Instability of periodic wavetrains in nonlinear dispersive systems

By T. B. Benjamin, F.R.S.

Department of Applied Mathematics and Theoretical Physics, University of Cambridge

[Plate 1]

The defining property of the class of physical systems under consideration herein is that, by striking a balance between nonlinear and frequency-dispersive effects, they can transmit periodic waves of finite amplitude but constant form. For any such system, therefore, in respect of propagation in the x direction relative to a state of rest, the dynamical equations have exact periodic solutions of the form $\eta(x, t) = H(x-ct)$, say, where c is a constant phase velocity depending on wave amplitude as well as on frequency or wavelength. This paper is concerned with the proposition that in many cases these uniform wavetrains are unstable to small disturbances of a certain kind, so that in practice they will disintegrate if the attempt is made to send them over great distances. The outstanding example only recently brought to light is that finite gravity waves on deep water are unstable: unmistakable experimental evidence of this property is now available, and it has also been demonstrated analytically.

In §2 the essential factors leading to instability are explained in general terms. A disturbance capable of gaining energy from the primary wave motion consists of a pair of wave modes at side-band frequencies and wavenumbers fractionally different from the fundamental frequency and wavenumber. In consequence of a nonlinear effect on these modes counteracting the detuning effect of dispersion on them, they are forced into resonance with second-harmonic components of the primary motion and thereafter their amplitudes grow mutually at a rate that is exponential in time or distance. In §3 a detailed stability analysis is presented for wavetrains on water of arbitrary depth h, and it is shown that they are unstable if the fundamental wavenumber k satisfies $kh > 1\cdot363$, but are otherwise stable. Finally, in §4, some experimental results regarding the instability of deep-water waves are discussed, and a few prospective applications to other specific systems are reviewed.

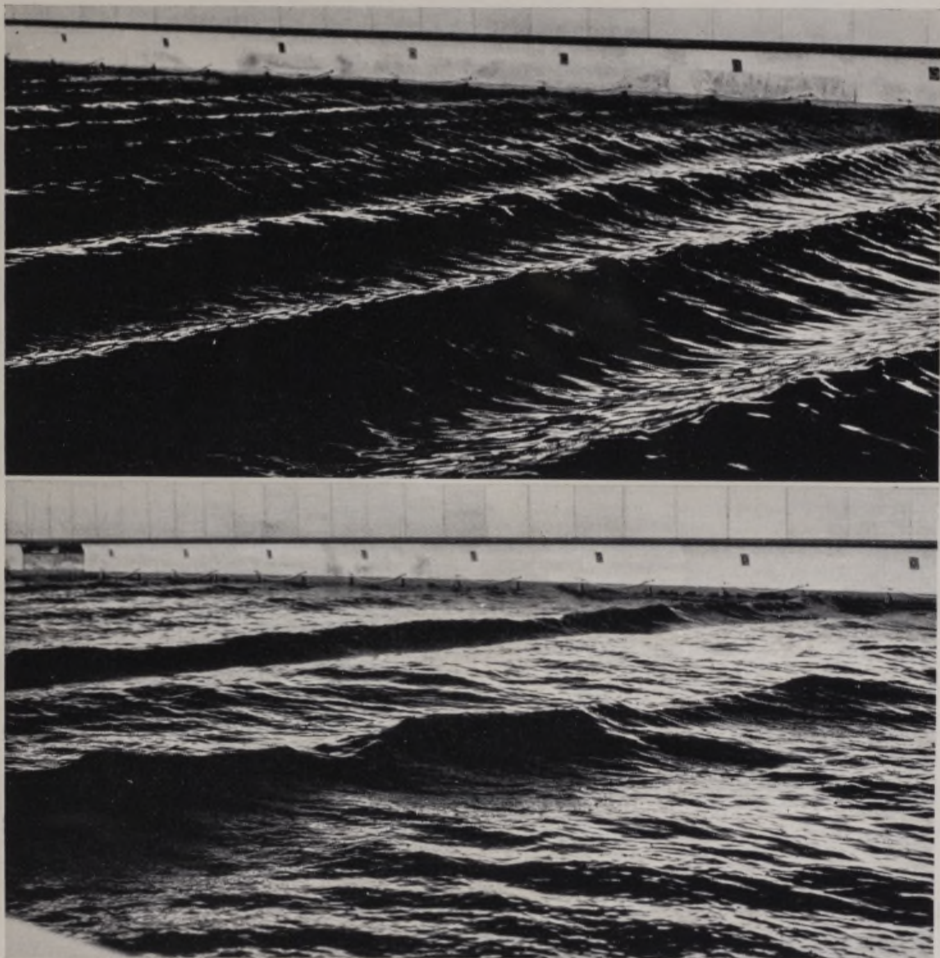
1. Introduction

This contribution to the Discussion focuses on the question whether uniform wavetrains in nonlinear dispersive systems are stable or not. The archetypal problem in view concerns straight-crested wavetrains on an infinite sheet of water over a horizontal bottom, but reference will be made later to several other physical problems upon which present ideas have bearing. One of my aims is to review the experimental and mathematical investigation that has been made in the last two years by Mr J. E. Feir and myself on the instability and consequent disintegration of wavetrains on deep water. A preliminary account of our work was presented at the Seventh British Theoretical Mechanics Colloquium in Leeds a year ago, and the details are reported in two papers by Benjamin & Feir (1967*a*, *b*) which will appear in the *Journal of Fluid Mechanics*. However, several extensions of this work are to be reported now for the first time: in particular, I shall show what has been found in applying a stability analysis to waves on water of arbitrary depth.

To fix the ideas at issue, the following familiar facts need first be recalled. For many wave systems that are both nonlinear and dispersive, a distinctive property is that the dynamical equations admit periodic, though nonsinusoidal, solutions representing progressive wavetrains of finite amplitude but with steady waveform. These solutions imply a balance between the effects of nonlinearity and of frequency dispersion, because in the absence of either factor such permanent wavetrains are generally impossible. In the case of water waves, the existence of these solutions was assumed by many writers since the earliest days of the subject, and it was supposed that they could be described by a series expansion in powers of the wave amplitude. This expansion is usually named after Stokes, who first worked out its leading terms. Definite proof of the existence of permanent water waves nevertheless presents a formidable task of analysis, and around the turn of this century there was a vigorous controversy on this point: that is, real doubt was entertained about the convergence of the Stokes expansion. The question was eventually settled in a famous paper by Levi-Civita (1925). For waves on infinitely deep water, he proved that the Stokes expansion converges when the ratio of amplitude to wavelength is sufficiently small, thus demonstrating that the nonlinear boundary conditions in the water-wave problem can be satisfied exactly by a permanent wavetrain. The proof was extended by Struik (1926) to small-amplitude waves on water of arbitrary depth, and the recent work of Kraskovskiĭ (1960, 1961) has finally established the existence of permanent periodic waves for all amplitudes less than the extreme at which the waves take a sharp-crested form. However, despite the effort given to proving the theoretical possibility of water waves with permanent shape, the separate question of their stability appears to have been ignored until now—except for a tentative study by Korteweg & de Vries in 1895 concerning long waves in shallow water. The surprising fact now brought to light is that the Stokes waves on sufficiently deep water are definitely unstable.

The practical implications of this fact have been demonstrated in the experiments reported by Benjamin & Feir (1967*b*). Deep-water wavetrains of fairly large amplitude (but still considerably smaller than white-capped waves) were generated at one end of a long tank and were observed travelling many wavelengths. It was found that such a wavetrain may develop conspicuous irregularities if it travels far enough, even when departures from periodicity can hardly be detected near the origin. And eventually, at great distances from the origin, the train may become completely disintegrated and its energy redistributed over a broad spectrum. The initial phase of these events, during which the irregularities are amplifying but are still fairly small, is amenable to analysis and is now more or less fully understood. Examples of the agreement that has been established between theory and observation will be given later.

The possible severity of the effects arising from instability is made clear by figure 1 (plate 1). The two photographs are views of a wavetrain in the large towing-basin (No. 3 Tank) of Ship Division, National Physical Laboratory, at Feltham. The fundamental wavelength was 7·2 ft., and the depth of water 25 ft. Figure 1 (*a*) shows the wavetrain close to the wavemaker at one end of the basin: here it was in a manifestly regular condition, except for small-scale roughness which is inevitably present when waves are produced on this scale and which is immaterial to the processes now in question. In contrast, figure 1 (*b*) shows the same wavetrain at a



(*a*)

(*b*)

FIGURE 1. Photographs of a progressive wavetrain at two stations, illustrating disintegration due to instability: (*a*) view near to wavemaker; (*b*) view at 200 ft. farther from wavemaker. Fundamental wavelength, 7·2 ft.

distance of 200 ft. (28 wavelengths) farther along the tank, and it can be seen that drastic distortions had occurred. Admittedly, this case was unnatural in that small disturbances in the form of discrete unstable modes were introduced by modulating the reciprocating motion of the wavemaker (more will be said about this later), and the results shown in figure 1 should definitely not be taken to reflect unfavourably on the N.P.L.'s excellent equipment. But in any sufficiently long wave-tank large distortions will also arise spontaneously from random background disturbances, once the wave amplitude is raised above a certain level at which the instability mechanism, depending on nonlinear effects and therefore intensifying with increasing amplitude, is not suppressed by the action of viscosity. This phenomenon has probably been noticed by many wave-tank operators in the past (it was certainly known to the staff of Ship Division, N.P.L.), even though the level of wave amplitudes at which it becomes strongly manifest is somewhat higher than that usual in ship-model testing.

The instability of wavetrains on deep water presents another remarkable facet, revealed in the recent theoretical studies by Lighthill (1965) and Whitham (1967). In their work equations were derived governing extremely gradual—but not necessarily small—variations in wave properties, and these equations for deep-water waves were shown to be of elliptic type. The encounter with elliptic equations with time as an independent variable is surprising at first, since an intuitively obvious aspect of the physical problem is the admissibility of arbitrary Cauchy conditions (i.e. hyperbolic boundary conditions) at an initial instant. On reflection from a physical viewpoint, the proper interpretation appears to be that if some very gradual distortion of a wavetrain is set going, it must first (as the fact of elliptic governing equations indicates) get steadily more severe without the wave properties passing through any extremum; inevitably, therefore, the wavetrain becomes in due course so much distorted that the theory breaks down. Thus, uniform validity of this approximate theory for an indefinitely long time is out of the question, but for a limited time one has a description of the unidirectional developments from an unstable situation. In this sense Lighthill's and Whitham's findings are perfectly in accord with the results of the instability theory to be outlined below, which shows that infinitesimal disturbances of a certain kind can undergo unbounded amplification. Balancing the advantage of their theories in dealing with large perturbations, the advantage of the present theory is a greater precision in describing the behaviour of small unstable perturbations. In particular, we find conditions for maximum amplification rates, and cutoff conditions beyond which a perturbation has stable behaviour.

The ingredients of this theory resemble a number of existing results in the theory of resonant interactions between water waves, as has been developed by Professor O. M. Phillips and others in recent years. But these ingredients need to be assembled rather carefully; and so, while the alliance with previous work on resonant interactions is freely acknowledged, the present analytical study makes its own way from first principles. First, in §2, the mechanism of instability is discussed generally in simple terms, and then in §3 a detailed stability analysis for water waves is presented.

2. The explanation for instability

In any system of the general type under discussion, let there be transmitted in the x direction a steady periodic wavetrain whose fundamental simple-harmonic component has (displacement) amplitude a and argument $\zeta = kx - \omega t$. Since the system is nonlinear, there must also be present harmonics with arguments $2\zeta, 3\zeta, \ldots$, all of which advance with the same phase velocity $c = \omega/k$ as the fundamental; and their respective amplitudes may generally be assumed to decrease in relative order of magnitude like successive integral powers of ka, which is taken to be a small fraction. Now consider that a disturbance is introduced consisting of two progressive wave modes with 'side-band' frequencies and wavenumbers adjacent to ω and k, so that their arguments have the forms expressed in the first and last lines of the accompanying diagram. Here κ and δ are understood to be small fractions and the amplitudes ϵ_1 and ϵ_2 of these modes are assumed to be much smaller than a. The diagram demonstrates the forms of two particular products arising from the nonlinear interaction between the disturbance and the primary wavetrain; specifically, these are the difference components generated between the side bands and the second harmonic.

Interaction diagram showing arguments { } and magnitudes of amplitudes [ ] of simple-harmonic components

$$\begin{array}{ll}
\textit{Upper side band} & \\
\quad \{\zeta_1 = k(1+\kappa)\,x - \omega(1+\delta)\,t - \gamma_1\}, [\epsilon_1] & \\
\textit{2nd harmonic of primary wavetrain} & \{\zeta_2 + (\gamma_1+\gamma_2)\}, [k^2a^2\epsilon_1], \\
\quad \{2\zeta = 2(kx-\omega t)\}, [ka^2] & \\
\textit{Lower side band} & \{\zeta_1 + (\gamma_1+\gamma_2)\}, [k^2a^2\epsilon_2]. \\
\quad \{\zeta_2 = k(1-\kappa)\,x - \omega(1-\delta)\,t - \gamma_2\}, [\epsilon_2] &
\end{array}$$

The vital fact revealed in the diagram is that, if

$$\theta = \gamma_1 + \gamma_2 \to \text{const.} \tag{1}$$

as the nonlinear processes develop, then the pair of interactions becomes mutually resonant. Each side-band mode thereafter suffers a synchronous forcing effect proportional to the amplitude of the other, and so the two amplitudes can grow exponentially. Thus the primary wave motion is unstable to this form of disturbance.

Realization of the property (1), which is necessary for instability, must derive from further effects of the primary motion upon the side-band modes. For, if the amplitude a were too small for there to be significant nonlinear coupling, then the net wavenumbers $\hat{k}_{1,2} = \partial\zeta_{1,2}/\partial x$ and frequencies $\hat{\omega}_{1,2} = -\partial\zeta_{1,2}/\partial t$ of these modes would satisfy the dispersion relation given by linearized theory, say

$$\hat{\omega} = f(\hat{k}). \tag{2}$$

To a first approximation for small κ and δ, evidently (2) is satisfied with constant γ_1 and γ_2 if

$$\frac{\delta\omega}{\kappa k} = c_g, \tag{3}$$

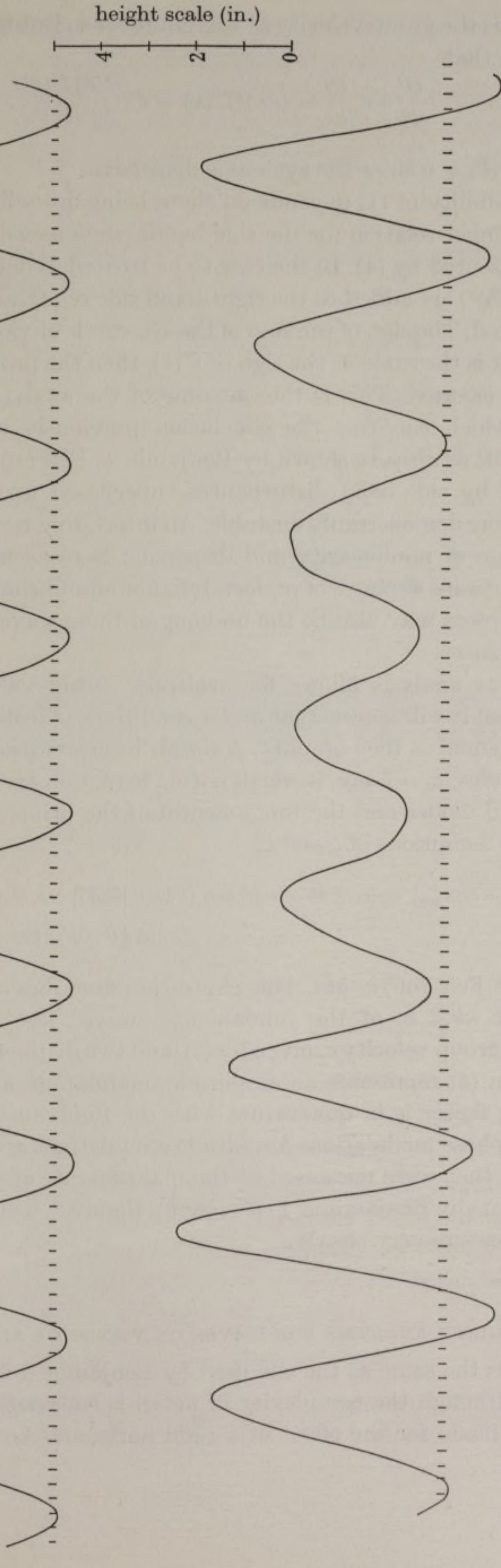


FIGURE 2. Experimental records of height of water surface as a function of time at two stations, showing spontaneous development of instability from background noise. Fundamental frequency, 0·85 c/s. Upper record taken at 200 ft. from wavemaker, lower at 400 ft. Interval between time marks, 0·1 s.

where $c_g = f'(k)$ is the group velocity at wavenumber k. But to a second approximation (2) requires that

$$\frac{\partial\theta}{\partial t} + c_g\frac{\partial\theta}{\partial x} = (\kappa k)^2 f''(k) = \delta^2\frac{f^2(k)f''(k)}{[f'(k)]^2}, \tag{4}$$

and generally $f''(k) \neq 0$ since the system is dispersive.

Thus the possibility of (1) depends on there being a nonlinear effect on the frequency-wavenumber relation for the side bands, such as will balance the effect of dispersion represented by (4). In the case to be treated in detail below, it turns out that terms $O(\omega k^2 a^2)$ are added to the right-hand side of (4), and this result is generally to be expected. The sign of the sum of these terms is obviously crucial as regards stability; for if it is the same as the sign of $f''(k)$, then the property (1) is impossible and stability is ensured. This is the outcome of the analysis respecting shallow-water waves, which confirms the conclusion previously reached by Whitham (1965, 1967). But, as already shown by Benjamin & Feir (1967*a*), the property (1) can be realized by side-band disturbances superposed upon deep-water waves, which are therefore demonstrably unstable. An interesting reflection is that whereas the counteraction of nonlinearity and dispersion is essential to the existence of permanent wavetrains as states of perfect dynamic equilibrium, opposition between the same two factors may also be the undoing of these wavetrains in the presence of small disturbances.

The subsequent analysis allows for arbitrary initial values of the side-band amplitudes ϵ_1, but it will appear that under conditions of instability the amplitudes tend to become equal as they amplify. A simple interpretation possible in the case of equal amplitudes, $\epsilon_i = \bar{\epsilon}$ say, is worth noting here. Considering the superposition of two side-band modes and the fundamental of the primary wavetrain, we may obtain from the definitions of ζ and ζ_i

$$\begin{aligned} a\cos\zeta + \bar{\epsilon}(\cos\zeta_1 + \cos\zeta_2) &= \{a + 2\bar{\epsilon}\cos\tfrac{1}{2}\theta\cos(\kappa kx - \delta\omega t)\}\cos(kx - \omega t) \\ &\quad + 2\bar{\epsilon}\sin\tfrac{1}{2}\theta\cos(\kappa kx - \delta\omega t)\sin(kx - \omega t) \end{aligned} \tag{5}$$

(cf. Benjamin & Feir 1967*a*, §3). This expression describes a gradual modulation, at wavenumber $\kappa k \ll k$, of the fundamental wave. The modulation envelope advances at the group velocity c_g given by (3), and two distinct effects are described: the first term in (5) represents an amplitude modulation, and the second, whose rapidly varying factor is in quadrature with the fundamental, is approximately equivalent to a phase modulation. Amplitude modulations are more easily detected in practice, and they were measured as the main feature of developing instability in the experiments by Benjamin & Feir (1967*b*). Figure 2, which will be discussed in §4, shows this feature very clearly.

2. Stability analysis for waves on water of arbitrary depth

The method is the same as the one used by Benjamin & Feir (1967*a*) for deep-water waves, although the complexity of detail is considerably increased by the allowance now made for the effect of a rigid horizontal bottom. Since the main steps are duplicated, the present account will keep to bare essentials and reference may be made to the previous paper for various points of supporting discussion.

As illustrated in figure 3, the axis x is drawn horizontally and y vertically upwards. The origin is taken at the level of the water in an undisturbed condition, when its depth is h, so that $y = -h$ denotes the bottom. The free surface is denoted by

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

The water is modelled as a perfect fluid whose motion is irrotational. Accordingly there is a velocity potential $\phi(x, y, t)$ satisfying

$$\phi_{xx} + \phi_{yy} = 0 \tag{7}$$

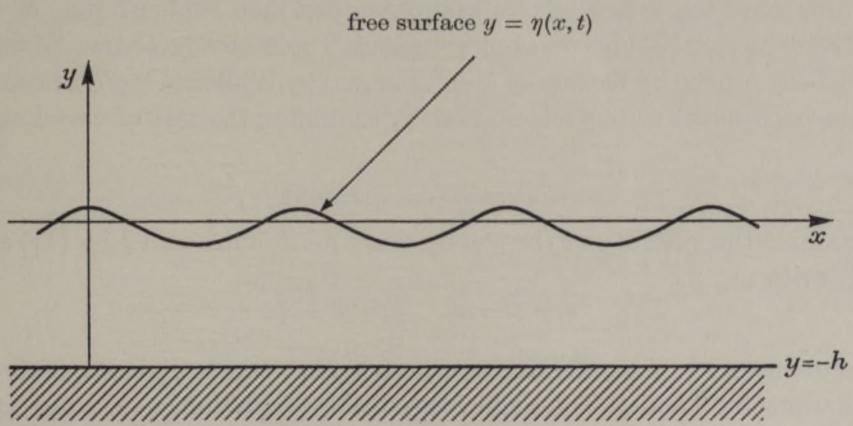


FIGURE 3. Illustration of coordinate system.

and subject to the boundary condition

$$\phi_y = 0 \quad \text{on} \quad y = -h. \tag{8}$$

At the free surface, the kinematical condition is

$$\eta_t + \eta_x[\phi_x]_{y=\eta} - [\phi_y]_{y=\eta} = 0, \tag{9}$$

and the condition of constant pressure is

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

(cf. Stoker 1957, §1·4).

Stokes waves of permanent form

It is known that the preceding equations have exact solutions in the form $\eta = H(x-ct)$, $\phi = \Phi(x-ct, y)$, where c is a constant phase velocity. The second approximation to these solutions as given originally by Stokes (1847, p. 205), supplemented by the second approximation to c (see, for example, Bowden 1948, p. 424), is sufficient for our purpose. Thus we consider

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

$$\phi = \Phi = \frac{\omega a}{k}\frac{\cosh k(y+h)}{\sinh K}\sin\zeta + \omega a^2 Q\frac{\cosh 2k(y+h)}{\sinh 2K}\sin 2\zeta, \tag{12}$$

in which

$$\left.\begin{aligned} &\zeta = kx-\omega t, \quad K = kh,\\ &\bar{\eta} = ka^2\Delta = -\tfrac{1}{2}ka^2 \operatorname{cosech} 2K,\\ &P = \tfrac{1}{4}\coth K(3\coth^2 K - 1), \quad Q = \tfrac{3}{4}\coth K \operatorname{cosech}^2 K,\end{aligned}\right\} \tag{13}$$

and

$$\omega^2 = k^2c^2 = gk\tanh K\left\{1+k^2a^2\left(1+\frac{1}{\sinh^2 K}+\frac{9-2\tanh^2 K}{8\sinh^4 K}\right)\right\}. \tag{14}$$

The validity of this approximation requires that $ka \ll 1$, or alternatively that $ka \ll K^3$ if K be small. The latter requirement places an excessive restriction on the fundamental amplitude a when the wavelength greatly exceeds the depth of water, and in this case the cnoidal-wave approximation due to Korteweg & de Vries (1895; Lamb 1932, §253) becomes more useful. The stability of cnoidal waves has virtually been proved by Korteweg & de Vries and by Whitham (1965), however, and so nothing of interest will be missed here by excluding the case of very long waves.

Perturbation equations

To examine the stability of the steady wave motion described by (11) and (12), we may write

$$\eta = H + \iota\tilde{\eta}, \quad \phi = \Phi + \iota\tilde{\phi}, \tag{15}$$

and, assuming ι to be a small number whose square can be ignored, derive linearized equations for $\tilde{\eta}$ and $\tilde{\phi}$. The coefficients of these equations need to be evaluated as far as terms in a^2. On the basis of the ideas explained in §2, the perturbation is assumed to consist of a pair of side-band modes, together with the products of their interaction with the primary wavetrain. Thus, defining the arguments ζ_i ($i = 1, 2$) as before, we take

$$\tilde{\eta} = \tilde{\eta}_1 + \tilde{\eta}_2 \tag{16}$$

with

$$\tilde{\eta}_i = \epsilon_i \cos\zeta_i + ka\{A_i\cos(\zeta+\zeta_i) + B_i\cos(\zeta-\zeta_i)\} + O(k^2a^2\epsilon_i). \tag{17}$$

The implied terms that are $O(k^2a^2\epsilon_i)$ have arguments $2\zeta+\zeta_i$ and do not need to be considered further, since they represent nonresonant, and therefore passive, effects of the interaction. Terms with the same order of magnitude and arguments $2\zeta-\zeta_i$ are vitally important but, as will appear later, they can be merged with terms having arguments ζ_i.

The small parameters κ and δ measuring the wavenumber and frequency deviations of the side bands are assumed to satisfy (3), where now c_g is the group velocity of infinitesimal water waves at wavenumber k. The required expression for c_g is obtainable from (14), when the correction for finite amplitude (i.e. the terms in k^2a^2) is omitted. Thus we obtain

$$\delta/\kappa = \tfrac{1}{2}(1+2K\operatorname{cosech} 2K) = \lambda, \quad \text{say}. \tag{18}$$

In anticipation of the final results, it is further assumed that

$$\kappa, \delta = O(ka), \tag{19}$$

and various coefficients depending on κ or δ will be simplified consistently with the adopted order of approximation in powers of ka. In keeping with these assumptions and again to be confirmed by the final results, the ϵ_i and γ_i are taken to be slowly varying functions of time alone, whose first derivatives have the properties

$$\dot{\epsilon}_i = O(\omega k^2 a^2 \epsilon_i), \quad \dot{\gamma}_i = O(\omega k^2 a^2). \tag{20}$$

The present model thus takes the characteristics of the modulation superposed on the primary wave train to be spatially uniform and so to develop simultaneously everywhere. The specification of only time-dependence is a convenient simplification for the analysis, but the generalization to cover x-dependence also, as already considered in §2, is easily accomplished later.

Corresponding to (16) and (17), the appropriate form of $\tilde{\phi}$ such that (7) and (8) are satisfied by the total velocity potential is

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

in which

$$\tilde{\phi}_i = \frac{\cosh k_i(y+h)}{k_i \sinh k_i h}\{\epsilon_i(\omega_i L_i + \dot{\gamma}_i M_i)\sin\zeta_i + \dot{\epsilon}_i N_i \cos\zeta_i\} + \omega a \epsilon_i \left\{C_i \frac{\cosh|(k+k_i)(y+h)|}{\sinh|(k+k_i)h|}\right.$$
$$\left. \times \sin(\zeta+\zeta_i) + D_i \frac{\cosh|(k-k_i)(y+h)|}{\sinh|(k-k_i)h|}\sin(\zeta-\zeta_i)\right\}, \tag{22}$$

with

$$k_i = k(1\pm\kappa), \quad \omega_i = \omega(1\pm\delta). \tag{23}$$

Note that the boundary conditions (9) and (10) imply

$$L_i = 1 + O(k^2a^2),$$

since to a first approximation the $\tilde{\phi}_i$ must satisfy the (linearized) forms of these conditions for $a \to 0$. The coefficients M_i, N_i, C_i, D_i, and A_i, B_i in (17), are also to be regarded as $O(1)$.

The procedure now is to substitute the perturbed solution (15) into the boundary conditions (9) and (10), linearize in ι, and then reduce all terms to simple-harmonic components. These conditions are supposed to apply over an unbounded range of x, and so they must be satisfied by each set of components at different wavenumber. Accordingly, the separation of components leads to independent equations for the coefficients of $\tilde{\eta}_i$ and $\tilde{\phi}_i$. The work is lengthy but straightforward, and the details can well be passed over here.

The first step is to find the coefficients proportional to $a\epsilon_i$ in (17) and (22). Separating from the boundary conditions all components with arguments $\zeta+\zeta_i$ (i.e. at wavenumbers $k\pm k_i$), we obtain pairs of simultaneous equations for A_i, C_i and B_i, D_i, the approximate solution of which is

$$\left.\begin{aligned} &A_{1,2} = 2P, \quad C_{1,2} = 2Q, \\ &B_{1,2} = -\frac{\frac{1}{2}K\,\mathrm{cosech}^2 K + \lambda\coth K}{K\coth K - \lambda^2}, \\ &D_{1,2} = \pm\frac{\frac{1}{2}\lambda K\,\mathrm{cosech}^2 K + K\coth^2 K}{K\coth K - \lambda^2}, \end{aligned}\right\} \tag{24}$$

if $O(\kappa, \delta)$ is neglected. This approximation is justified in accord with (19), since these coefficients only appear multiplied by k^2a^2 in our final equations. Note that here P and Q are given by (13), and λ by (18).

Next, components at the wavenumbers k_i are separated from the boundary conditions (9) and (10), and the approximation is taken to $O(\omega k^2 a^2 \epsilon_i)$ and $O(\omega^2 k a^2 \epsilon_i)$ respectively. The terms evaluated in (24) contribute to the results, since, for instance, the product $\sin\zeta\cos(\zeta+\zeta_i)$ yields the component $-\frac{1}{2}\sin\zeta_i$. And a point of great importance at this stage is the one demonstrated by the diagram in §2, namely that

$$2\zeta-\zeta_{1,2}=\zeta_{2,1}+\theta,$$

where $\theta=\gamma_1+\gamma_2$. Thus components with the arguments $2\zeta-\zeta_i$ contribute to the final equations. The result obtained from (9) after much reduction is the pair of equations

$$\begin{aligned}\epsilon_{1,2}\{\omega_{1,2}(1-L_{1,2})+\dot\gamma_{1,2}(1-M_{1,2})\}\sin\zeta_{1,2}+\dot\epsilon_{1,2}(1-N_{1,2})\cos\zeta_{1,2}&\\ =\tfrac{1}{2}\omega k^2a^2\{\epsilon_{1,2}R\sin\zeta_{1,2}+\epsilon_{2,1}S\sin(\zeta_{1,2}+\theta)\},&\end{aligned}\tag{25}$$

where

$$\left.\begin{aligned}R&=\tfrac{3}{2}+(2\Delta+A+B)\coth K+2C\coth 2K+|D|\,K^{-1},\\ S&=\tfrac{3}{4}+(P+B)\coth K+2Q\coth 2K+|D|K^{-1}\end{aligned}\right\}\tag{26}$$

are functions of K only, in which A, B, C and $|D|$ are given by (24). And the result from (10) is

$$\begin{aligned}\epsilon_{1,2}[\omega_{1,2}^{-1}\{gk_{1,2}\tanh(k_{1,2}h)-\omega_{1,2}^2\}-\dot\gamma_{1,2}(1+N_{1,2})]\cos\zeta_{1,2}+\dot\epsilon_{1,2}(1+N_{1,2})\sin\zeta_{1,2}&\\ =\tfrac{1}{2}\omega k^2a^2\{\epsilon_{1,2}U\cos\zeta_{1,2}+\epsilon_{2,1}V\cos(\zeta_{1,2}+\theta)\},&\end{aligned}\tag{27}$$

where

$$\left.\begin{aligned}U&=-\tfrac{5}{2}+(2\Delta+A+B)\tanh K-2C\operatorname{cosech}2K-|D|K^{-1},\\ V&=-\tfrac{5}{4}+(P+B)\tanh K-2Q\operatorname{cosech}2K-|D|K^{-1}.\end{aligned}\right\}\tag{28}$$

The boundary conditions must be satisfied independently by simple-harmonic components in quadrature. Hence from each of (25) and (27) two equations are obtained, making eight in all, by separation of terms in $\sin\zeta_i$ and $\cos\zeta_i$. The constants L_i, M_i and N_i can then be eliminated simply by additions and subtractions, and the outcome is four equations with known parameters for the functions $\epsilon_i(t)$ and $\gamma_i(t)$.

Adding the coefficients of $\cos\zeta_i$ in (25) to those of $\sin\zeta_i$ in (27), and finally reducing the quantity $S-V$ after substituting for P, Q, Δ, A, B, C and $|D|$, we obtain

$$\frac{d\epsilon_{1,2}}{dt}=\{\tfrac{1}{2}\omega k^2a^2X(K)\sin\theta\}\,\epsilon_{2,1},\tag{29}$$

where

$$X(K)=\tfrac{1}{2}(S-V)=\frac{9-10\tanh^2K+9\tanh^4K}{8\tanh^4K}+\frac{4+2\operatorname{sech}^2K+3K\coth K\operatorname{sech}^4K}{1-2K\tanh K\operatorname{sech}^2K\cosh 2K+K^2\operatorname{sech}^4K}.\tag{30}$$

An integral of the pair of simultaneous equations (29) is

$$\epsilon_{1,2}(t) = \epsilon_{1,2}(0)\cosh\left\{\tfrac{1}{2}\omega k^2a^2 X\int_0^t \sin\theta\,\mathrm{d}t\right\} + \epsilon_{2,1}(0)\sinh\left\{\tfrac{1}{2}\omega k^2a^2X\int_0^t \sin\theta\,\mathrm{d}t\right\}, \quad (31)$$

which precisely confirms the behaviour suggested in §2, that the side-band amplitudes undergo unbounded amplification if $\theta \to$ const. $(\neq 0, \pi)$.

From the remaining components of (25) and (27), we obtain

$$\frac{\mathrm{d}\gamma_{1,2}}{\mathrm{d}t} = \tfrac{1}{2}\,\omega_{1,2}^{-1}\{gk_{1,2}\tanh(k_{1,2}h) - \omega_{1,2}^2\} + \tfrac{1}{2}\omega k^2a^2\left\{\tfrac{1}{2}(R-U) + \left(\frac{\epsilon_{1,2}}{\epsilon_{2,1}}\right)X\cos\theta\right\}, \quad (32)$$

and these two equations can be added to give one for θ. The sum involves two factors of the form

$$F = f^2(k_i) - \omega_i^2 = f^2(k_i) - \omega^2(1\pm\delta), \quad (33)$$

where

$$f(k_i) = \{gk_i\tanh(k_ih)\}^{\frac{1}{2}} \quad (34)$$

is the expression given by linearized theory for frequency as a function of wavenumber (cf. equation (2) in §2). Recalling that $\delta\omega = \kappa k f'(k)$, we obtain from (33) to a second approximation

$$\begin{aligned} F &= f^2(k) \pm \kappa k f^{2\prime}(k) + \tfrac{1}{2}(\kappa k)^2 f^{2\prime\prime}(k) - \omega^2(1\pm\delta) \\ &= f^2(k) - \omega^2 - \omega^2\delta^2 Y(K), \end{aligned} \quad (35)$$

where

$$\begin{aligned} Y(K) &= 1 - \frac{1}{2}\frac{f^{2\prime\prime}(k)}{[f'(k)]^2} = -\frac{k^2f''(k)}{\lambda^2 f(k)} \\ &= 1 - \frac{8K(1-K\tanh K)\operatorname{cosech}2K}{(1+2K\operatorname{cosech}2K)}, \end{aligned} \quad (36)$$

and the difference $f^2(k)-\omega^2$ can be found immediately to $O(\omega^2k^2a^2)$ from equation (14). Hence, after reduction of the coefficient $R-U$, the two equations (32) lead to

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \omega k^2a^2X(K)\left\{1 + \frac{\epsilon_1^2+\epsilon_2^2}{2\epsilon_1\epsilon_2}\cos\theta\right\} - \omega\delta^2Y(K). \quad (37)$$

The last term in this equation is equivalent to the right-hand side of (4), and thus we now have the extra terms mentioned in §2, which represent the nonlinear effects possibly opposing the detuning effect of dispersion on the side bands.

Conditions of stability or instability

The system of equations given by (29) and (37) have been investigated fully by Benjamin & Feir (1967*a*) in the case of deep-water waves, $K \to \infty$, and explicit solutions were written down. A recapitulation of the principal conclusions is sufficient here.

We first observe that the function $Y(K)$, which is proportional to the negative curvature of the frequency-wavenumber relation according to linearized theory, is positive for all nonzero values of K. For deep-water waves, we have $Y(\infty) = 1$, and Y diminishes steadily towards zero as K is reduced (i.e. as the ratio of water depth h to wavelength $2\pi/k$ is decreased). Hence the argument explained by Benjamin & Feir (1967a) proves that if

$$2k^2a^2X(K) < \delta^2 Y(K), \tag{38}$$

then the solutions $\epsilon_i(t)$ are periodic and finitely bounded. The fact that unbounded amplification of the ϵ_i is impossible in this case is indeed obvious from (31) and (37): the first equation shows that $\epsilon_1 \to \epsilon_2$ if $\epsilon_2 \to \infty$, but (37) shows that $d\theta/dt \to 0$ is then impossible under the condition (38). This is, therefore, the condition of stability.

On the other hand, if

$$2k^2a^2X(K) > \delta^2 Y(K), \tag{39}$$

then for any choice of initial values, with but one exception, the solutions of (29) and (37) have the asymptotic properties

$$\theta \to \cos^{-1}\{(\delta^2 Y/k^2a^2X) - 1\} \quad \text{in} \quad (0, \pi), \tag{40}$$

$$\epsilon_1 \sim \epsilon_2 \sim \exp\left(\tfrac{1}{2}k^2a^2X \sin\theta\, \omega t\right) = \exp\{\tfrac{1}{2}\delta Y^{\frac{1}{2}}(2k^2a^2X - \delta^2 Y)^{\frac{1}{2}}\, \omega t\} \tag{41}$$

for $t \to \infty$. The exceptional case, which obviously has no physical importance, occurs when the initial values of ϵ_1 and ϵ_2 are made equal and $\theta(0)$ is given the value $\cos^{-1}\{(\delta^2 Y/k^2a^2X) - 1\}$ in $(\pi, 2\pi)$: the disturbance then dies away exponentially.

When (39) is replaced by an equality between the two terms, the ϵ_i are again found to have unbounded asymptotic growth, but this is linear rather than exponential in time. Thus, including this marginal case, the condition of instability is

$$2k^2a^2X(K) \geqslant \delta^2 Y(K). \tag{42}$$

We now see that the stability of the primary wavetrain depends in the last resort on the sign of $X(K)$. If $X(K) < 0$, the stability condition (38) is satisfied for all values of the other parameters; but if $X(K) > 0$ a range of values of δ exists satisfying the condition of instability (42). Evaluating the expression (30) for $X(K)$, one finds that

$$\left.\begin{aligned} X(K) > 0 \quad &\text{for} \quad K > 1{\cdot}363, \\ X(K) < 0 \quad &\text{for} \quad K < 1{\cdot}363. \end{aligned}\right\} \tag{43}$$

Thus the Stokes wavetrains on sufficiently deep water, specifically when wavelength $< (2\pi/1{\cdot}363)\,h = 4{\cdot}61h$, are unstable, whereas those for which wavelength $> 4{\cdot}61h$ are stable. The critical value of K, and its meaning as regards stability, has been deduced in a totally different way by Whitham (1967), and it is pleasant to record this nice agreement between the two theoretical approaches.

Note that when $K > 1{\cdot}363$ there is still a 'cutoff' value of δ, namely

$$\delta_c = ka(2X/Y)^{\frac{1}{2}}, \tag{44}$$

above which unbounded growth of the side-band amplitudes does not occur. In other words, for unstable growth the side-band frequencies need to be sufficiently close to the fundamental frequency, so that the modulation of the primary wavetrain is sufficiently gradual. The unstable range $(0, \delta_c)$ of δ narrows down to vanishing point as K is reduced towards 1·363. Note also that the asymptotic growth rate given by (41) is a maximum for $\delta = \delta_c/\sqrt{2}$. The maximum value of the exponent is

$$\left(\frac{1}{\epsilon}\frac{\mathrm{d}\epsilon}{\mathrm{d}t}\right)_{\text{max.}} = \tfrac{1}{2}k^2a^2\omega X(K), \tag{45}$$

which is largest for waves on very deep water ($X = 1$) and diminishes steadily towards zero as K is reduced towards 1·363.

Finally, we note the generalization to the case where the modulation of the primary wavetrain is not spatially uniform. By a well-known argument it can be concluded, for example, that the asymptotic behaviour described by (41) is equivalent more generally to

$$\frac{\partial\epsilon_i}{\partial t} + c_g\frac{\partial\epsilon_i}{\partial x} = \{\tfrac{1}{2}\delta Y^{\frac{1}{2}}(2k^2a^2X - \delta^2 Y)^{\frac{1}{2}}\,\omega\}\,\epsilon_i. \tag{46}$$

In the experiments by Benjamin & Feir (1967*b*), side-band modes were introduced by modulating the movement of a wavemaker generating the primary wavetrain at one end of a tank; the properties of these modes therefore became approximately steady ($\partial\epsilon_i/\partial t = 0$) and developed with distance along the tank. Accordingly, the above formula was used to estimate $\mathrm{d}\epsilon_i/\mathrm{d}x$ for comparison with the measured spatial growth rates.

4. Discussion

The preceding results are, of course, only first approximations to the properties of unstable disturbances for small ka, and correspondingly for small δ. But they are significant theoretically as the *exact* representations of these properties in the limit $ka \to 0$. For instance, (42) is precisely the instability condition for Stokes waves of limitingly small amplitude on a frictionless liquid, if it is understood that a time much greater than $(\omega k^2a^2)^{-1}$ [or a distance much greater than $c_g/\omega k^2a^2 = \lambda/k^3a^2$] is allowed for instability to develop. In practice, however, instability will be suppressed by the action of viscosity if ka is small enough, because viscous damping rates are approximately independent of amplitude. It can be assumed that the mechanisms of damping and of energy transfer to the side bands are virtually independent if both are weak. Hence the practical condition of instability is likely to be

$$\tfrac{1}{2}k^2a^2\omega X(K) > \beta, \tag{47}$$

where the left-hand side of the inequality is the maximum possible rate of amplification according to the inviscid model, as given by (45), and β is the temporal damping rate (equal to c_g times the spatial damping rate) that is suffered by waves of extremely small amplitude at wavenumber k.

For deep-water waves, Benjamin & Feir's experimental results agree reasonably well with the predictions of the theory, leaving no doubt about its essential correctness in describing the process of instability. As already explained, in most of the experiments discrete side-band modes having a predetermined frequency deviation δ were generated with the primary wavetrain, simply by imposing a slight modulation on the reciprocating motion of a wavemaker. Their initial amplitudes were made very small, though large enough for the disturbance to be clearly distinguishable above the noise level in the system, and were usually made about equal. However, in one experiment only a single mode was present initially: the aim then was to confirm the theoretical prediction that the other mode would be produced by the subsequent nonlinear interaction. In another experiment, the results of which are shown in figure 2, the action of the wavemaker was made as regular as possible, with no modulation imposed, so that instability of the radiated wavetrain was allowed to develop spontaneously from background noise. Capacitance probes signalling the instantaneous height of the water surface were placed at various distances along the tank, and from the various records of height as a function of time the developing properties of the disturbance modes were measured. Exponential growth of amplitude with distance was commonly observed through 30 dB or more. Extensive results were obtained from a small tank at Cambridge, and some others in substantial agreement with the first set were obtained on a much larger scale from No. 3 Tank in the Ship Division, National Physical Laboratory.

Figure 2 shows a pair of records obtained at the N.P.L., in an experiment with fundamental frequency $\omega/2\pi = 0{\cdot}85$ c/s, wavelength $2\pi/k = 7{\cdot}2$ ft. and $ka = 0{\cdot}17$. No initial disturbance was introduced artificially. The record presented uppermost in the figure was taken at 200 ft. from the wavemaker, and the lower one was taken at 400 ft. The remarkable feature of these results is that whereas only very slight irregularities can be traced in the first record, the second shows a pronounced amplitude modulation. By measuring the average period of the modulation, it was estimated that $\delta = 0{\cdot}15$, and this value was confirmed by a spectral analysis of the second record, which gave high peaks at the side-band frequencies 0·72 and 0·98 c/s. Thus it appeared that a pair of side-band modes had emerged distinctly by a process of selective amplification from the background noise. Equation (41) or (46) predicts that for deep-water waves, with $X = Y = 1$, the rate of amplification is largest for $\delta = ka$; and so the experimental finding $\delta = 0{\cdot}15$ compares reasonably well with the basic property $ka = 0{\cdot}17$ of the wavetrain. The agreement seems quite fair on consideration that the present theory attempts only a first approximation and that, as is well known from other examples in fluid mechanics, a process of selective amplification is in practice liable to discriminate the optimum wave mode only roughly.

As a second example, figure 4 shows the results of a series of experiments at Cambridge, in which the spatial growth of the side-band amplitudes was measured as a function of the fundamental amplitude with the fundamental and side-band frequencies fixed, specifically with $\omega/2\pi = 2{\cdot}5$ c/s and $\delta = 0{\cdot}1$. For each experimental point, records of the modulated wavetrain had to be taken at a large number of positions along the tank, and hence the logarithmic growth rate $\mathrm{d}(\ln \epsilon)/\mathrm{d}x$ was estimated. The theory indicates that for deep-water waves no amplification of side-band modes at a given δ will occur until ka is greater than $\delta/\sqrt{2}$, which is 0·71 in the present case. The figure shows this cutoff value to be confirmed quite closely; but for larger values of ka the theory overestimates the observed growth rates by a significant margin. The rate of viscous damping was measured to be about 0·01 in the units of the ordinate in the figure, and allowance for this accounts for about a third of the discrepancy. The remainder of the discrepancy can be attributed mainly to the inaccuracy of the first approximation given by the theory; when a correction $O(\delta)$ is made to the coefficient X in (41) or (46) [the error in Y is only $O(\delta^2k^2a^2)$], the agreement with the experimental results is greatly improved. However, even without such adjustments, the comparison shown in figure 4 seems fair, certainly fair enough to inspire general confidence in the theoretical explanation.

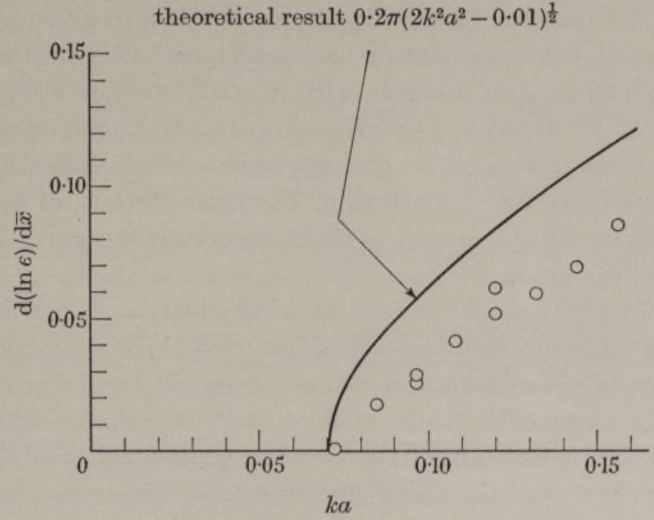


FIGURE 4. Theoretical and experimental values of spatial amplification rate of side-band amplitudes for $\delta = 0{\cdot}1$, expressed as a function of dimensionless fundamental amplitude ka. The ordinate measures $\mathrm{d}(\ln\epsilon)/\mathrm{d}\bar{x}$, where $\bar{x} = (2\pi)^{-1}kx$ is horizontal distance in units of wavelength.

In conclusion, let us leave the problem of straight-crested water waves in uniform channels and review some other applications of present ideas. While avoiding the need to specify the physical system, the general considerations presented in §2 took us quite far towards understanding the envisaged mechanism of instability; and in particular it was shown how the 'dispersive component' of the crucial equation (37) for the phase function θ could be inferred very simply from the relation, written implicitly as $\omega = f(k)$, that holds between frequency and wavenumber in the limit $a \to 0$. Hence an attractive possibility is that a correspondingly simple and general argument might provide the 'nonlinear component' of the θ equation, in which case a universal criterion of instability for systems of the type in question could at once be deduced. I have tried to discover such a criterion by generalizing the essential steps in the analysis for water waves, but nothing simple has come to light. In this connexion it must be noted that Lighthill (1965) has given an elegant and remarkably simple result determining whether, under certain restrictions, very gradual variations in the properties of a wavetrain are governed by elliptic or hyperbolic differential equations; and, of course, in any specific example instability is virtually proved if these equations are shown to be elliptic. But the aforementioned restrictions, which make the dependent variables keep constant mean values, appear to be unnatural in some problems (e.g. waves on water of arbitrary depth, though not in the case of infinite depth), and so Lighthill's result certainly does not provide a general criterion of instability. At this time, therefore, a good deal may still be learnt by examining an increasing variety of physical systems to which these theories can be applied; and it seems particularly worth while to look for fairly simple ones that can also be investigated experimentally.

As a possible case in point, Binnie (1960) has reported some surprising experimental observations on the flow of water along an open channel with corrugated sides. As would be expected, the corrugations were seen to give rise to stationary waves on the steadily flowing stream. But under certain conditions this steady situation appeared to become unstable, with the outcome that a regular succession of travelling waves developed and progressed upstream. Though a definite analysis has not yet been accomplished, I am strongly inclined to believe this is an instance of the type of instability under discussion. The basic state is, of course, akin to a Stokes wavetrain in being a steady periodic structure through which there is a relative motion of the fluid.

To introduce another prospective sphere of application, I mention that I have recently begun experiments on finite extensional waves in long rubber cords. Rubber has the advantage of withstanding enormous strains elastically without significant plastic flow, and consequently its use enables easily repeatable experiments to be made involving waves of large amplitude whose propagation depends in an important way on nonlinear effects and also on effects of frequency dispersion. Now, extensional waves in the first Pochhammer–Chree mode (for which, except at very large wavenumber k, longitudinal displacements are approximately uniform over the cross section) have the property that phase and group velocities are a maximum at zero wavenumber (i.e. $f'(k) \to kf(k), f''(k) < 0$, for $k \to 0$). Also, a very long wave generated by a release of tension (so that the cross section is expanded) has almost all its energy in this first mode, and the forward parts of its waveform suffer a tendency to steepen, due to the effect of amplitude on local propagation speed. Both these properties are analogous to properties of long waves in shallow water; and consistently with this fact it has been observed that the tendency of an extensional wave towards shock formation is resisted by dispersion in the same way as for an open-channel surge, so that a phenomenon precisely analogous to an undular bore arises. Furthermore, the dynamical equations for the elastic system can be transformed to yield the equation of Korteweg & de Vries (see Whitham 1965, §7), which governs the propagation of long water waves. The question of the stability of periodic wavetrains in the new system can be answered immediately, therefore, by reference to the perturbation analysis that Whitham applied to the periodic (i.e. cnoidal wave) solutions of the Korteweg–de Vries equation: his results establish, for both physical systems, that uniform wave trains are stable. For extensional waves at large wavenumber the possibility of instability definitely remains, however, particularly in view of the fact that $f''(k)$ changes sign twice as k is increased from zero. (But for $k \to \infty$ steady wavetrains of finite amplitude become impossible, since $f'(k) \to$ const. and the wave system becomes non-dispersive.) Instabilities of finite flexural and torsional waves in rubber cords are also interesting possibilities which might repay investigation.

I am greatly indebted to Mr J. E. Feir, who performed the experiments on the instability of deep-water waves. I am grateful also to Mr A. Silverleaf, Superintendent of Ship Division of the National Physical Laboratory, for making available the facilities of the towing basin at Feltham where the experimental observations presented in figures 1 and 2 were made.

References (Benjamin)

Benjamin, T. B. & Feir, J. E. 1967*a* The disintegration of wave trains on deep water. Part 1. Theory. *J. Fluid Mech.* **27**, 417.

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

Bowden, K. F. 1948 Some observations of waves and other fluctuations in a tidal current. Appendix: Surface waves of finite amplitude in water of finite depth. *Proc. Roy. Soc.* A **192**, 403.

Binnie, A. M. 1960 Self-induced waves in a conduit with corrugated walls. I. Experiments with water in an open horizontal channel with vertically corrugated sides. *Proc. Roy. Soc.* A **259**, 18.

Korteweg, D. J. & de Vries, G. 1895 On the change of form of long waves advancing in a rectangular channel and a new type of long stationary waves. *Phil. Mag.* (5), **39**, 422.

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. (Dover reprint 1945.)

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

Lighthill, M. J. 1965 Contributions to the theory of waves in non-linear dispersive systems. *J. Inst. Math. Applic.* **1**, 269.

Stoker, J. J. 1957 *Water waves*. New York: Interscience.

Stokes, G. B. 1847 On the theory of oscillatory waves. *Trans. Camb. Phil. Soc.* **8**, 441. *Papers*, **1**, 197.

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

Whitham, G. B. 1965 Non-linear dispersive waves. *Proc. Roy. Soc.* A **283**, 238.

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

Discussion

By K. Hasselmann

The demonstration by Professor Whitham and Dr Brooke Benjamin that a periodic Stokes wave is unstable to small perturbations of frequency and amplitude is extremely interesting, and I wish to congratulate the authors on their fine work. The result appears revolutionary in view of the sustained efforts to prove the mathematical existence of a Stokes wave. But from another aspect the result is perhaps less surprising. In any wave spectrum, the resonant nonlinear interactions have an irreversible tendency to spread the wave energy evenly over all wavenumbers. A narrow spectral peak may therefore be expected to broaden. That this occurs for gravity waves in deep water has been verified by computations (Hasselmann 1963). This is equivalent, however, to saying that an almost periodic Stokes wave is unstable to its side-band energies.

Reference (Hasselmann)

Hasselmann, K. 1963 *J. Fluid Mech.* **15**, 273 (see footnote on p. 281).