J. Fluid Mech. (1967), *vol.* 27, *part* 3, *pp.* 417–430
Printed in Great Britain

The disintegration of wave trains on deep water

Part 1. Theory

By T. BROOKE BENJAMIN AND J. E. FEIR†

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge

(Received 21 June 1966)

The phenomenon in question arises when a periodic progressive wave train with fundamental frequency ω is formed on deep water—say by radiation from an oscillating paddle—and there are also present residual wave motions at adjacent side-band frequencies $\omega(1 \pm \delta)$, such as would be generated if the movement of the paddle suffered a slight modulation at low frequency. In consequence of coupling through the non-linear boundary conditions at the free surface, energy is then transferred from the primary motion to the side bands at a rate that, as will be shown herein, can increase exponentially as the interaction proceeds. The result is that the wave train becomes highly irregular far from its origin, even when the departures from periodicity are scarcely detectable at the start.

In this paper a theoretical investigation is made into the stability of periodic wave trains to small disturbances in the form of a pair of side-band modes, and Part 2 which will follow is an account of some experimental observations in accord with the present predictions. The main conclusion of the theory is that infinitesimal disturbances of the type considered will undergo unbounded magnification if

$$0 < \delta \leqslant (\sqrt{2})ka,$$

where k and a are the fundamental wave-number and amplitude of the perturbed wave train. The asymptotic rate of growth is a maximum for $\delta = ka$.

1. Introduction

The object of this paper is to establish analytically that progressive waves of finite amplitude on deep water (that is, Stokes waves) are unstable. This proposition implies that in practice, where perturbations from the ideal wave motion are inevitably present, a train of such waves will disintegrate if it travels far enough. Experimental evidence of this remarkable property has been found by us, and our observations will be presented in Part 2 of this study (Benjamin & Feir 1967).

The present findings are perhaps most striking when considered as an epilogue to the famous controversy about the existence of water waves of permanent form. The history of the controversy is summarized in Lamb's (1932, p. 420) text-book,

† On leave from the Division of Mechanical Engineering, National Research Council, Ottawa, Canada.

where the issue is explained to depend on the convergence of the power series in wave amplitude whose leading terms were first obtained by Stokes as an approximate solution of the non-linear problem. For periodic waves on infinitely deep water, Levi-Civita (1925) proved these series to be convergent if the ratio of amplitude to wavelength is sufficiently small, and thus the existence of permanent waves satisfying the exact non-linear boundary conditions was definitely established. Soon afterwards Struik (1926) extended the proof to waves on water of finite depth. Among recent work on the subject, the greatest advances have been made by Krasovskiĭ (1960, 1961), who proved the existence of permanent periodic waves subject only to the restriction that their maximum slope is less than the limiting value of 30°. For a long time no doubt has remained, therefore, that water waves of unchanging form are theoretically possible as states of perfect dynamic equilibrium. Until now, however, it has apparently not been suspected that for waves on deep water the equilibrium so long in question is in fact unstable.

Although the detailed analysis of the instability is necessarily complicated, the essential factors can be simply explained as follows. Consider the various simple-harmonic modes present in a slightly disturbed wave motion. We have first, for the basic wave train, the fundamental component with amplitude a and argument $\zeta = kx - \omega t$, say, and harmonics with arguments $2\zeta, 3\zeta, \ldots$, which all advance in the horizontal x-direction with the phase velocity $c = \omega/k$. For the disturbance, we take a pair of progressive-wave modes which have 'side-band' frequencies and wave-numbers adjacent to ω and k, so that their arguments may be expressed by

$$\left.\begin{aligned}\zeta_1 &= k(1+\kappa)x - \omega(1+\delta)t - \gamma_1,\\ \zeta_2 &= k(1-\kappa)x - \omega(1-\delta)t - \gamma_2,\end{aligned}\right\} \tag{1}$$

where κ and δ are small fractions. The respective amplitudes are denoted by ϵ_1, ϵ_2 and are assumed to be much smaller than a. Now, among the products of a non-linear interaction between these disturbance modes and the basic wave train, there will be components with arguments

$$\left.\begin{aligned}2\zeta - \zeta_1 &= \zeta_2 + (\gamma_1+\gamma_2),\\ 2\zeta - \zeta_2 &= \zeta_1 + (\gamma_1+\gamma_2),\end{aligned}\right\} \tag{2}$$

and with amplitudes proportional to $a^2\epsilon_1$ and $a^2\epsilon_2$, respectively. Thus, if it happens that

$$\theta = \gamma_1 + \gamma_2 \to \text{const.} \tag{3}$$

as the non-linear processes develop in time, each mode will generate effects that become resonant with the other. Thereafter, if $\theta \neq 0, \pi$, each mode suffers a synchronous forcing action proportional to the amplitude of the other, so that the two can grow mutually at an exponential rate.

The crucial task of the analysis is to show that, for given specifications a, k, ω of the basic wave train, the property (3) is possible for some non-zero κ and δ. This property would be impossible in the absence of the basic wave train, or if the amplitude a were too small for there to be significant non-linear coupling with the side-band modes, because then the *net* frequencies $\omega_{1,2}$ and wave-numbers $k_{1,2}$ of these modes would have to obey the dispersion relation

$$\omega_{1,2}^2 = gk_{1,2} \tag{4}$$

given by linearized theory (Lamb 1932, §229). Letting $k_{1,2} = k(1 \pm \kappa)$ precisely, for instance, we see that the frequencies $\omega(1 \pm \delta)$ appearing explicitly in (1) cannot both satisfy (4), and the discrepancy must be accomodated by allowing γ_1 and γ_2 to be slowly-varying functions of time. If we put $\delta = \frac{1}{2}\kappa$, which means that $\delta\omega/\kappa k = \frac{1}{2}c = \frac{1}{2}(g/k)^{\frac{1}{2}}$ is the group velocity for an infinitesimal wave with wave-number k, then (4) is satisfied to a first approximation for small δ, even if γ_1 and γ_2 are constant; but to a second approximation (4) requires that

$$d\theta/dt = -\omega\delta^2. \tag{5}$$

Thus the effect of dispersion, in so far as it may be independent of non-linear effects, is to detune the prospective resonance between second-harmonic components of the basic wave motion and the side-band modes.

However, it is already known from the work of Longuet-Higgins & Phillips (1962) and Benney (1962) that the presence of one train of waves on deep water will affect the phase velocity of another train. The change produced was shown to be of second order in the amplitude of the waves responsible for it; and so in the present instance we may expect that terms in a^2 will be added to (5) when the non-linear interaction is analysed to the same order of approximation as in the previous work. The property (3) appears to be possible, therefore, if these particular non-linear effects act so as to balance the effect of dispersion represented by (5). And it may be expected that $\delta = O(ka)$ if disturbances of the kind in view are to manifest instability.

Despite the ground in common with previous studies of water-wave interactions, the presentation of a new treatment from first principles is judged desirable. There are several essential results in the present theory that are not readily accessible from existing analyses: in particular, we need to find the actual asymptotic value of the phase function θ in order to predict the ultimate rate of amplification of an unstable disturbance. In Part 2 this prediction will be shown to compare favourably with our experimental results.

2. Analysis

We consider two-dimensional irrotational motion in infinitely deep water, which is modelled as an inviscid fluid. The axis x is drawn horizontally and y vertically upwards, with $y = 0$ denoting the mean level of the free surface. The equation of the free surface is written

$$y = \eta(x,t), \tag{6}$$

so that η denotes the elevation of the surface above its mean level.

The velocity potential $\phi(x,y,t)$ satisfies

$$\nabla^2\phi = \phi_{xx} + \phi_{yy} = 0 \tag{7}$$

everywhere on and below the free surface; and, assuming there to be no motion at infinite depths, we have

$$\nabla\phi \to 0 \quad \text{for} \quad y \to -\infty. \tag{8}$$

The kinematical boundary condition at the free surface is

$$D(\eta - y)/Dt = \eta_t + \eta_x[\phi_x]_{y=\eta} - [\phi_y]_{y=\eta} = 0; \tag{9}$$

and, surface tension being supposed absent, the condition of constant pressure on the free surface may be expressed

$$g\eta + [\phi_t]_{y=\eta} + \tfrac{1}{2}[\phi_x^2 + \phi_y^2]_{y=\eta} = 0. \tag{10}$$

Here it is understood that an arbitrary function of t may be added to ϕ (Lamb 1932, §§20, 227).

The Stokes approximation to waves of permanent form

The non-linear boundary-value problem defined by the preceding equations is known to have exact periodic solutions in the form $\eta = H(x-ct)$, $\phi = \Phi(x-ct, y)$, where c is a constant phase velocity. As we recalled in §1, the existence of such solutions was proved by Levi-Civita (1925), who demonstrated the convergence of the series expansions for η, ϕ and the parameter c whose leading terms comprise the approximation first given by Stokes. At present we need to take the approximation only as far as the first terms representing the effects of finite wave amplitude a.

The required result is (cf. Lamb, §250)

$$\eta = H = a\cos\zeta + \tfrac{1}{2}ka^2\cos 2\zeta, \tag{11}$$

$$\phi = \Phi = \omega k^{-1} a\, e^{ky} \sin\zeta, \tag{12}$$

in which $\zeta = kx - \omega t$, and

$$\omega^2 = gk(1 + k^2a^2). \tag{13}$$

This describes, with sufficient accuracy if ka is very small, the steady wave motion whose stability is now to be examined.

Perturbation equations

The question of stability is approached in the usual way by deriving the equations satisfied by a small perturbation and hence investigating its development in time. Accordingly we write

$$\phi = \Phi + \epsilon\tilde{\phi}, \qquad \eta = H + \epsilon\tilde{\eta}, \tag{14}$$

and aim to prove instability, if it exists, by demonstrating the asymptotic properties of $\tilde{\eta}$.

Since (7) is a linear equation, it follows that

$$\tilde{\phi}_{xx} + \tilde{\phi}_{yy} = 0, \tag{15}$$

and

$$\nabla\tilde{\phi} \to 0 \quad \text{for} \quad y \to -\infty. \tag{16}$$

After substitution of (14) into the boundary conditions (9) and (10), linearization in ϵ gives

$$\tilde{\eta}_t + \tilde{\eta}_x[\Phi_x]_{y=H} + \tilde{\eta}[-\Phi_{yy} + H_x\Phi_{xy}]_{y=H} + [-\tilde{\phi}_y + H_x\tilde{\phi}_x]_{y=H} = 0, \tag{17}$$

$$g\tilde{\eta} + \tilde{\eta}[\Phi_x\Phi_{xy} + \Phi_y\Phi_{yy} + \Phi_{ty}]_{y=H} + [\tilde{\phi}_t + \Phi_x\tilde{\phi}_x + \Phi_y\tilde{\phi}_y]_{y=H} = 0. \tag{18}$$

These are the exact linearized perturbation equations in so far as Φ and H represent the exact basic solution. After substitution of the approximate expressions (11) and (12), however, it is consistent to simplify (17) and (18) by reducing the coefficients to approximations of the same order—namely, as far as terms in a^2. At the same time, an analytical continuation of $\tilde{\phi}$ over the neighbourhood of the free surface can be assumed, and so the various derivatives of $\tilde{\phi}$ at $y = H$ can be evaluated in the form of Taylor series about $y = 0$. In this way, (17) and (18) lead to

$$\begin{aligned}\tilde{\eta}_t - (\tilde{\phi}_y)_{y=0} = {} & a[k\omega\sin\zeta\,\tilde{\eta} - \omega\cos\zeta\,\tilde{\eta}_x + (\cos\zeta\,\tilde{\phi}_{yy} + k\sin\zeta\,\tilde{\phi}_x)_{y=0}] \\ & + \tfrac{1}{2}a^2[2k^2\omega\sin 2\zeta\,\tilde{\eta} - k\omega(1+\cos 2\zeta)\tilde{\eta}_x \\ & + \{k\sin 2\zeta(2k\tilde{\phi}_x + \tilde{\phi}_{xy}) + k\cos 2\zeta\,\tilde{\phi}_{yy} + \tfrac{1}{2}(1+\cos 2\zeta)\tilde{\phi}_{yyy}\}_{y=0}],\end{aligned} \tag{19}$$

$$\begin{aligned}g\tilde{\eta} + (\tilde{\phi}_t)_{y=0} = {} & a[\omega^2\cos\zeta\,\tilde{\eta} - (\omega\cos\zeta\,\tilde{\phi}_x + \omega\sin\zeta\,\phi_y + \cos\zeta\,\phi_{yt})_{y=0}] \\ & - \tfrac{1}{2}a^2[k\omega^2(1-\cos 2\zeta)\tilde{\eta} + \{\omega\sin 2\zeta(k\tilde{\phi}_y + \tilde{\phi}_{yy}) \\ & + (1+\cos 2\zeta)(k\omega\tilde{\phi}_x + \omega\tilde{\phi}_{xy} + \tfrac{1}{2}\tilde{\phi}_{yyt}) + k\cos 2\zeta\,\tilde{\phi}_{yt}\}_{y=0}].\end{aligned} \tag{20}$$

These forms of the boundary conditions enable us readily to develop the solution $\tilde{\eta}$, $\tilde{\phi}$ to second order in powers of a.

Assumed form of solution

The character of unstable perturbations can be foreseen on the basis of the ideas explained in §1, and it would pointlessly disguise the main issue of this analysis to allow for a more general type of perturbation. The solution $\tilde{\eta}$, $\tilde{\phi}$ is therefore assumed to consist of two side-band modes, together with the products of their interaction with the basic wave train, and certain simplifications permitted by this assumption are introduced. The symbol ϵ has now served its purpose as an ordering parameter in the derivation of linearized equations, and no confusion should arise from its being used again, as in §1, for the amplitudes of the side-band modes.

We accordingly take

$$\tilde{\eta} = \tilde{\eta}_1 + \tilde{\eta}_2, \tag{21}$$

where the two components have the following form, with $i = 1, 2$ respectively:

$$\tilde{\eta}_i = \epsilon_i\cos\zeta_i + ka\epsilon_i\{A_i\cos(\zeta+\zeta_i) + B_i\cos(\zeta-\zeta_i)\} + O(k^2a^2\epsilon_i). \tag{22}$$

Here the arguments ζ_i are given by (1), and the coefficients A_i, B_i are understood to be $O(1)$. The implied terms that are $O(k^2a^2\epsilon_i)$ have arguments $2\zeta+\zeta_i$ and do not need to be considered in analysing the instability to the adopted order of approximation: they represent effects of the interaction with the basic wave train that are non-resonant and therefore passive. Terms with arguments $2\zeta-\zeta_i$

play a crucial role, on the other hand; but, as will appear presently, they can be merged into the leading terms of the present expansions.

It is assumed that the ϵ_i and γ_i are slowly-varying functions of time, such that their derivatives have the properties

$$\dot{\epsilon}_i = O(\omega k^2 a^2 \epsilon_i), \quad \dot{\gamma}_i = O(\omega k^2 a^2). \tag{23}$$

These orders of magnitude can be guessed from the considerations outlined in § 1, and they will be confirmed by the final results of the analysis. The form of $\tilde{\phi}$ corresponding to (21) must include terms in $\dot{\epsilon}_i$ and $\dot{\gamma}_i$, but, as before, terms also proportional to a^2 but with arguments $2\zeta + \zeta_i$ can be ignored. Hence the appropriate solution of (15) satisfying (16) can be expressed as the sum

$$\tilde{\phi} = \tilde{\phi}_1 + \tilde{\phi}_2, \tag{24}$$

in which

$$\tilde{\phi}_i = k_i^{-1} e^{k_i y}\{\epsilon_i(\omega_i' L_i + \dot{\gamma}_i M_i)\sin\zeta_i + \dot{\epsilon}_i N_i \cos\zeta_i\} + \omega a \epsilon_i\{C_i e^{|k+k_i|y}\sin(\zeta+\zeta_i) + D_i e^{|k-k_i|y}\sin(\zeta-\zeta_i)\}, \tag{25}$$

with

$$k_i = k(1 \pm \kappa), \quad \omega_i' = \omega(1 \pm \delta). \tag{26}$$

As already noted in § 1, we put

$$\delta = \tfrac{1}{2}\kappa \tag{27}$$

in order that k_i and ω_i' should satisfy the dispersion relation (4) to a first approximation for small δ. Recalling another point made in § 1, we assume

$$\delta = O(ka), \tag{28}$$

and anticipate this will be confirmed by the final results. Hence, in keeping with the scheme of approximation already defined in powers of ka, coefficients depending on δ can be simplified in the ensuing work. The assumption (28) is by no means necessary to the completion of the analysis, but the gain in generality by not using it (i.e. the coverage of frequencies well outside the unstable range) is judged not to be worth the attendent complication of the results. Note that in consequence of these assumptions the boundary conditions (19) and (20) imply that

$$L_i = 1 + O(k^2 a^2). \tag{29}$$

The coefficients M_i, N_i, C_i, D_i are also to be regarded as $O(1)$.

Evaluation of coefficients

The boundary conditions (19) and (20) are to be satisfied over a continuous and unbounded range of x. Therefore, if all the terms in them are reduced to simple-harmonic components, each set of components at every different wave-number must satisfy these conditions independently. The next part of the analysis, leading to differential equations for $\epsilon_i(t)$ and $\gamma_i(t)$, proceeds on this well-known principle.

The first task is to find the coefficients of the terms proportional to $a\epsilon_i$ in (22) and (25). By the separation of components with arguments $\zeta \pm \zeta_i$ in (19) and (20), these coefficients are gathered on the left-hand sides of the equations, while terms of the required order of magnitude are given on the right-hand sides by substituting the 'zeroth' approximation

$$\tilde{\eta}_i = \epsilon_i \cos\zeta_i, \quad \tilde{\phi}_i = k_i^{-1}\omega_i\epsilon_i e^{k_i y}\sin\zeta_i \quad (L_i = 1)$$

and making reductions of the kind $\sin\zeta\cos\zeta_i = \frac{1}{2}\{\sin(\zeta+\zeta_i)+\sin(\zeta-\zeta_i)\}$. Separating the components at wave-numbers $k \pm k_i$, we thus obtain pairs of simultaneous equations for A_i, C_i and B_i, D_i, solution of which gives

$$\left.\begin{aligned} A_{1,2} = 1, \quad B_{1,2} = 0, \\ C_{1,2} = 0, \quad D_{1,2} = \pm 1 \end{aligned}\right\} \tag{30}$$

if we neglect $O(\delta)$. This approximation is justified since these constants only appear multiplied by k^2a^2 in our final results: the neglect of $O(\delta)$ is therefore consistent with (28) and the extent of the approximation in powers of ka.

The next step is to separate components at the wave-numbers k_i from the boundary conditions (19) and (20), the approximation being taken to $O(\omega k^2a^2\epsilon_i)$ and $O(\omega^2ka^2\epsilon_i)$, respectively. The terms just evaluated now contribute to the right-hand sides, since, for instance, the product $\sin\zeta\cos(\zeta+\zeta_i)$ yields the component $-\frac{1}{2}\sin\zeta_i$. And, as a point of great importance, note is taken of the relationships shown by (2) in § 1. Thus, for instance, we must put

$$\begin{aligned} \sin 2\zeta \cos\zeta_{1,2} &= \tfrac{1}{2}\sin(2\zeta-\zeta_{1,2}) + \tfrac{1}{2}\sin(2\zeta+\zeta_{1,2}) \\ &= \tfrac{1}{2}\sin(\zeta_{2,1}+\theta) + \tfrac{1}{2}\sin(2\zeta+\zeta_{1,2}), \end{aligned}$$

ignoring the second component but recognizing that the first contributes at the other of the two wave-numbers. In this way, after a lot of reduction, (19) leads to the pair of equations

$$\begin{aligned} \epsilon_{1,2}\{\omega'_{1,2}(1-L_i)+\dot{\gamma}_{1,2}(1-M_{1,2})\}\sin\zeta_{1,2}+\dot{\epsilon}_{1,2}(1-N_{1,2})\cos\zeta_{1,2} \\ = \omega k^2a^2\{\tfrac{5}{4}\epsilon_{1,2}\sin\zeta_{1,2}+\tfrac{5}{8}\epsilon_{2,1}\sin(\zeta_{1,2}+\theta)\}, \end{aligned} \tag{31}$$

and (20) leads to the pair

$$\begin{aligned} \epsilon_{1,2}\{\omega_1'^{-1}(gk_{1,2}-\omega_{1,2}'^{2}L_{1,2})-\dot{\gamma}_{1,2}(1+M_{1,2})\}\cos\zeta_{1,2}+\dot{\epsilon}_{1,2}(1+N_{1,2})\sin\zeta_{1,2} \\ = -\omega k^2a^2\{\tfrac{3}{4}\epsilon_{1,2}\cos\zeta_{1,2}+\tfrac{3}{8}\epsilon_{2,1}\cos(\zeta_{1,2}+\theta)\}. \end{aligned} \tag{32}$$

Furthermore, the boundary conditions must be satisfied separately by components at the same wave-number but in quadrature. Hence each one of the equations (31) and (32) yields two simultaneous equations upon separation of the coefficients of $\sin\zeta_i$ and $\cos\zeta_i$. The constants L_i and M_i occur only in the combination $\omega'_i L_i + \dot{\gamma}_i M_i$, and so they can be eliminated between only four equations. Therefore, since the N_i are the only other unknown constants, the eight equations given by (31) and (32) can be reduced to four equations with known parameters for the four functions ϵ_i and γ_i.

Adding the coefficients of $\cos\zeta_i$ in (31) and those of $\sin\zeta_i$ in (32), we obtain

$$d\epsilon_{1,2}/dt = (\tfrac{1}{2}\omega k^2a^2\sin\theta)\epsilon_{2,1}. \tag{33}$$

The other components of (31) and (32) lead to

$$\frac{d\gamma_{1,2}}{dt} = \frac{1}{2}\left\{\frac{gk_{1,2}}{\omega'_{1,2}} - \omega'_{1,2}\right\} + \omega k^2a^2\left\{1 + \frac{1}{2}\frac{\epsilon_{2,1}}{\epsilon_{1,2}}\cos\theta\right\}, \tag{34}$$

and this pair of equations can be added to give an equation for $\theta = \gamma_1 + \gamma_2$; thus

$$\frac{d\theta}{dt} = \tfrac{1}{2}g\left(\frac{k_1}{\omega'_1} + \frac{k_2}{\omega'_2}\right) - \tfrac{1}{2}(\omega'_1 + \omega'_2) + \omega k^2a^2\left\{2 + \frac{\epsilon_1^2 + \epsilon_2^2}{2\epsilon_1\epsilon_2}\cos\theta\right\}. \tag{35}$$

We also have, substituting (26), (27) and then (13),

$$\begin{aligned}\tfrac{1}{2}g\left(\frac{k_1}{\omega'_1} + \frac{k_2}{\omega'_2}\right) - \tfrac{1}{2}(\omega'_1 + \omega'_2) &= \frac{gk}{\omega}(1 - \delta^2) - \omega \\ &= -\omega(k^2a^2 + \delta^2).\end{aligned} \tag{36}$$

The error in (36) is $O(\delta^4, k^4a^4, \delta^2k^2a^2)$, being therefore negligible in the present scheme of approximation. Hence (35) is reduced to

$$\frac{d\theta}{dt} = \omega k^2a^2\left\{1 + \frac{\epsilon_1^2 + \epsilon_2^2}{2\epsilon_1\epsilon_2}\cos\theta\right\} - \omega\delta^2. \tag{37}$$

This equation should be compared with (5), which was given by ignoring the effect of the basic wave train on the side-band properties.

Equations (33) and (37) express secular properties that the perturbed motion will have if the system is released (i.e. the water surface is kept at constant pressure) after an infinitesimal disturbance of the specified kind has been introduced. The stability or instability of the system to a disturbance of this kind can be decided, therefore, from the asymptotic behaviour of the ϵ_i determined by these equations. If, for some choice of finite initial values at $t = 0$, the property $|\epsilon_i| \to \infty$ for $t \to \infty$ is forthcoming, then the system is proved to be unstable in the usual sense. The question of stability as thus answered is, of course, separate from the practical question of what happens in the case of instability when a disturbance grows so large that the condition $\epsilon_i \ll a$ assumed here is no longer satisfied. The range over which the properties predicted by the present theory are manifested in practice, and the final outcome of instability, are matters that will be dealt with in Part 2.

Demonstration of instability

An integral of the simultaneous equations (33) is

$$\epsilon_{1,2}(t) = \epsilon_{1,2}(0)\cosh\left\{\tfrac{1}{2}\omega k^2a^2\int_0^t \sin\theta\, dt\right\} + \epsilon_{2,1}(0)\sinh\left\{\tfrac{1}{2}\omega k^2a^2\int_0^t \sin\theta\, dt\right\}. \tag{38}$$

Although θ is as yet an unknown function of t, several important conclusions can be drawn immediately from this expression. First, in confirmation of the property suggested in § 1, we see that both side-band amplitudes will undergo unbounded amplification if $\theta \to$ const. $(\neq 0, \pi)$ as $t \to \infty$. We also see that this behaviour can ensue even when one of the side-band components is absent initially: the missing component is subsequently generated by the interaction of the other component with the basic wave train, and the two amplitudes tend ultimately to be equalized by the process of amplification. (Some experimental observations of this situation, where only one component is present initially, will be described in Part 2.)

No generality is lost by assuming henceforth that the initial value $\epsilon_1(0)$ is positive, and that $\epsilon_2(0)$ is positive or zero. Since changing the sign of ϵ_i is equivalent to adding π to γ_i, the case where the $\epsilon_i(0)$ have opposite signs is obviously covered by allowing $\theta(0)$ any value in $(0, 2\pi)$. And, in respect of the equations now under consideration, the case where the $\epsilon_i(0)$ are both negative is indistinguishable from the specified case. Under the present assumption, (33) shows that the amplitudes grow steadily if θ begins and remains in $(0, \pi)$, but they diminish at first if $\theta(0)$ lies in the open interval $(\pi, 2\pi)$.

We proceed to derive an explicit solution for ϵ_1, first writing for convenience

$$T = k^2a^2\omega t, \quad \alpha = \frac{k^2a^2-\delta^2}{k^2a^2}. \tag{39}$$

After multiplication by $\epsilon_1\epsilon_2\sin\theta$, (37) then becomes

$$-\epsilon_1\epsilon_2\frac{d(\cos\theta)}{dT} = \alpha\epsilon_1\epsilon_2\sin\theta + \tfrac{1}{2}(\epsilon_1^2+\epsilon_2^2)\sin\theta\cos\theta. \tag{40}$$

But (33) gives

$$\frac{d\epsilon_1^2}{dT} = \frac{d\epsilon_2^2}{dT} = \epsilon_1\epsilon_2\sin\theta. \tag{41}$$

Hence (40) is seen to be equivalent to

$$d(\epsilon_1\epsilon_2\cos\theta + \alpha\epsilon_1^2)/dT = 0,$$

and so

$$\epsilon_1\epsilon_2\cos\theta + \alpha\epsilon_1^2 = \text{const.} = \rho, \quad \text{say.} \tag{42}$$

The equations (41) also show that

$$\epsilon_1^2-\epsilon_2^2 = \text{const.} = 2\alpha\rho(1-v), \quad \text{say.} \tag{43}$$

Both constants ρ and v are determined by the initial values of ϵ_1, ϵ_2 and θ.

Using (42) and (43) to express the right-hand side of (41) in terms of the variable ϵ_1 alone, we obtain

$$(d\epsilon_1^2/dT)^2 = (1-\alpha^2)\epsilon_1^4 + 2\alpha v\rho\epsilon_1^2 - \rho^2. \tag{44}$$

Let this quadratic function of ϵ_1^2 be denoted by Q for short. Since ϵ_1 must be real, the solution ϵ_1^2 of (44) is restricted to the range of positive values over which $Q > 0$, and a positive root of Q represents an extremum of $\epsilon_1^2(T)$. The two roots of Q may be expressed by $A \pm B$, where

$$A = -\frac{\alpha v\rho}{1-\alpha^2}, \quad B = \frac{\rho(1-\alpha^2+\alpha^2v^2)^{\frac{1}{2}}}{|1-\alpha^2|} \quad (>0). \tag{45}$$

Three cases may now be distinguished:

The case of instability: $-1 < \alpha < 1$

Only one root, $A+B$, is then positive, and any value of ϵ_1^2 greater than this makes $Q > 0$, so that unbounded growth of ϵ_1^2 with increasing T is possible.

Writing the quadratic in the form

$$Q = (1-\alpha^2)\{(\epsilon_1^2 - A)^2 - B^2\},$$

we obtain from (44) upon integration

$$(1-\alpha^2)^{\frac{1}{2}}T = \int_{\epsilon_1^2(0)}^{\epsilon_1^2} \{(\psi - A)^2 - B^2\}^{-\frac{1}{2}} d\psi = \left[\cosh^{-1}\left(\frac{\psi - A}{B}\right)\right]_{\epsilon_1^2(0)}^{\epsilon_1^2}.$$

And if the initial value of the expression on the right-hand side is denoted by $(1-\alpha^2)^{\frac{1}{2}}\tau$, this result can be rearranged to give

$$\epsilon_1^2 = A + B\cosh\{(1-\alpha^2)^{\frac{1}{2}}(T+\tau)\}. \tag{46}$$

We reflect that the constants A, B and τ in (46) are all determined by the initial values of ϵ_1, ϵ_2 and θ. Also, referring to (41), we see that τ will have the sign of $\sin\theta(0)$. If $\tau < 0$, the solution ϵ_1^2 will first decrease with increasing T, reach the minimum value $A+B$ when $T = -\tau$, and thereafter increase steadily without bound. If $\tau > 0$, there will be steady growth from the start. Comparing (38), we may conclude from these results that θ always remains in $(0, \pi)$ if its initial value lies in this interval, but θ generally leaves $(\pi, 2\pi)$ if started there.

For large T, the asymptotic behaviour of ϵ_1 according to (46) is

$$\epsilon_1 \sim \exp\{\tfrac{1}{2}(1-\alpha^2)^{\frac{1}{2}}T\}. \tag{47}$$

From (38) or (43) it follows that $\epsilon_2 \to \epsilon_1$ concomitantly with (47). And it is readily seen that $\theta \to \cos^{-1}(-\alpha)$ in $(0, \pi)$.

An exception to the preceding conclusions must be allowed in the instance when $\epsilon_1(0) = \epsilon_2(0)$ and $\theta(0) = \cos^{-1}(-\alpha)$. Then both ρ and $v\rho$ vanish, θ is constant by (42) and (43), and the solution of (44) is seen immediately to be

$$\epsilon_1 = \epsilon_1(0)\exp\{\pm\tfrac{1}{2}(1-\alpha^2)^{\frac{1}{2}}T\}. \tag{48}$$

The positive sign applies when $0 < \theta(0) < \pi$; thus, as is obviously to be expected for these initial conditions, the properties achieved asymptotically in the general case are manifested from the start. The negative sign applies when $\pi < \theta(0) < 2\pi$, and thus we have one special set of initial conditions for which the disturbance will die away. This form of disturbance is incapable of practical realization, of course, being in effect an '*unstable*' singular solution of the system of differential equations.†

The case of marginal instability: $\alpha = -1$

The right-hand side of (44) reduces in this case to a linear function of ϵ_1^2, with $\alpha v\rho = -v\rho > 0$ (cf. (50) below). The equation is then easily solved, giving

$$\epsilon_1^2 = -\tfrac{1}{2}v\rho\{(1/v^2) + (T+\tau)^2\}. \tag{49}$$

† A familiar interpretation of the two exceptional solutions (48) can be made if the system of differential equations, (33) and (37), is reduced to a pair of equations in the two dependent variables $E = \epsilon_1/\epsilon_2$ and θ. The values $E = 1$, $\theta = \cos^{-1}(-\alpha)$ then define two singular points of the system: the first, with θ in $(0, \pi)$, is a 'stable node' towards which all neighbouring solutions converge as $T \to \infty$, whereas the second, with θ in $(\pi, 2\pi)$, is an 'unstable node' (cf. Stoker 1950, p. 44).

Hence we have that $\epsilon_1 \sim T$ asymptotically for large T, and this unbounded linear growth must be classed as an instability.

The case of stability: $\alpha < -1$

The quadratic function Q then has two positive roots, and the solution ϵ_1^2 of (44) must range between them. It is not obvious from the original definitions of ρ and v that the coefficients of Q have the properties

$$v\rho < 0, \quad 1-\alpha^2+\alpha^2 v^2 \geqslant 0, \tag{50}$$

which, in addition to $\alpha < -1$, are necessary for positive real roots; but the fact that (44) is satisfied by the arbitrary initial values of ϵ_1^2 and $d\epsilon_1^2/dT$ is enough to establish these properties. Thus we have that $A > B \geqslant 0$, and by putting $(1-\alpha^2)^{\frac{1}{2}} = i(\alpha^2-1)^{\frac{1}{2}}$ in the previous solution (46) we obtain for the present case

$$\epsilon_1^2 = A + B\cos\{(\alpha^2-1)^{\frac{1}{2}}(T+\tau)\}. \tag{51}$$

This shows, as expected, that ϵ_1^2 varies periodically between the finite extrema $A \pm B$, and hence (43) shows that ϵ_2^2 also remains bounded. The system is therefore stable to disturbances as now specified.

For the particular choice of initial values

$$\theta(0) = 0, \quad \epsilon_2(0) = \{-\alpha \pm (\alpha^2-1)^{\frac{1}{2}}\}\epsilon_1(0), \tag{52}$$

it is found that the equality included in the second member of (50) holds, so that $B = 0$. Then all three variables remain constant for all T. These two quiescent solutions may be described as the singular points of the system of differential equations in the case $\alpha < -1$, being of the type commonly termed *centres* (cf. Stoker 1950, pp. 38, 44).

3. Discussion

In §3, the stability of Stokes wave trains on infinitely deep inviscid liquid has been investigated by means of a linearized perturbation analysis. The condition of instability—that is, the condition under which the amplitudes of the side-band wave modes composing the disturbance appear to undergo indefinitely great magnification—was established as $-1 \leqslant \alpha < 1$, which means, according to the definition of α in (39),

$$0 < \delta \leqslant (\surd 2)ka. \tag{53}$$

In principle, therefore, all wave trains of the type specified are unstable, since in accordance with the assumptions of the analysis a choice of δ satisfying (53) can be made for every finite value of ka, however small. But an obvious reservation must be made regarding the practical application of the theory: since viscous damping rates are approximately independent of wave amplitude, the effect of dissipation can be expected to suppress the instability if ka is sufficiently small. The practical aspects are to be taken up in Part 2, and for now it will suffice to note that (53) has been confirmed very closely as the natural condition of instability for water waves with $ka = O(10^{-1})$ and wavelengths about 1 ft. or greater.

Another result with especial interest, having been checked experimentally, is the asymptotic growth of the side-band amplitudes. From (47), after the definitions of T and α have been substituted, we have

$$\epsilon_i \sim \exp\{\tfrac{1}{2}\delta(2k^2a^2-\delta^2)^{\frac{1}{2}}\omega t\} \tag{54}$$

if $0 < \delta < (\sqrt{2})ka$. For a given ka, the exponent in (54) is a maximum at $\delta = ka$; that is, the optimum frequency deviation $\omega' - \omega$ of the side bands is a fraction $1/\sqrt{2}$ of the cut-off value beyond which no amplification occurs. Thus for a

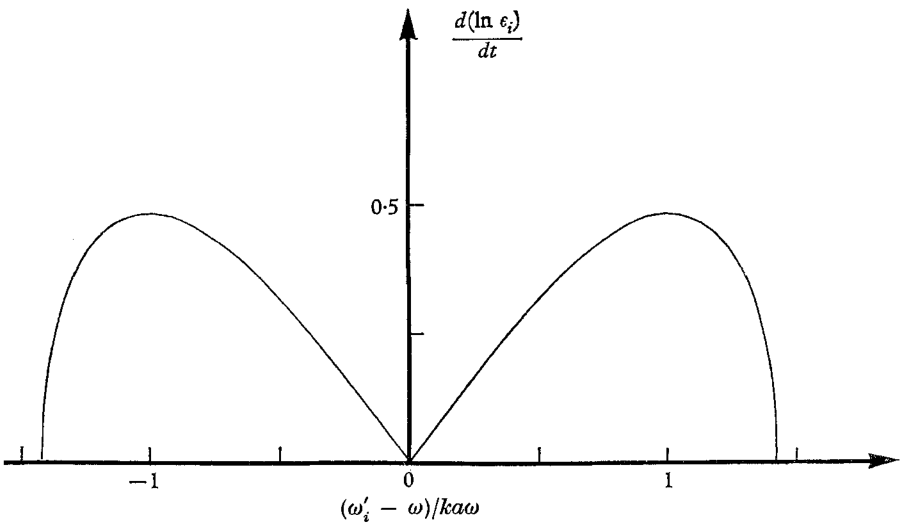

FIGURE 1. The asymptotic growth rate of the side-band amplitudes as a function of frequency.

specific Stokes wave there is a disturbance that is most unstable, comprising a pair of side-band modes with frequencies $\omega_i' = \omega(1 \pm ka)$ and wave-numbers $k_i = k(1 \pm 2ka)$ [see (26) and (27)]. This disturbance will emerge distinctly by selective amplification if the instability develops from a sufficiently low level of random background noise. To emphasize the present facts, particularly the existence of optimum and cut-off frequencies, the logarithmic growth rate given by (54) is plotted as a function of ω_i' in figure 1.

When the side-band amplitudes become approximately equal, as they always do after sufficient amplification, the assumed form of disturbance is equivalent to a uniform, forward-travelling modulation of the primary wave train. To demonstrate this point, let $\bar{\epsilon}$ denote the absolute value of the displacement amplitude for each side-band component (i.e. $\bar{\epsilon} \equiv \epsilon\epsilon_i$ according to the strict meanings of ϵ and ϵ_i as introduced in §2), and consider the sum of the leading terms of the primary displacement H and perturbation $\epsilon\tilde{\eta}$. Thus, taking account of (11), (21) and (22), we express the second of (14) in the approximate form

$$\begin{aligned}\eta &= a\cos\zeta + \bar{\epsilon}(\cos\zeta_1 + \cos\zeta_2)\\ &= a\cos(kx-\omega t) + 2\bar{\epsilon}\cos(\kappa kx - \delta\omega t)\cos(kx - \omega t - \tfrac{1}{2}\theta)\\ &= \{a + 2\bar{\epsilon}\cos\tfrac{1}{2}\theta\cos(\kappa kx - \delta\omega t)\}\cos(kx-\omega t)\\ &\qquad + 2\bar{\epsilon}\sin\tfrac{1}{2}\theta\cos(\kappa kx - \delta\omega t)\sin(kx-\omega t).\end{aligned} \tag{55}$$

Here the terms in $\bar{\epsilon}$ represent a gradual modulation, at wave-number $\kappa k \ll k$, of the fundamental $a\cos(kx-\omega t)$: the first term, proportional to $\cos\frac{1}{2}\theta$, describes an amplitude modulation; and the second, in which the rapidly-varying factor is in quadrature with the fundamental, describes a phase modulation. The modulation envelope advances at the group velocity $c_g = \delta\omega/\kappa k = \frac{1}{2}c$. Note that the type of modulation, amplitude or phase, changes from one extreme to the other as δ is varied through the unstable range given by (53). For, recalling that the asymptotic value of θ is $\cos^{-1}(-\alpha)$ in $(0,\pi)$, we deduce that as δ is raised from near zero up to the cut-off value $(\sqrt{2})ka$, $\frac{1}{2}\theta$ falls from near $\frac{1}{2}\pi$ (mainly phase modulation) to zero (pure amplitude modulation).

As is illustrated clearly by (55), the assumed physical model of the perturbed wave train is spatially uniform and its features (e.g. the depth of the modulations in amplitude and phase) develop simultaneously everywhere, having been expressed in the analysis as slowly-varying functions of t alone. In an alternative model, the non-linear processes are assumed to act for an unlimited time, but to originate from a certain position—say $x = 0$—and so to develop with distance x in the direction of propagation of the primary wave train. This is the situation depicted in the first paragraph of the summary beginning this paper, and it is the closest model for the experiments described in Part 2. By a well-known principle, the present results can be adapted simply to apply to this second state of affairs: we need only to replace t in (54) by $x/c_g = 2kx/\omega$. (Various formal arguments are available to justify this step: for instance, see Gaster (1962).) Thus, for the spatial growth of side-band modes that are generated with fixed frequencies $\omega(1\pm\delta)$ at $x = 0$, we have

$$\epsilon_i \sim \exp\{\delta(2k^2a^2-\delta^2)^{\frac{1}{2}}kx\}. \tag{56}$$

The condition of instability is (53) as before.

The present theory has been extended to deal with waves on water of arbitrary depth h, and the details will be reported in another paper (Benjamin 1967). It is found that a Stokes wave train is unstable, in the manner described here, only if $kh > 1{\cdot}363$. The crucial property (3) turns out to be impossible if $kh < 1{\cdot}363$, and the consequent inability of the interaction between disturbance modes and basic wave train to become resonant is sufficient to ensure stability. This significant division of the range of kh has also been discovered by Whitham (1966) in applying his non-linear theory of wave dispersion to Stokes waves. He showed that equations governing extremely gradual—but not necessarily small—variations in wave properties are of elliptic type if $kh > 1{\cdot}363$, but are hyperbolic if $kh < 1{\cdot}363$. Some comments on the connexion between Whitham's theory and the present stability theory are included in his paper (see also Benjamin 1967).

REFERENCES

Benjamin, T. B. 1967 Instability of periodic wave trains in non-linear dispersive systems. *Proc. Roy. Soc.* A (to appear).

Benjamin, T. B. & Feir, J. E. 1967 The disintegration of wave-trains on deep water. Part 2. Experiments. *J. Fluid Mech.* (to appear).

Benney, D. J. 1962 Non-linear gravity wave interactions. *J. Fluid Mech.* **14**, 577.

GASTER, M. 1962 A note on the relation between temporally-increasing and spatially-increasing disturbances in hydrodynamic stability. *J. Fluid Mech.* **14**, 222.

KRASOVSKIĬ, Y. P. 1960 On the theory of steady waves of not small amplitude (in Russian). *Dokl. Akad. Nauk SSSR* **130**, 1237.

KRASOVSKIĬ, Y. P. 1961 On the theory of permanent waves of finite amplitude (in Russian). *Zh. vȳchisl. Mat. mat. Fiz.* **1**, 836.

LAMB, H. 1932 *Hydrodynamics*, 6th ed. Cambridge University Press.

LEVI-CIVITA, T. 1925 Détermination rigoureuse des ondes permanentes d'ampleur finie. *Math. Ann.* **93**, 264.

LONGUET-HIGGINS, M. S. & PHILLIPS, O. M. 1962 Phase velocity effects in tertiary wave interactions. *J. Fluid Mech.* **12**, 333.

STOKER, J. J. 1950 *Nonlinear vibrations*. New York: Interscience.

STRUIK, D. J. 1926 Détermination rigoureuse des ondes irrotationnelles périodiques dans un canal à profondeur finie. *Math. Ann.* **95**, 595.

WHITHAM, G. B. 1966 Non-linear dispersion of water waves. *J. Fluid Mech.* **27**, 399.