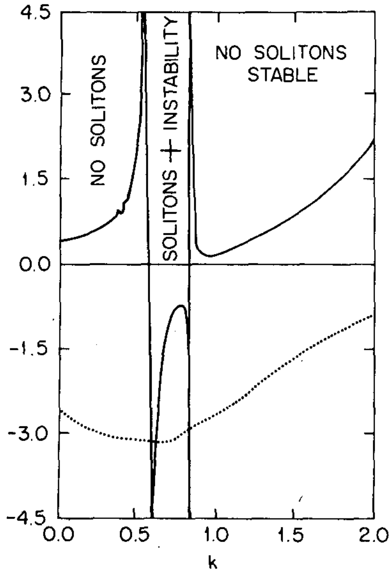

FIG. 5. Gravity wave $n = 1$ propagating toward the west ($w < 0$): Landau constant $\nu(k)$ (solid) and dispersion coefficient (dashed). For clarity, the dispersion coefficient was multiplied by 10. The Landau constant increases very rapidly for large k while the dispersion coefficient is asymptotic to zero from below.

$$\nu_1(k) = \frac{-0.146}{k - 0.577} + \left\{\frac{0.0168}{k - 0.834} + \frac{0.00154}{k - 0.524} + \frac{0.0000586}{k - 0.441} + \cdots\right\} + g_1(k), \quad (5.2)$$

where the braces enclose the infinite sum of long-wave resonances. The most striking features of (5.2) are (i) the residues of the long-wave resonances decrease very rapidly with j and (ii) the residue of the second harmonic pole is more than eight times larger than that of the gravest long wave resonance and nearly 100 times the size of the second long-wave resonance. This implies that in a practical sense, the situation for the $n = 1$ and higher gravity modes is not much different from that for the $n = 0$ mixed Rossby–gravity wave: the behavior of the Landau constant is dominated by the single pole associated with second-harmonic resonance. This is borne out by Fig. 5, which graphs the Landau constant for $n = 1$ (solid curve). Despite the rather fine graphing scale (0.02 in k), only the first long-wave resonance (at $k = 0.834$) is visible on the figure and the width of the region in k, where it is important, is very narrow: the first long-wave pole in (5.2) is greater than the second harmonic pole in (5.2) only when $|k - 0.834| \leq 0.03$. In theory, the higher long-wave resonances would show themselves (in *very* narrow regions of wavenumber) if the graphing interval would be made very, very fine. In a practical sense, however, these resonances would be almost unobservable even in an idealized inviscid fluid. In the real, dissipative ocean, these higher long-wave resonances would be invisible.

In a practical sense, therefore, the parameter space for the $n = 1$ gravity may be divided into three regions with solitons and sideband instability occurring only in the middle region $0.577 < k < 0.834$. The only difference from the $n = 0$ mixed Rossby–gravity wave is that the upper limit of the soliton region is now given by the location of the strongest long-wave resonance instead of being a non-singular change of sign in $\nu(k)$. Behavior for higher order gravity waves is similar: only the second-harmonic resonance and perhaps the lowest long-wave resonance are important and there are no solitons for large k. The Landau constant increases very rapidly with k for $k \gg 1$ (roughly as $O(k^4)$ or $O(k^5)$) just as for gravity waves travelling in the opposite direction.

Rossby waves also have an infinite number of long-wave resonances, but unfortunately the residues are much larger than for the westward-travelling gravity waves, presumably because the resonances are between waves of the same type (short Rossby with long Rossby) instead of different types (packets of short gravity waves with long Rossby waves). The Landau constants for the Rossby modes are of the form

$$\nu_n(k) = \frac{R_{\mathrm{KDV}}}{k} + \sum_{l=1}^{[n/2]} \frac{R_l^{\mathrm{SH}}}{k - k_l^{\mathrm{SH}}} + \sum_{j=1}^{\infty} \frac{R_j}{(k - k_j)} + g_n(k), \quad (5.3)$$

where $[n/2]$ denotes the smallest integer greater than or equal to half the mode number n and is the number of second-harmonic resonances, and where the infinite sum gives the poles and residues of the long-wave resonances with $g_n(k)$ a function analytic on the real k axis. By explicit computation

$$R_l^{\mathrm{SH}} \leq 0.061, \quad \text{for all } n, l, \quad (5.4)$$

so the second harmonic resonances for Rossby waves generate only poles with rather small residues. At least some of the poles for the lowest long-wave resonances are larger than the bound in (5.4) for all the modes examined. There is also an additional pole (for odd n only) at $k = 0$, associated with the regime described by the Korteweg–deVries (KDV) equation, which usually has a large residue, too. (This residue R_{KDV} can be calculated by combining (4.4) with the KDV coefficients in Table 1).

For the lowest, $n = 1$ Rossby wave one finds

$$\nu_1(k) = \frac{-0.333}{k} + \left\{\frac{-0.077}{(k - 0.949)} - \frac{0.057}{(k - 1.155)} - \frac{0.0236}{(k - 1.266)} - \frac{0.00676}{(k - 1.338)} - \cdots\right\} + g_1(k). \quad (5.5)$$

There are no second harmonic resonances for this

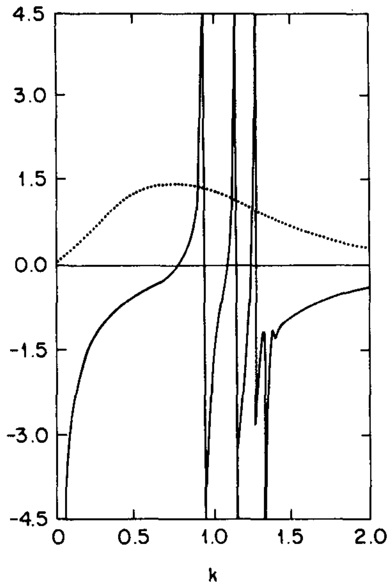

FIG. 6. Rossby wave $n = 1$: Landau constant $\nu(k)$ and dispersion coefficient (the dispersion coefficient is magnified by a factor of 10). The Landau constant changes sign for large k and increases linearly. The dispersion coefficient also changes sign for large k and is asymptotic to zero.

mode; instead, the KDV pole at the origin is dominant. The long-wave poles in the braces have residues that decrease with j, but not nearly as fast as in the corresponding infinite series of poles for the gravity waves. The result is that all five of the poles shown explicitly in (5.4) produce clearly visible spikes in Fig. 6. Parameter space for the Rossby wave is therefore chopped up into many narrow regions in which solitons are possible, alternating with regions where solitons are impossible, and plane waves are stable—a confusing situation.

Nonetheless, a few general statements are possible. First, the long-wave resonances, though infinite in number, are invariably confined within a finite interval in k—$k \epsilon [0.949, 1.732]$ for $n = 1$, for example. Second, just as for the westward travelling gravity waves, only a handful of poles have residues large enough to be of practical significance, roughly five poles for $n = 1$. Third, the Landau constant asymptotes to a *linear* slope for $k \gg 1$ for Rossby waves: $\nu(k) \approx -2.27 + 1.44k$ for $n = 1$. Fourth, the dispersion coefficients are positive for small k but negative and small (asymptotic to 0 from below as $k \rightarrow \infty$) for larger k. There is only a single change of sign for w'' which occurs at $k = 2.94$ (off the graph in Fig. 6) for $n = 1$ and larger k for larger n. For all Rossby waves (and indeed all waves travelling toward the west), solitons and sideband instability are impossible for large k. Fifth, envelope solitons are also impossible for Rossby wave packets of small k, whether the mode number is even or odd.

6. Solutions of the Nonlinear Schrödinger equation

a. Introduction and reduction to canonical form

The oceans are bounded by the continents, but the Nonlinear Schrödinger equation has been previously studied only for (i) an unbounded domain with a localized initial condition or (ii) periodic boundary conditions. However, the NLS equation is not an appropriate model for equatorial waves reflecting off the continents because it describes the evolution of only a single mode whereas the classic thesis of Moore (1968) showed that equatorial waves reflect as a sum, usually an infinite sum, of different modes. Upon reflection, for example, a Kelvin wave becomes a series of Rossby modes plus coastal Kelvin waves, while a Rossby wave reflects part of its energy as a Kelvin wave plus gravity waves plus very short Rossby waves. In consequence, one has no recourse but to apply the NLS equation away from the continental boundaries and ignore the wave reflections that occur there.

A useful preliminary step, before discussing the spatially localized and spatially periodic solutions, is to reduce the NLS equation from the general form

$$i(A_{01})_\tau + \tfrac{1}{2}w''(A_{01})_{\zeta\zeta} + \nu|A_{01}|^2A_{01} = 0 \quad (6.1)$$

to the canonical form

$$iA_t + A_{xx} \pm 2|A|^2A = 0, \quad (6.2)$$

via the change of variables

$$\tau = t\,\mathrm{sgn}[w''], \quad (6.3)$$

$$\zeta = x(|w''|/2)^{1/2}, \quad (6.4)$$

$$A_{01} = A\nu^{-1/2}. \quad (6.5)$$

Note that the plus sign occurs in (6.2) if w'' and ν have the same sign while the minus sign is necessary in (6.2) if w'' and ν have opposite signs. Short of using imaginary coordinates, it is impossible to convert these two cases, represented by the different signs in (6.2) into one another; as explained below, this is because the two cases are physically different as shown in Fig. 7.

Please bear in mind that in this section, x and t are defined by (6.3) and (6.4) and are not the nondimensional space and time scales given in Section 3.

b. The linear solution

The starting point for a discussion of the nonlinear solutions of (6.2) is the solution of its linearized form

$$iA_t + A_{xx} = 0. \quad (6.6)$$

Defining $\tilde{A}(k)$ to be the Fourier transform of the initial conditions, i.e.,

$$\tilde{A}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(x, 0)e^{-ikx}dx, \quad (6.7)$$

the exact solution of (6.6) is given by the Fourier integral

$$A(x, t) = \int_{-\infty}^{\infty} \tilde{A}(k)e^{ikx-ik^2t}dx, \quad (6.8)$$

which for large values of the time is given by the asymptotic, steepest descent approximation

$$A(x, t) \sim \tilde{A}(x/[2t])\sqrt{\pi/t}\exp[ix^2/(4t) - i\pi/4]. \quad (6.9)$$

Two features of (6.9), both general to dispersive waves, are that the envelope $\tilde{A}[x(2t)^{-1}](\pi/t)^{1/2}$ is increasing in width linearly with time while its magnitude is decreasing as the inverse square root of time. In the limit $t \rightarrow \infty$, the wavetrain has dispersed completely so that there is no detectable wave disturbance at all.

c. *Localized initial conditions and solitons*

Zakharov and Shabat (1972) showed that when the initial conditions were such that the pulse was localized in some finite region of space, decaying exponentially for large positive and negative x, the Nonlinear Schrödinger equation could be formally solved through a sequence of linear steps via the so-called "inverse scattering method." This procedure is a generalization of the Fourier transform applied to the linearized equation in the previous subsection, but the details are complicated so that only the qualitative conclusions will be described here.

When the signs of the dispersive (second spatial derivative) and nonlinear terms in the NLS equation are opposite—the minus sign in (6.2)—the asymptotic solution is simply a wavetrain qualitatively similar to the linear solution given by (6.9). The slight nonlinear modifications to (6.9) have been calculated by Segur and Ablowitz (1976), Segur (1976) and Manakov (1974), but the most striking difference is that the nonlinearity also acts to widen the wave packet, giving "defocusing" or "superlinear" dispersion in the sense that wavetrain spreads out more rapidly than in a linearized theory.

When the dispersive and nonlinear terms in the NLS equation are of the same sign—the positive sign in (6.2)—then the asymptotic solution to the NLS equation consists of a wavetrain again qualitatively similar to (6.9) plus a finite number (possibly zero) of envelope solitary waves. The isolated single solitons are of the form

$$A(x, t) = ae^{ia^2t}\operatorname{sech} ax, \quad (6.10)$$

which form a one-parameter family with a, the amplitude of the soliton, as the parameter. Please bear in mind that (6.10) is merely the *envelope* of the soliton; to obtain the actual variation of the currents and surface height of the water with x and t, one must multiply $A(x, t)$ as given by (6.10) by the factor $\exp[ikx - iw(k)t]$ to obtain a shape qualitatively like

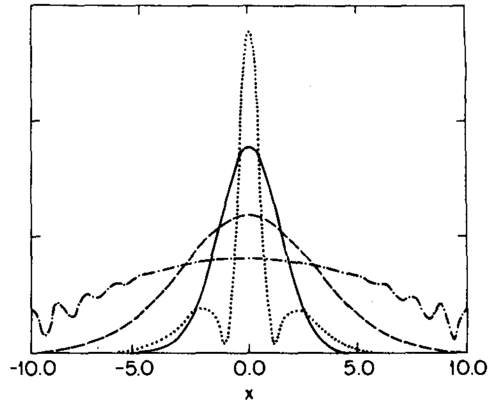

FIG. 7. A comparison of the integration of the Nonlinear Schrödinger equation for the same initial condition (solid curve) but three different values of the Landau constant. Dashed curve: $\nu = 0$ (linear), $t = 2.0$; dot-dashed curve: $\nu = -2$ ("superlinear" dispersion or "defocusing"), $t = 2.0$; dotted curve: $\nu = 2$ (soliton forming, "focusing"), $t = 0.84$.

that shown by the solid line in Fig. 1b. Eq. (6.10) by itself merely describes the shape of the dotted line in this same figure. One can generalize (6.10) by using the general Galilean invariance theorem which states that if $A(x, t)$ is a solution of the NLS equation—any solution, not necessarily a soliton—then $\tilde{A}(x, t) = \exp[i(V/2)x - i(V^2/4)t]A(x - Vt, t)$ is also a solution of the Nonlinear Schrödinger equation for any constant V. For the special case of the single stationary soliton [(6.10)], this gives the two-parameter family of travelling single solitons

$$A(x, t) = a\exp[i(V/2)x - i(V^2/4)t + ia^2t] \times \operatorname{sech}[a(x - Vt)]. \quad (6.11)$$

One can of course generalize (6.10) still further by adding constant phase factors to the spatial dependence.

In physical terms, the parameter V in (6.11) corresponds to making a slight alteration in the central wavenumber of the packet and then computing the resulting changes in the phase speed and group velocity via the perturbative approximations inherent in the derivation of the NLS equation rather than from the original dispersion relation $w(k)$. In considering a single soliton by itself, Eq. (6.11) is rather pointless because its physical content is the same as that of the one-parameter family given in Eq. (6.10). However, an initial disturbance of sufficiently large amplitude may evolve into two or more solitons with different speeds. When these have become sufficiently separated, each can be described by a member of the two-parameter family [(6.11)] with different values for V and a.

Analytic formulas for the multi-solitons when they are still interacting with one another were derived by Zakharov and Shabat (1972), but are too complicated

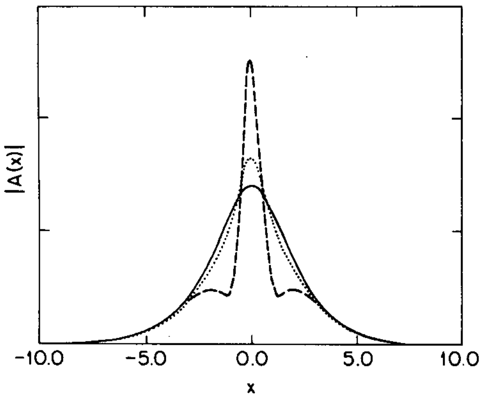

FIG. 8. The absolute value of the envelope for a typical "breather" double soliton or "bion" of the Nonlinear Schrödinger equation. The bion is periodic in time and pulsates between the two extreme shapes shown.

to write down here. It is perfectly possible, however, for two or more solitons to have precisely the same speed, forming a bound state known as a "breather" soliton. Although this equality of soliton velocities is impossible for many other types of solitary waves such as those of the KDV equation in Boyd (1980b), for the Nonlinear Schrödinger equation a sufficient condition is that $A(x, t = 0)$ is real.[1] Fig. 8 shows the time development of a double breather or "bion" as numerically calculated by the author. The bion is periodic in time and pulsates between a single broad peak and a taller, narrower peak flanked by two smaller, broader crests. Satsuma and Yajima (1974) give a similar graph for a triple breather as well and write down the analytic expression for a bion which will be omitted here.

Unlike the Korteweg–deVries equation where the wavetrain and the solitons always rapidly separate in space—the former drifts to the left while the solitons travel steadily to the right—the Nonlinear Schrödinger equation wavetrain may remain centered on the soliton. However, for either equation, the wavetrain decays algebraically with time. Kaup (1977) shows graphically (his Fig. 3) that the principle of "soliton dominance" discussed in Boyd (1980b) also holds for the NLS equation: a smooth initial condition of large enough amplitude to give one or more solitons will place most of the energy into the solitons, leaving very little for the decaying wavetrain. Satsuma and Yajima (1974) are able to make this notion qualitative for the special case of a hyperbolic secant initial condition: if N is the number of solitons, then the maximum fraction of energy which goes into the wavetrain is bounded by

$$0.25/(N + \tfrac{1}{2})^2 \leqslant \tfrac{1}{9}, \quad \text{for all } N.$$

[1] Note that $A(x, t)$ is usually complex since the physical variables are obtained by taking the real part of the complex solution.

Consequently, the decaying wavetrain will not mask the solitons very much unless the initial condition has a lot of small-scale structure.

The soliton amplitudes and speeds can be obtained from the initial condition by solving a non-self-adjoint eigenvalue problem similar to the one-dimensional Schrödinger equation of quantum mechanics with the initial condition playing the role of the potential energy. The number of solitons is equal to the number of bound states of the potential/initial condition and the eigenvalues, which may be complex because the potential problem is not self-adjoint, give both the amplitude and speed of each soliton. Obtaining the soliton phases and information about the wavetrain is considerably more difficult and requires performing additional steps in the inverse scattering algorithm. Kaup (1977) presents graphs of the needed eigenvalues for several real initial conditions and shows that the WKB approximation is very accurate.

The most important conclusion from the eigenvalue problem is the discovery that there is a minimum amplitude necessary to create a soliton. Specifically, Ablowitz *et al.* (1974) show that no solitons will form if

$$\int_{-\infty}^{\infty} |A(x, 0)| dx < 0.904. \tag{6.12}$$

This is in marked contrast to the Korteweg–deVries equation where at least one soliton will form no matter how small the initial disturbance, so long as it is of the right sign. The same authors and also Satsuma and Yajima (1974) present a number of theorems about the solutions and eigenvalues of the NLS equation and the associated Zakharov–Shabat potential problem, but in general the situation is much more confused than for the Korteweg–deVries equation. What is clear, however, from specific cases and numerical integrations is that it seems to be as easy to generate solitons for the Nonlinear Schrödinger equation as for the KDV equation, provided that the initial condition is energetic enough to violate (6.12).

d. Sideband instability

The Nonlinear Schrödinger equation (6.1) has the exact, nonlinear solution

$$A_0 = ae^{i\nu a^2 \tau},$$

where a is a constant, which physically is a plane wave since one must always multiply $A(x, t)$ by $\exp[ikx - iw(k)t]$ to obtain the complete x and t dependence of the waves. When the NLS equation is applied to water waves, the amplitude-dependent correction to the phase speed which is evident in (6.13) is merely that found by Stokes in 1847.

Though exact for the NLS equation, the plane waves described by (6.13) may be unstable in various parameter ranges. When the conditions for triad resonance are approximately satisfied—exact resonance is not necessary if the wave is of sufficiently large amplitude—then the plane wave may be unstable through triad resonance with growth rates that are linearly proportional to a. Even if the conditions for triad resonance are not satisfied—as is always implicit when the NLS equation is used as a model—there is an escape clause. The "sidebands" of the central packet wavenumber, that is to say, waves with wavenumbers only slightly larger or slightly smaller than k, are necessarily in approximate resonance with wavenumber k if the dispersion relation $w(k)$ is smooth. Simply because the sidebands differ only slightly in wavenumber, they must have approximately equal phase and group velocities, and this permits an instability involving only a single meridional mode and a narrow spectrum of wavenumbers. However, as noted in Section 3, the self-interaction of plane wave of wavenumber k leads only to forced oscillations of wavenumbers 0 and $2k$. These in turn interact with the fundamental (or in this case, with the sidebands) to give resonant forcing for the fundamental at third order. Thus, sideband instability, like the NLS equation itself, involves the self-interaction of a single meridional mode at $O(a^3)$ where a, as in (6.13), is the wave amplitude. This is why the NLS equation can describe sideband or "modulational" instability, as it is sometimes alternatively named, but the consequence of this $O(a^3)$ dependence on the wave amplitude is that the growth rate is $O(a^2)$ versus the $O(a)$ growth rates of the resonant triad instability. Thus, when the latter occurs it is usually able to mask sideband instability quite effectively.

Often, however, triad instability does not occur, and sideband instability was in fact first discovered by J. Feir in laboratory experiments with water waves. The theoretical explanation was provided in an accompanying paper (Benjamin and Feir, 1967), and the instability is sometimes referred to as the Benjamin–Feir instability for this reason. Although historically, Benjamin and Feir did not employ the Nonlinear Schrödinger equation for the simple reason it had not yet been applied to water waves, it is much easier to derive the characteristics of the instability from the NLS equation directly than to employ their original procedure.

The linearized stability analysis takes the plane wave A_0 as given by (6.13) and subjects it to small perturbations $A_\pm$ which depend on x as $\exp(\pm i\mu x)$. One can then derive two coupled constant coefficient ordinary differential equations in time to describe these sidebands $A_\pm$. One finds the following:

- Instability is possible if and only if the Landau constant ν and the dispersive coefficient w'' have the *same sign*—which is also the necessary condition for solitary waves.
- The wavelength of maximum instability is

$$\mu_{\max} = a\sqrt{2\nu/w''}. \tag{6.14}$$

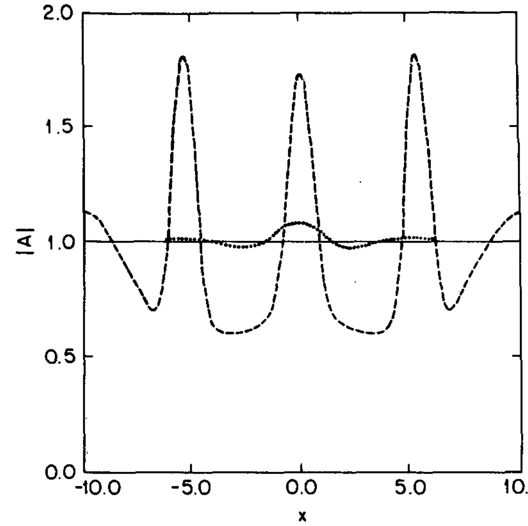

FIG. 9. An example of sideband instability. The Landau constant is $\nu = 2$. The initial condition is the plane wave $|A| = 1$ plus the small Gaussian-shaped perturbation $0.01 \exp(-2\pi x^2/25)$. The periodic boundary condition is $u(x + 20.0) = u(x)$. Solid curve, $t = 0$; dotted curve, $t = 2.0$; dashed curve, $t = 4.0$.

- The maximum growth rate, with $A_\pm$ growing as $e^{\lambda t}$ is

$$\lambda_{\max} = \nu a^2. \tag{6.15}$$

- All sidebands which satisfy the inequality

$$|\mu| < 2a\sqrt{\nu/w''} \tag{6.16}$$

are unstable with growth rates given by

$$\lambda = \frac{1}{2} w''\mu^2\left(\frac{4\nu a^2}{w''\mu^2} - 1\right)^{1/2}. \tag{6.17}$$

Fig. 9 shows the early time development of a typical case of sideband instability. Initially, the curve (note that only $|A|$ is plotted) is completely flat; the Gaussian perturbation is too small to be seen on the scale of the graph. At later times, a peak develops flanked by two shallow troughs. Instead of widening and ultimately dispersing, as in linear theory, these features are "focused" or "pinched in" and exaggerated by the nonlinearity, leading to the ultimate development of three large peaks at the final time step shown. This process of peaks developing from an initially featureless envelope is strongly reminiscent of the development of multiple solitons from a localized initial condition with but a single maximum, and indeed, the same mechanism is responsible for both as evidenced by the fact that soliton formation and sideband instability both occur when—but only when—the Landau constant and dispersion coefficient have the same sign, the so-called "focusing" case.

This suggests that sideband instability does not make the waves dissolve into ever-finer-scale turbulence, cascading forever, but rather that the instability is checked when the peaks develop a sufficiently narrow scale so that dispersion is stronger than nonlinearity. Indeed, Hasimoto and Ono (1972), among others, have speculated that the final end product of sideband instability would be the formation of solitons.

Later work, however, has shown that this is only a half-truth. The sharp narrow peaks shown in the last curve in Fig. 9 do have many of the characteristics of solitons, and if the graph were used to supply the initial condition for an integration of the NLS equation on an unbounded domain, these peaks would simply separate and go their separate ways as solitary waves. On a periodic domain, however, the solitons cannot separate. Even when the peaks have become narrow, there is still a weak interaction between the solitons on one periodicity interval and those on adjacent periodicity intervals. The result, as reviewed by Yuen and Lake (1980), is that the solitons recombine to approximately reproduce the initial plane wave, and this process repeats indefinitely, the so-called Fermi–Pasta–Ulam or FPU recurrence. Yuen and Ferguson (1978) have shown that whether this recurrence pattern is simple and periodic in time or very complicated and only quasi-periodic is dependent upon the perturbation that triggers the instability. If the perturbation is sinusoidal, then the recurrence will be simple if only the original perturbation is unstable according to the Benjamin–Feir theory summarized above. If harmonics of the perturbation, as well as the perturbation itself, are all unstable according to (6.16), then the recurrence will be complex and display several temporal periods. Infeld (1981), in a note only two pages long, has analytically solved the full nonlinear triad of equations for A_0 and $A_\pm$ (neglecting higher harmonics) to explicitly display simple FPU recurrence as an elliptic sine function of time.

e. Polycnoidal waves and theta functions

Like the Korteweg–deVries equation, the NLS equation has exact, periodic solutions which can be expressed in terms of elliptic functions and are known therefore as "cnoidal waves." Hasimoto and Ono (1972) give the cnoidal waves as

$$A(x, t) = a\, dn[a(\nu/w'')x; m]e^{i\alpha t}, \qquad (6.18)$$

where a is a constant, $dn(x; m)$ is the usual elliptic function, and where the elliptic modulus m is given by

$$m = 2 - 2\alpha/(\nu a^2). \qquad (6.19)$$

In physical terms, Eqs. (6.18) and (6.19) describe a spatially periodic modulation of the envelope which propagates at the linear group velocity. In the limit $m \to 1$, the dn function passes over into a hyperbolic secant and (6.18) reduces to the stationary envelope soliton described above. In the opposite limit $m \to 0$, $dn = 1$ and (6.18) reduces to the nonlinear plane wave (6.13).

An analogue of the inverse scattering method for periodic boundary conditions was developed around 1974 by independent groups of workers in the United States, principally P. Lax and H. P. McKean, and in the Soviet Union where S. P. Novikov made the original breakthrough. As reviewed by Dubrovin *et al.* (1976), most of the work to date has concentrated on the Korteweg–deVries equation, but the "Hill's spectrum" method has been extended to the sine–Gordon equation, the Toda lattice, the modified Korteweg–deVries equation, and the Nonlinear Schrödinger equation among others. For each of these equations, there exist generalizations of the cnoidal waves which can be expressed in terms of either N-dimensional hyper-elliptic functions or, more compactly, in terms of N-dimensional Riemann theta-functions. These waves, dubbed N-polycnoidal waves in Boyd (1982), are dense on the set of the spatially periodic solutions of the evolution equation (Korteweg–deVries, Nonlinear Schrödinger, etc.) in the sense that the solution for an arbitrary periodic initial condition can be approximated to an arbitrary degree of accuracy for an arbitrary finite time interval by an N-polycnoidal wave of sufficiently high N. McKean and Trubowitz (1978) prove this by developing the theory of theta functions of "infinite genus" which are the limit of the N-polycnoidal waves as $N \to \infty$.

The 1-polycnoidal wave is merely the ordinary cnoidal wave given by (6.18) and (6.19). The N-polycnoidal wave is a function of the N arguments

$$\eta_j = k_j(x - c_j t) + \phi_j, \quad j = 1, 2, \ldots, N, \qquad (6.20)$$

where the k_j are wavenumbers, the c_j are phase speeds, and the ϕ_j are constant phase factors. In the limit of small amplitude, the N-polycnoidal wave is simply the superposition of N linear plane waves of different wavenumbers and phase speeds. As done by Stokes in 1847, one can calculate nonlinear corrections in this small-amplitude regime using perturbation methods and express the result as a Fourier series. Boyd (1982) gives the Stokes' series for the 2-cnoidal wave for the Korteweg–deVries equation. In the limit of large amplitude, the N-polycnoidal wave consists of N solitons on each periodicity interval and the parameters k_j and c_j give the width and speed of each solitary wave. Besides their Fourier series, the theta functions can also be represented as an infinite sum of Gaussian functions. Boyd (1982) has shown for the Korteweg–deVries equation that the lowest 2^N terms of the Gaussian series generate the exact N-soliton solution and that the Gaussian series can be the basis for a perturbation scheme that corrects for the effects of the spatial periodicity on the solitary waves in the same way that Stokes' Fourier series corrects for the nonlinear interactions of the N-plane waves. Boyd (1982) also shows that there is a high degree of overlap between the small-amplitude, Fourier series expansion and the large-amplitude Gaussian series so that for moderately nonlinear waves, the lowest term of either series is acceptably accurate. Boyd (1983c) extends this work to the Nonlinear Schrödinger equation.

The usefulness of this work on the Hill's spectrum method and polycnoidal waves is that there exists a fairly complete understanding of the spatially periodic solutions of the Nonlinear Schrödinger equation. In particular, it appears that the "simple recurrence" of Yuen and Ferguson (1978) for sideband instability is basically a 2-cnoidal wave and the more complex recurrences they observe can be approximated by polycnoidal waves of higher N. If one represents the polycnoidal waves via the Gaussian series and then truncates at lowest order, one obtains an approximation in terms of interacting solitary waves. Physically as well as mathematically, the mechanism for sideband instability is one of soliton formation.

7. Summary and conclusions

This present work has extended the theory of equatorial waves by calculating the nonlinear effects upon a strongly dispersive wave packet due to its self-interaction. It has been shown that the wave amplitude satisfies the Nonlinear Schrödinger equation. When the Landau constant and dispersion coefficients of this equation are of the same sign, envelope solitons will readily form in which the usual spreading of the envelope of the wave packet is exactly balanced by nonlinearity to give a finite amplitude disturbance of permanent form. Under these same conditions, an initial plane wave is unstable due to sideband modulations, tending to form solitons. If the boundary conditions are periodic, however, the solitons will recombine to recreate the initial conditions, the so-called FPU recurrence. If the signs of the Landau constant and dispersion coefficient are opposite, then the nonlinearity will cause the wave packet to spread more quickly than in linear theory—"superlinear" dispersion.

Fig. 7 illustrates how dramatic these different effects of nonlinearity can be by comparing three numerical integrations of the Nonlinear Schrödinger equation with the same initial condition but with three different values of the nonlinear coefficient ($\nu = 0, 2, -2$). Negative ν has greatly broadened the wave packet in comparison with the linear integration, while with $\nu > 0$ a narrow solitary wave has formed.

The graphs of Section 5 show that which of these two alternatives is realized for a given mode depends on the wavenumber k of the wave packet. Gravity waves propagating toward the east [$w(k) > 0$] have no resonances, but the Landau constant $\nu(k)$ does change sign at some intermediate value of k so that solitons and sideband instability occur only for large k. The mixed Rossby–gravity or "Yanai" wave has solitons only for an intermediate range of k which is bounded from below by a second harmonic resonance at $k = 0.288$ and above by a zero of ν at $k = 0.624$.

For the other negative frequency, westward propagating waves, both gravity and Rossby, it is also true that envelope solitons are impossible for both small and large k. The intermediate range of wavenumbers, however, is confused by the existence of a finite number of second-harmonic resonances together with an infinite number of discrete long-wave/short-wave resonances. Fortunately, for the gravity waves, only the single second-harmonic resonance for each gravitational mode is important; the long-wave/short-wave resonances have very small interactions coefficients so that the neighborhood in k where each such resonance is significant is almost vanishingly small. For all practical purposes, there is a single continuous range of intermediate wavenumber for each westward-propagating gravity wave where solitons occur.

For Rossby waves, envelope solitons also form when the wavenumber is intermediate, but the graph of the Landau constant ν is quite complicated because the long-wave/short-wave resonances now are very strong. The difference is simply that the long waves are always Rossby waves, and short Rossby waves interact with them much more effectively than the short gravity wave packets do. The implication is that a strongly nonlinear Rossby wave packet will evolve in a rather complicated and confusing pattern if its wavenumber falls within the range of the resonances.

Application of these results to the real ocean is both easy and difficult. It is easy in the sense that the wavenumber k is the only nondimensional parameter; by unscrambling the nondimensionalization given in Section 3 and the canonical rescaling of the Nonlinear Schrödinger equation presented in Section 6, one can easily apply the numbers and figures of this work to equatorial waves on any scale for any baroclinic mode. Application is hard in the sense that one must have concrete information about the wavenumber k and the width and amplitude of the initial condition before quantitative results can be obtained, and data on equatorial waves is still very limited. Numerical models are, of course, very good at generating the necessary numbers, but the model results are normally presented in the form of contour plots of the total currents and height field, which are the summation of all the equatorial modes. It is difficult to use the Nonlinear Schrödinger equation to analyze the effects of nonlinearity and dispersion on a single mode if that mode is never seen in isolation.

Nevertheless, the decomposition of model data into the free oscillations of the system as done by Kasahara (1976) on the sphere is certainly possible on the equatorial β plane also. When this is done, the present work will make it possible to go a step beyond linear theory in understanding the evolution of equatorial waves.

Acknowledgments. This work was supported by Grants OCE7901919 and OCE8108530 from the National Science Foundation and by NASA under Grant NGL-22-007-228. A previous paper in this series, "Equatorial Solitary Waves, Part I: Rossby Solitons," was also supported by NASA Grant NGL-22-007-228, but the acknowledgment was inadvertently omitted from the paper itself. I wish to thank Prof. Richard Lindzen for his hospitality at Harvard where much of this work was completed, and also Prof. Stanley Jacobs for sharing with me his considerable knowledge of nonlinear waves and solitons.

APPENDIX A

Numerical Details

Calculation of the Hermite series coefficients of the forcing for u^{10}, u^{12} and the other $O(\epsilon^2)$ fields involves integrals of triple products of Hermite functions which are too difficult for practical analytical calculation. It is, however, easy to perform the calculation numerically using Gauss–Hermite quadrature. To expand

$$F(y) = e^{-(1/2)y^2} \sum_{n=0}^{\infty} \bar{f}_n \bar{H}_n(y), \tag{A1}$$

where the H_n are the normalized Hermite polynomials, one writes

$$\bar{f}_n = \int_{-\infty}^{\infty} e^{-y^2} \bar{H}_n(y) F(y) dy \tag{A2}$$

$$= \sum_{i=1}^{N} w_i e^{-y_i^2} \bar{H}_n(y_i) F(y_i), \tag{A3}$$

where the Gaussian quadrature weights w_i and abscissas y_i are tabulated for various N in Abramowitz and Stegun (1965). $N = 32$, i.e., 32 quadrature points, was used for most of the calculations reported in this paper. However, it is only necessary to sum over half the quadrature points in (A3) [those $y_i > 0$] because all the functions $F(y)$ are either symmetric or antisymmetric about $y = 0$. It is shown in Table 2 of Boyd (1980b) that the Hermite coefficients $\bar{f}_n$ must decrease like those of a geometric series; it can be shown that the quadrature error decreases exponentially fast with N also. For the lowest modes, the coefficients of the Nonlinear Schrödinger equation are accurate to five or six decimal places.

For ease of exposition, unnormalized Hermite polynomials were employed in the main body of this paper, but to avoid machine overflow—the unnormalized polynomials increase exponentially with n—normalized Hermite polynomials were employed in the computer program. The unnormalized series expansion coefficients f_n are related to those normalized coefficients $\bar{f}_n$ defined by (A1) and (A2) via

$$f_n = \frac{\bar{f}_n}{\pi^{1/4} 2^{n/2} [(2n)!]^{1/2}} . \tag{A4}$$

The normalized Hermite polynomials satisfy the recursion

$$\bar{H}_{n+1} = 2^{1/2}(n+1)^{-1/2} y \bar{H}_n - [n/(n+1)]^{1/2} \bar{H}_{n-1}, \tag{A5}$$

with the starting values

$$\bar{H}_0 = \pi^{-1/4}, \tag{A6}$$

$$\bar{H}_1 = \pi^{-1/4} 2^{1/2} y. \tag{A7}$$

The remaining technical problems are posed by the complex arithmetic. Since all quantities can, by suitable choice of phase for the wave packet, be taken as either real or imaginary, one can perform all the calculations in real arithmetic and reduce the computational cost by a factor of 4. One must, however, keep careful track of which quantities are real and which are imaginary so that when two imaginary terms are multiplied, the sign change due to squaring $i = (-1)^{1/2}$ is inserted.

It is necessary to be careful in any case, however, because every wave field consists of a term plus its complex conjugate. Thus, a term like

$$v^{01} v_y^{01}, \tag{A8}$$

which would seem to have a minus because v^{01} is pure imaginary, actually yields

$$-A_{01}^2 E^2 + 2|A_{01}|^2 + \text{c.c.}; \tag{A9}$$

the sign change appears only for the second harmonic in (A9) because the long-wave component is the product of iA_{01} with its complex conjugate $-iA_{01}^*$. A term like

$$u^{01} v_x^{01} \tag{A10}$$

has a negative sign for *both* components on the other hand because the x-derivative of $iA_{01}E$ is $i^2 k A_{01} E$ while u^{01} is real. One must carefully write out each product in terms of both A_{01} and A_{01}^* to avoid mistakes.

The relationships between the Korteweg–deVries and Nonlinear Schrödinger equations, derived in Appendix C, provided useful numerical checks as did the relationships between the residues of the poles of $\nu(k)$ at the resonant values of k and the coefficients of the resonant dyad equations and long-wave/short-wave equations.

APPENDIX B

Errors in Approximate Equatorial Dispersion Relations

Both the dispersion coefficient $w''(k)/2$ here and the calculation of long-wave and second-harmonic resonances in Boyd (1982a,b) depend on the dispersion relation $w(k)$. Unfortunately, this is determined by the cubic equation

$$w^3 - (k^2 + 2n + 1)w - k = 0, \quad \text{(B1)}$$

which does not have simple closed form solutions except for $n = 0$ when (B1) reduces to a quadratic. However, the approximate roots

$$w \approx \frac{-k}{2n + 1 + k^2} \text{ [Rossby waves]}, \quad \text{(B2)}$$

$$w \approx \pm(2n + 1 + k^2)^{1/2} \text{ [gravity waves]}, \quad \text{(B3)}$$

or the higher approximation [Domarcki and Loesch, (1977) who give a third approximation, too],

$$w \approx \pm(k^2 + 2n + 1)^{1/2} \times \left[1 \pm \frac{k}{2(k^2 + 2n + 1)^{3/2}}\right], \quad \text{(B4)}$$

are all very accurate.

To show this for Rossby waves, we define Δ by writing

$$w = \frac{-k}{k^2 + 2n + 1} [1 + \Delta], \quad \text{(B5)}$$

so that Δ is the relative error in the approximation (B2). Substituting (B5) into (B1) gives the exact relationship

$$\Delta = \epsilon(k, n)[1 + \Delta]^3, \quad \text{(B6)}$$

where

$$\epsilon(k, n) \equiv \frac{k^2}{(k^2 + 2n + 1)^3}. \quad \text{(B7)}$$

By differentiating (B7) with respect to k to find the maximum, one can easily show that

$$\epsilon \leq \frac{1}{27(n + 1/2)^2} \quad \text{(B8)}$$

$$\leq 0.0165, \quad \text{for all } n \geq 1. \quad \text{(B9)}$$

Thus, ϵ is a very small quantity even for the lowest mode so that (B6) can be made the basis of the iteration

$$\Delta^{n+1} = \epsilon(k, n)[1 + \Delta^n]^3, \quad \text{(B10)}$$

which gives

$$\Delta \approx \epsilon + 3\epsilon^2 + O(\epsilon^3). \quad \text{(B11)}$$

The smallness of ϵ should make it clear that the iteration converges very rapidly and that Δ, the relative error in (B2), is roughly equal to ϵ. One can express this mathematically by applying the methods of functional analysis to the iteration (B10); the proof is taken from Sawyer (1972) and is therefore omitted, but the final result is given below.

In a similar way for gravity waves, we let

$$w = \pm(k^2 + 2n + 1)^{1/2}[1 + \Delta]. \quad \text{(B12)}$$

The iteration is

$$\Delta^{n+1} = \lambda - \tfrac{3}{2}(\Delta^n)^2 + \tfrac{1}{2}(\Delta^n)^3, \quad \text{(B13)}$$

where

$$\lambda = \pm \frac{k}{2(k^2 + 2n + 1)^{3/2}}. \quad \text{(B14)}$$

One can show, again by differentiating with respect to k, that

$$\lambda \leq \frac{1}{3(3)^{1/2}(2n + 1)}. \quad \text{(B15)}$$

Like ϵ, λ is a small quantity and the relative error is approximately given by the first iteration, i.e., $\Delta \approx \lambda$ so that the error in the lowest gravity wave approximation [(B3)] is roughly bounded by the right-hand side of (B15). Functional analysis gives the tools to prove the following theorem. Details are given in Sawyer (1972).

THEOREM: The errors in the approximations (B2)–(B4) have the following bounds:

(i) $$w = \frac{-k}{k^2 + 2n + 1} \times [1 + \Delta] \text{ (Rossby waves)}, \quad \text{(B2)}$$

where

$$|\Delta| \leq \frac{0.039}{(n + 1/2)^2}, \quad n \geq 1. \quad \text{(B16)}$$

The relative error is no worse than 1.7% even for $n = 1$.

(ii) $$w \approx \pm(k^2 + 2n + 1)^{1/2} \times [1 + \Delta] \text{ (gravity waves)} \quad \text{(B3)}$$

$$|\Delta| \leq \frac{0.141}{(n + 1/2)}. \quad \text{(B17)}$$

The relative error is no worse than 10% even for $n = 1$.

(iii) The relative error in the second approximation (B4) for gravity waves is bounded by

$$|w - [\pm(k^2 + 2n + 1)^{1/2}(1 + \lambda)]| \leq \frac{0.021}{(n + 1/2)^2}. \quad \text{(B18)}$$

For the mixed Rossby-gravity ($n = 0$), the cubic dispersion relation collapses to a quadratic which can be solved exactly, so we lose nothing by demanding that $n \geq 1$ in the theorem.

APPENDIX C

The Relationship Between the Nonlinear Schrödinger Equation and the Korteweg–deVries Equation

In Boyd (1980b), it was shown that the Korteweg–deVries equation is a consistent description of ultra-long Rossby waves of small amplitude. To derive the Nonlinear Schrödinger equation for this same class of waves, one can bypass the complicated modal analysis of Section 3 and attack the Korteweg–deVries equation

$$u_t + auu_x + bu_{xxx} = 0 \tag{C1}$$

directly, thus keeping to but a single spatial dimension. The multiple scales analysis, however, is exactly the same as for Section 3. Thus, a derivation of the NLS equation from (C1) is presented here so that the reader may see the skeleton of the perturbation theory without the Hermite function analysis that makes Section 3 rather heavy reading.

As in Section 3, we assume that

$$\epsilon \ll 1, \tag{C2}$$

$$k \gg \epsilon, \tag{C3}$$

introduce slow space and time scales via

$$\tau = \epsilon^2 t, \tag{C4}$$

$$\zeta = \epsilon(x - c_g t), \tag{C5}$$

and take the lowest order solution as the wave packet

$$u^0 = A_{01}(\zeta, \tau)e^{ik(x-c_p t)} + \text{c.c.}, \tag{C6}$$

where c.c. denotes the complex conjugate. A uniformly valid perturbation expansion can then be obtained in the form

$$u = \epsilon \sum_{n=0}^{\infty} \epsilon^n \sum_{m=0}^{n+1} A_{nm}(\zeta, \tau)E^m + \text{c.c.}, \tag{C7}$$

where

$$E \equiv e^{ik(x-c_p t)},$$

with c_p and c_g as the linear phase and group velocities, respectively. Derivatives transform according to the chain rules:

$$\frac{\partial}{\partial t} = -ikc_p - \epsilon c_g \frac{\partial}{\partial \zeta} + \epsilon^2 \frac{\partial}{\partial \tau}, \tag{C8}$$

$$\frac{\partial}{\partial x} = ik + \epsilon \frac{\partial}{\partial \zeta}. \tag{C9}$$

The lowest [O(ϵ)] equation is

$$u_t^0 + bu_{xxx}^0 = 0, \tag{C10}$$

with the solution given by (C6) with the dispersion relation

$$c_p = -bk^2. \tag{C11}$$

The function of this lowest order equation is simply to determine the dispersion relation. For notational simplicity, the second superscript on u^0 has been suppressed because the whole lowest order solution is proportional to E.

At O(ϵ^2), however, we have three separate equations in E^0, E^1 and E^2 (plus the complex conjugates E^{-1} and E^{-2} but these need not be explicitly solved). By multiplication, the nonlinear term $u^0u_x^0$ is the sum of terms proportional to E^0 and E^2 and thus provides the forcing for u^{10} and u^{12}. The forcing for u^{11} is linear, i.e.,

$$u_t^{11} + bu_{xxx}^{11} = c_g u_\zeta^0 - 3bu_{\zeta xx}^0 \tag{C12}$$

$$= (c_g + 3bk^2)u_\zeta^0. \tag{C13}$$

Eq. (C12) can have a bounded solution, i.e., the "non-secularity" condition can be satisfied only if the right-hand side of (C13) is identically equal to zero, which implies

$$c_g = -3bk^2. \tag{C14}$$

However, the usual definition of the group velocity,

$$c_g \equiv \frac{\partial w}{\partial k}, \tag{C15}$$

where $w = kc_p$ is the frequency, gives (C14) again if the dispersion relation (C11) is used. Consequently, the O($\epsilon^2 E^1$) equation merely tells us what we knew already, namely (C15) and

$$u^{11} \equiv 0. \tag{C16}$$

In the two-dimensional problems such as those solved in Section 3, however, u^{11} need not (and generally is not) identically equal to zero. Eq. (C15), and its justification via the non-secularity condition at this order, are both general results, however.

The lowest order nonlinear term is

$$u^0u_x^0 = (A_{01}E + A_{01}^*E^{-1})(ikA_{01}E - ikA_{01}^*E^{-1}), \tag{C17}$$

$$= ikA_{01}^2E^2 + (ik - ik)|A_{01}|^2 - ikA_{01}^{*2}E^{-2}, \tag{C18}$$

$$= ikA_{01}^2 + \text{c.c.}, \tag{C19}$$

where the asterisk and c.c. both denote the complex conjugate. Thus, to O(ϵ^2) there is no forcing of the long waves, but this should not be interpreted to mean that there is no O(ϵ^2) long-wave component. This rather subtle point will be explained in a moment.

The second harmonic, however, presents no complication if we assume that

$$u_t^{12} + bu_{xxx}^{12} = -ikaA_{01}^2E^2 + \text{c.c.} \tag{C20}$$

has a particular solution proportional to E^2 $\{=\exp[2ik(x - c_p t)]\}$ to obtain

$$u^{12} = \frac{aA_{01}^2E^2}{6bk^2} + \text{c.c.} \tag{C21}$$

By assumption, there is no second harmonic which is not directly forced by the fundamental wave packet u^0, so there is no need to add a homogeneous solution of (C20) to the particular solution (C21). More important, the free second harmonic travels at a speed different from that of (C21) and would therefore be irrelevant to the Landau constant even if the harmonic had a non-zero amplitude.

The subtlety involving u^{10} may be called "deferred determination of the long waves": although u^{10} itself is $O(\epsilon^2)$, the terms in the perturbation equation that determines it are $O(\epsilon^3)$, i.e.,

$$-c_g u_\zeta^{10} + 3bu_{\zeta xx}^{10} = -\tfrac{1}{2}a(\overline{u^0u^0})_\zeta, \qquad (C22)$$

where the overbar denotes that only that portion of $(u^0u^0)_\zeta$ which is independent of E is to be taken. The reason for this peculiar state of affairs is that u^{10} is independent of the "fast" space and time scales and is therefore a function of the "slow" variables ζ and τ. The ζ-differentiations introduce the extra ϵ in (C22). The derivatives with respect to x cause the second term in (C22) to vanish, reducing it to

$$-c_g u_\zeta^{10} = -\tfrac{1}{2}a(2|A_{01}|^2)_\zeta, \qquad (C23)$$

which implies

$$u^{10} = -\frac{a}{3bk^2}|A_{01}|^2, \qquad (C24)$$

where, as in Section 3, it was necessary to integrate with respect to ζ in the final step.

We are now ready to proceed to the final step which is to attack the $O(\epsilon^3E^1)$ equation

$$u_t^{21} + bu_{xxx}^{21} = -\{[u_\tau^0 + 3bu_{\zeta\zeta x}^0] + a[u^0u_x^{10} + u^{10}u_x^0] + a[u^0u_x^{12} + u^{12}u_x^0]\}. \qquad (C25)$$

The terms in the first set of brackets come from applying the chain rules (C8) and (C9) to the linear terms in the Korteweg–deVries equation (C1) and supply the two linear terms in the Nonlinear Schrödinger equation.

The first term in the second set of brackets is zero because u^{10} is independent of the "fast" variable x, so the forcing due to the interaction of the fundamental with its forced long wave is simply

$$u^{10}u_x^0 = \frac{-iaA_{01}|A_{01}|^2}{3bk}. \qquad (C26)$$

By direct multiplication, the corresponding second-harmonic forcing is

$$u^0u_x^{12} + u^{12}u_x^0 = \frac{iaA_{01}^3}{2bk}E^3 + \frac{iaA_{01}|A_{01}|^2E}{6bk} + \text{c.c.} \qquad (C27)$$

The first term is proportional to E^3 and must be discarded from (C25); it provides the forcing for an $O(E^3\epsilon^3)$ third harmonic. The surviving term is equal to $-\tfrac{1}{2}$ the value of its long-wave counterpart (C26).

The non-secularity condition is that the forcing in (C25) vanish (except the E^3 part), which demands that

$$\frac{A_{01}}{\partial\tau} + 3ibk\frac{\partial^2A_{01}}{\partial\zeta^2} - \frac{ia^2}{6bk}|A_{01}|^2A_{01} = 0, \qquad (C28)$$

which is the usual Nonlinear Schrödinger equation.

The corresponding derivation for the Modified Korteweg–deVries equation

$$u_t + au^2u_x + bu_{xxx} = 0 \qquad (C29)$$

is similar but essentially trivial because the cubic nonlinearity in (C29) implies that $u^{10} = u^{12} = 0$ so that the nonlinear term in the Schrödinger equation is determined directly from that term in $(u^0)^2u_x^0$ which is proportional to E. We find

$$\frac{\partial A_{01}}{\partial\tau} + 3ibk\frac{\partial^2A_{01}}{\partial\zeta^2} + iak|A^{01}|^2A^{01} = 0. \qquad (C30)$$

Note that the Landau constant vanishes as $k \to 0$ instead of blowing up as in (C28); resonance as $k \to 0$ occurs only for the Korteweg–deVries equation as discussed previously in Section 4.

REFERENCES

Ablowitz, M. J., *et al.*, 1974: The inverse scattering transform-Fourier analysis for nonlinear problems. *Stud. Appl. Math.*, **53**, 249–315.

Abramowitz, M., and I. A. Stegun, 1965: *Handbook of Mathematical Functions.* Dover, 1043 pp.

Benjamin, T. B., and J. E. Feir, 1967: The disintegration of wavetrains on deep water. Part 1. Theory. *J. Fluid Mech.*, **27**, 417–430.

Boyd, J. P., 1980a: The nonlinear equatorial Kelvin wave. *J. Phys. Oceanogr.*, **10**, 1–11.

——, 1980b: Equatorial solitary waves, Part 1: Rossby solitons, *J. Phys. Oceanogr.*, **10**, 1699–1717.

——, 1982: Theta functions, Gaussian series, and the spatially periodic solutions of the Korteweg–deVries equation. *J. Math. Phys.*, **23**, 375–387.

——, 1983a: Long wave/short wave resonance in equatorial waves *J. Phys. Oceanogr.*, **13**, 450–458.

——, 1983b: Second harmonic resonance in equatorial waves. *J. Phys. Oceanogr.*, **13**, 459–466.

——, 1983c: Theta functions and FPU recurrence for the Nonlinear Schrödinger equation. (In preparation.)

Domaracki, A., and A. Loesch, 1977: Nonlinear interactions among equatorial waves. *J. Atmos. Sci.*, **31**, 486–498.

Dubrovin, B. A., V. B. Matveev and S. P. Novikov, 1976: Nonlinear equations of Korteweg-deVries type, finite zone linear operators, and Abelian varieties. *Russian Math. Surveys*, **31**, 59–146.

Hasimoto, H., and H. Ono, 1972: Nonlinear modulation of gravity waves. *J. Phys. Soc. Japan*, **33**, 805–811.

Infeld, E., 1981: Quantitative theory of the Fermi-Pasta-Ulam recurrence in the Nonlinear Schrödinger equation. *Phys. Rev. Lett.*, **47**, 717–718.

Johnson, R. S., 1976: Nonlinear, strongly dispersive water waves in arbitrary shear. *Proc. Roy. Soc. London*, **A338**, 101–114.

Kasahara, A., 1976: Normal modes of ultralong waves in the atmosphere. *Mon. Wea. Rev.*, **104**, 669–690.

Kaup, D. J., 1977: Coherent pulse propagation: A comparison of the complete solution with the McHall-Hahn theory and others. *Phys. Rev.*, **A16**, 704–719.

Loesch, A., and R. D. Deininger, 1979: Dynamics of closed systems of resonantly interacting equatorial waves. *J. Atmos. Sci.*, **36**, 1490–1497.

McKean, H. P., and E. Trubowitz, 1978: Hill's surfaces and their theta functions. *Bull. Amer. Math. Soc.*, **84**, 1042–1084.
Manakov, S. V., 1974: Nonlinear Fraunhofer diffraction. *Sov. Phys. JETP*, **38**, 693–696.
Moore, D. W., 1968: Planetary gravity waves in an equatorial ocean. Ph.D. thesis, Harvard University.
Ripa, P., 1982: Nonlinear wave-wave interactions among equatorial waves. *J. Phys. Oceanogr.*, **12**, 97–111.
——, 1983a: Resonant triads of equatorial waves in a one-layer model. Part I. Non-local triads and triads of waves with the same speed. *J. Phys. Oceanogr.*, **13** (in press).
——, 1983b: Resonant triads of equatorial waves in a one-layer model. Part II: General classification of resonant triads and double triads. *J. Phys. Oceanogr.*, **13** (in press).
Satsuma, J., and N. Yajima, 1974: Initial value problems of one-dimensional self-modulation of nonlinear waves in dispersive media. *Suppl. Prof. Theor. Phys.*, **55**, 284–306.
Sawyer, W. W., 1972: *Functional Analysis in Numerical Analysis.* Methuen, 160 pp.
Segur, H., 1976: Asymptotic solutions and conservation laws for the Nonlinear Schrödinger equation, II. *J. Math. Phys.*, **17**, 714–716.
——, and M. J. Ablowitz, 1976: Asymptotic solutions and conservation laws for the Nonlinear Schrödinger equation, I. *J. Math. Phys.*, **17**, 710–713.
Stokes, G. G., 1847: On the theory of oscillatory waves. *Cambridge Trans.*, **8** (Reprinted in Stokes' *Collected Papers*, Cambridge University Press, 6 vols.).
Yuen, H. C., and B. M. Lake, 1980: Instabilities of waves on deep water. *Annual Reviews of Fluid Mechanics*, Vol. 12, Annual Reviews, Inc., 303–334.
——, and W. E. Ferguson, Jr., 1978: Relationship between Benjamin-Feir instability and recurrence in the Nonlinear Schrödinger equation. *Phys. Fluids*, **21**, 303–334.
Zakharov, V. E., and A. B. Shabat, 1972: Exact theory of two-dimensional self-focusing and one-dimensional modulational instability of waves in nonlinear media. *Sov. Phys. JETP*, **34**, 62–69.