PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Nonlinear Wave Packets in the Kelvin-Helmholtz Instability

M. A. Weissman

Phil. Trans. R. Soc. Lond. A 1979 **290**, 639-681
doi: 10.1098/rsta.1979.0019

NONLINEAR WAVE PACKETS IN THE KELVIN–HELMHOLTZ INSTABILITY

By M. A. WEISSMAN†
The National Maritime Institute, Teddington, Middlesex, U.K.

(*Communicated by J. T. Stuart, F.R.S. – Received* 1 *June* 1977– *Revised* 3 *April* 1978)

CONTENTS

The instability of two layers of immiscible inviscid and incompressible fluids in relative motion is studied with allowance for small, but finite, disturbances and for spatial as well as temporal development. By using the method of multiple scaling, a generalized formulation of the amplitude equation is obtained, applicable to both stable and marginally unstable regions of parameter space. Of principal concern is the neighbourhood of the critical point for instability, where weakly nonlinear solutions can be found for arbitrary initial conditions. Among the analytical results, it is shown that (1) the nonlinear effects can be stabilizing or destabilizing depending on the density ratio, (2) the existence of purely spatial instability depends upon the frame of reference, the density ratio, and whether the nonlinear effects are stabilizing, (3) exact nonlinear solutions of

† Now at Flow Research Company, 21414 68th Ave. 5, Kent, Washington 98031, U.S.A.

 [Published 21 February 1979

the amplitude equation exist representing modulations of permanent form travelling faster than the signal velocity of the linear equation (in particular, a solution is found that represents a solitary wave packet), and (4) the linear solution to the impulsive initial value problem has 'fronts' which travel with the two (multiple) values of the group velocity (the packet as a whole moves with the mean of the two values).

Numerical solutions of the amplitude equation (a nonlinear, unstable Klein–Gordon equation) are also presented for the case of nonlinear stabilization. These show that the development of a localized disturbance, in one or two dimensions, is highly dependent on the precise form of the initial conditions, even when the initial amplitude is very small. The exact solutions mentioned above play an important rôle in this development. The numerical experiments also show that the familiar uniform solution, an oscillatory function of time only, is unstable to spatial modulation if the amplitude of oscillation is large enough.

1. Introduction

The nonlinear development of the Kelvin–Helmholtz instability has been studied previously by Drazin (1970) and Nayfeh & Saric (1971, 1972) for the case where the amplitude of an unstable wave is uniform in space and growing only in time. Here, following the lead of Newell & Whitehead (1969) (for thermal convection), Stewartson & Stuart (1971) (for plane Poiseuille flow), Lange & Newell (1971) (for the buckling problem), and Pedlosky (1972) (for baroclinic instability), we study the development and propagation of *packets* of waves, wave trains in which the amplitude is a function of space as well as time. (Many of the results presented here have been summarized in Weissman (1972). See also Weissman (1973).)

The Kelvin–Helmholtz model considered here is the classical one of Lord Kelvin (1871): two layers of immiscible, incompressible, inviscid fluid in relative, irrotational motion. The basic velocity and density profiles are uniform in each layer but discontinuous at the interface, where surface tension exists. Of major concern is the nonlinear development of arbitrary disturbances when this flow is 'slightly' unstable, that is, when the velocity difference exceeds the critical value by a small amount. However, the derivation of the governing equation is general enough to include stable and marginally unstable waves at other points in parameter space.

This highly simplified model is well known as an instructive example in the theory of shear flow instability, where perhaps its greatest value still lies. (The results presented here are believed to be typical of flow instabilities of 'inviscid' character.) However, Thorpe (1969) has demonstrated that the model is directly applicable to (at least) some real flows. In his experiment, the onset of instability was predicted well by the inviscid discontinuous model, even though viscous boundary layers were, of course, present at the interface.

Thorpe used immiscible fluids of similar density; for fluids of large density difference, such as air over water, the situation is not so clear. Kelvin himself recognized that the model does not predict the onset of waves when the wind flows over a body of water. This is due to the presence of other mechanisms for wave generation, such as that in the shear flow model of Miles (1957). However, the Kelvin–Helmholtz *mechanism*† can still operate in more complicated flows, and in some circumstances it could be dominant. For example, Miles (1959) has shown that when the shear flow mechanism is stabilized by a large amount of dissipation in the heavier fluid, Kelvin–Helmholtz instability can still occur. We shall return to the important question of the relevance of the Kelvin–Helmholtz model to the flow of air over water in the summary. In any case, in the bulk of the following analysis, the density ratio of the two layers is allowed to be arbitrary.

† That is, the occurrence of a pressure component in phase with the wave crests. Kelvin–Helmholtz instability results when this pressure becomes so large that it overwhelms the restoring forces of gravity and surface tension.

Kelvin (1871) determined the linear stability characteristics of his model by using what is now known as the method of normal modes (in fact, this appears to have been the first time the method was used to study hydrodynamic stability (Rayleigh 1880)). By linearizing the equations and assuming a sinusoidal disturbance, one obtains a stability diagram as given in figure 1, where U is the velocity difference, or shear, between the two layers (U, suitably non-dimensionalized, is the 'stability parameter' for this problem) and k is the wavenumber in the U-direction of the disturbance. The so-called 'neutral curve' (or 'neutral surface' if oblique waves are allowed) separates the stable and unstable regions. In the linear problem, the solution for an arbitrary disturbance can be found by taking an appropriate sum of these normal modes.

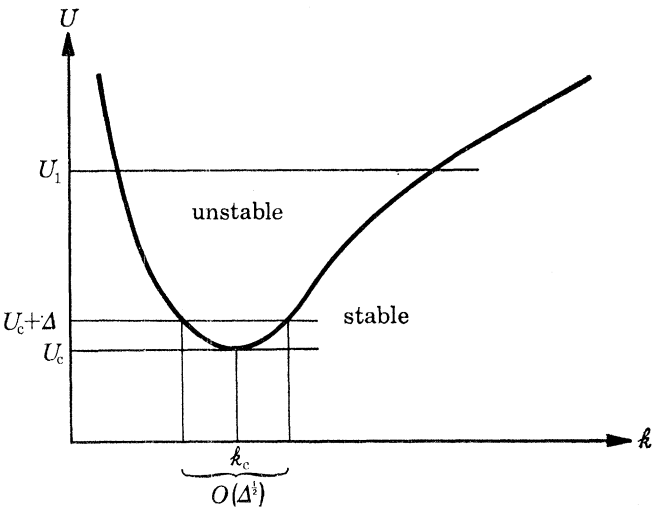

FIGURE 1. The neutral curve.

Whenever the shear is above the critical level, U_c in figure 1, there is a band of waves that can grow exponentially, and the linear solution is valid only for a short period of time. When the disturbance becomes large (in a sense that can be well defined, see § 3), the linear approximation is no longer valid (i.e. the solution has become 'finite amplitude' as opposed to 'infinitesimal'). Thus it is especially important to consider nonlinear effects for the unstable modes. This has not yet been accomplished in general; however, in certain regions of parameter space, nonlinear solutions can be found by use of the methods of asymptotic expansions and multiple scaling (see, for example, Cole 1968).

Multiple scaling relies on the wave amplitude being slowly modulated in time or space. Solutions may be found in the neighbourhood of the neutral curve, where the (linear) growth rate is small. If Δ is the perturbation in U above the neutral curve, the growth rate is proportional to $\Delta^{\frac{1}{2}}$, and so the variation of the amplitude can be expected to be on an $O(\Delta^{-\frac{1}{2}})$ time scale. (This is typical of 'inviscid' instabilities. For instabilities of 'viscous' character the growth rate is $O(\Delta)$. We will compare these two types of flow instabilities as we proceed.)

Formally, solutions may be found anywhere along the neutral curve, but, unless the wavenumber spectrum is quantized in some way, they are useful only near the minimum of the curve, where $U = U_c$. For values of $U \gg U_c$ (e.g. $U = U_1$ in figure 1) there is a wide band of unstable wavenumbers. For an arbitrary initial disturbance, waves in the interior of the unstable region, with large growth rates, would rapidly overwhelm any solutions found for waves near the neutral curve.

Thus it is in the neighbourhood of the critical point where the small-amplitude slowly-varying nonlinear theory is most relevant. In addition, in this region one may attempt a nonlinear solution for arbitrary initial conditions. Consider the velocity difference to be a small amount above the critical value, say $U = U_c + \Delta$, $0 < \Delta \ll 1$, and suppose that the entire band of unstable waves are present in the initial conditions. Then, since the width of the band in Fourier space is $O(\Delta^{\frac{1}{2}})$ (see figure 1; the shape of the curve is parabolic), that part of the disturbance associated with the instability† will appear in physical space as a wavetrain of wavenumber k_c, modulated in amplitude (and perhaps phase) on a spatial scale $\Delta^{-\frac{1}{2}}$.‡ Therefore, including the $O(\Delta^{-\frac{1}{2}})$ temporal scale again, the solution for the entire packet of unstable waves near the critical point is expected to take the form of the real part of

$$A(X, T) \exp[\mathrm{i}(k_c x - \sigma_c t)],$$

where $X = \Delta^{\frac{1}{2}}x$ and $T = \Delta^{\frac{1}{2}}t$. $A(X, T)$ is the 'amplitude function' (which may be complex), and σ_c is the (real) frequency corresponding to the critical point $U = U_c$, $k = k_c$.

In the nonlinear analysis there is another small parameter, ϵ, the scale of the amplitude. The relationship between Δ and ϵ is determined by requiring the instability to appear at the same order in the analysis as the nonlinear 'self-interaction'. This yields $\Delta = O(\epsilon^2)$ and thus, in terms of the amplitude, the time and space scales of the wave-packet are $O(\epsilon^{-1})$. We shall see that this scaling leads to a nonlinear partial differential equation for the amplitude function that is of hyperbolic type: second order in time and second order in space (a nonlinear Klein–Gordon equation).

For values of $U < U_c$, all wave solutions are stable, neither growing nor decaying. The model now supports conservative dispersive wave-trains, which have been considered by numerous authors, e.g. Benney & Newell (1967). It has been found that for these waves, the above scaling is not correct, and a different amplitude equation results. The temporal scale changes to $O(\epsilon^{-2})$, but the spatial scale remains $O(\epsilon^{-1})$. This leads to a nonlinear parabolic equation for the amplitude: first order in time and second order in space. One aim of the present work has been to understand why this changeover occurs when going from the unstable to the stable region. To this end, a rather general solution is sought in which amplitude modulation is allowed on both ϵ^{-1} *and* ϵ^{-2} time and space scales. This results in a system of two partial differential equations for the amplitude, equations which take different forms depending upon the point of interest in parameter space. The coefficients of the linear terms in these equations bear a simple relation to the lowest order linear problem (see (2.01), (2.02)).

The derivation of the general amplitude equations, being rather tedious, is relegated to appendix A; in the text we will concern ourselves mainly with the forms the equations take under various conditions and the corresponding solutions for the amplitude function. In §2, we will find that there are three regions of parameter space in which the equations take on different forms: (1) in the stable region, (2) on the neutral surface but away from the minimum and (3) in the neighbourhood of the minimum of the neutral surface, the critical point for the instability. It is the latter region that will be of primary concern for the rest of the paper. We will find that the

† Stable waves may also be present in the initial conditions. Although we will neglect their influence in this study, it is possible that, regardless of their initial amplitude, they can grow because of nonlinear interactions with the unstable waves. We return to this point in §6.

‡ Note that this argument for the *spatial* scale does not depend on the inviscid character of the instability, but only on the shape of the neutral curve at the minimum.

group velocity is multi-valued here and that unstable wave packets propagate with its *average* value.

In §§ 3 and 4, special solutions are considered which depend only on time or only on space. The uniform, time-dependent solutions are rather familiar; they have been found before by Drazin (1970) and Nayfeh & Saric (1971, 1972) for the Kelvin–Helmholtz instability and by Pedlosky (1970) for the baroclinic instability. However, they are briefly reviewed to provide a background for the more general solutions which follow. In § 4, we consider whether or not a 'spatial' instability is possible, that is, whether a steady-state wave (linear or nonlinear) can exist that is growing only in space. We will see that the answer depends upon the density difference, the frame of reference, and whether the nonlinear effects are stabilizing. The points raised here are related to Briggs's (1964) discussion of 'absolute' versus 'convective' instability (see also Bers 1975). Under linear theory, if an instability is 'absolute', spatial instability is not possible. However, when nonlinearity is included and is stabilizing, we shall find steady-state solutions that represent spatial growth.

In the rest of the paper, we study waves that are modulated in both time and space; however, we restrict ourselves to the case of perhaps greatest physical and mathematical interest, the one in which linear effects are destabilizing, and nonlinear effects stabilizing. In § 5, we will find, as has Fleishman (1959) before us, that wave-like modulations, travelling without change of form, are possible solutions to the nonlinear Klein–Gordon equation. These can travel much faster than the signal velocity of the linear equation. In particular, soliton-like solutions appear, modulation envelopes which take the shape of a solitary wave (these were not studied by Fleishman).

These exact solutions will be seen to play an important rôle in the numerical calculations of § 7. However, before a numerical solution can be attempted, appropriate initial conditions must be considered. This is accomplished in § 6, where the linear development of a localized disturbance is determined (by use of the amplitude equation, which holds for linear as well as nonlinear waves). We will find that the wave packet has 'fronts' which move with the multiple values of the group velocity.

Section 7, 'numerical solutions', has three parts. In the first, localized disturbances in one dimension are considered. We will see here that the initial conditions, even when the initial disturbance is extremely small, have a decisive effect on the final solution. Whenever the initial condition 'contains' the hyperbolic secant in some way, the solitary solution (which has that shape) emerges. Next, the stability of the spatially uniform, time-dependent solution is examined. It is found that spatial modulation is a destabilizing influence. And finally two-dimensional localized solutions are calculated with the simplifying assumption of axisymmetry. Again the initial conditions are found to dramatically affect the form of the solution.

I have included in appendix B a brief discussion of sub-harmonic resonance in the Kelvin–Helmholtz instability. As Nayfeh & Saric (1972) point out, this occurs at precisely the same wavenumber as for ordinary gravity-capillary waves without wind (see, for example, McGoldrick 1970). Nayfeh & Saric consider sub-harmonic resonance in the stable region, but their analysis does not hold for marginally unstable waves. The proper balance between Δ and ϵ is found to be $\Delta = O(\epsilon)$, implying that nonlinear effects are larger (for a given Δ) than in the self-interaction case. We see in appendix B that, once the general form of the operators in the amplitude equations is known, it is a relatively simple task to determine the proper form of the equations for this situation.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

2. The general form of the amplitude equations

If $z = \epsilon\zeta(x, y, t)$ is the elevation of the interface [ϵ is a measure of the magnitude of the disturbance ($0 < \epsilon \ll 1$)], the perturbation expansion presented in Appendix A yields a solution for ζ in the form

$$\zeta = \zeta^{(1)} + \epsilon\zeta^{(2)} + \epsilon^2\zeta^{(3)} + \ldots,$$

where $$\zeta^{(1)} = A(X_1, Y_1, T_1, X_2, Y_2, T_2) \exp[\mathrm{i}(kx + ly - \sigma t)] + \text{complex conjugate},$$

and $\zeta^{(2)}$ (a second-harmonic term) is given in (A 19). The expansion is not expected to be valid if there is growth on the $O(1)$ scales, that is, if the imaginary parts of σ, $-k$ or $-l$ take on positive values. The complex amplitude function A is assumed to depend upon the two sets of slow variables

$$(X_1, Y_1, T_1) = \epsilon(x, y, t)$$

and

$$(X_2, Y_2, T_2) = \epsilon^2(x, y, t).$$

As shown in appendix A, the amplitude function must then satisfy the following system of partial differential equations:

$$-F_\sigma \frac{\partial A}{\partial T_1} + F_k \frac{\partial A}{\partial X_1} + F_l \frac{\partial A}{\partial Y_1} = 0, \tag{2.01}$$

$$\mathrm{i}\left(-F_\sigma \frac{\partial A}{\partial T_2} + F_k \frac{\partial A}{\partial X_2} + F_l \frac{\partial A}{\partial Y_2}\right) + \tfrac{1}{2}F_{\sigma\sigma}\frac{\partial^2 A}{\partial T_1^2} - F_{\sigma k}\frac{\partial^2 A}{\partial T_1 \partial X_1} + \tfrac{1}{2}F_{kk}\frac{\partial^2 A}{\partial X_1^2}$$
$$+ F_{kl}\frac{\partial^2 A}{\partial X_1 \partial Y_1} + \tfrac{1}{2}F_{ll}\frac{\partial^2 A}{\partial Y_1^2} - F_{\sigma l}\frac{\partial^2 A}{\partial T_1 \partial Y_1} = (\Delta/\epsilon^2) F_U A + N|A|^2 A. \tag{2.02}$$

The coefficients of the linear terms are simply the derivatives of the characteristic function,

$$F(\sigma, k, l, U) = \rho_2 \sigma^2 + \rho_1(\sigma - Uk)^2 - (\tilde{g}\kappa + \tilde{\tau}\kappa^3),$$

and are given in (A 23).† (The definitions of ρ_1, ρ_2, $\tilde{g}$, and $\tilde{\tau}$ are given in (A 04); U is the magnitude of the (vector) velocity difference, assumed to be in the x-direction; and $\kappa = (k^2 + l^2)^{\frac{1}{2}}$.) The coefficients must be evaluated at the point of interest in parameter space subject to the condition

$$F(\sigma, k, l, U) = 0. \tag{2.03}$$

Solving (2.03) for $\sigma = \sigma(k, l, U)$ yields the dispersion relation, (A 13).

The term $(\Delta/\epsilon^2) F_U A$ in (2.02) arises by replacing U by $U + \Delta$ in the original equations and assuming that $\Delta = O(\epsilon^2)$. In the stable region, this term can be shown to yield only a small shift in frequency or wavenumber and can be taken to be zero without loss of generality. However, for U marginal [that is, $U = U_m + \Delta$, where $U_m(k, l)$ is the neutral surface], this term will give us the linear instability and must be retained. The nonlinear term $N|A|^2A$ arises by self-interaction; the coefficient, which can be positive or negative, is given in (A 24).

Let us now apply (2.01) and (2.02) in the three regions of parameter space mentioned in the introduction. These regions are distinguished by the vanishing or non-vanishing of the first derivatives of F.

2.1. *The stable region*: $F_\sigma, F_k, F_l \neq 0$

In the region of (U, k, l)-space below the neutral surface, wave packets propagate in much the same way as dispersive waves in other conservative systems. Equation (2.01), rewritten as

$$\frac{\partial A}{\partial T_1} + \sigma_k \frac{\partial A}{\partial X_1} + \sigma_l \frac{\partial A}{\partial Y_1} = 0, \tag{2.04a}$$

† This fundamental form for the amplitude equations arises because the method of multiple scaling is basically an expansion of the linear operators in the original system of equations. This is discussed further in the summary.

shows that modulations on the time scale ϵ^{-1} propagate without change of shape with the group velocity,

$$(\sigma_k, \sigma_\ell) = -(F_k/F_\sigma, F_\ell/F_\sigma) \tag{2.04b}$$

However, (2.02) implies that there will be a change on the ϵ^{-2} time scale due to nonlinear self-interaction. By using (2.04), derivatives in either T_1, X_1 or Y_1, can be eliminated from (2.02). Let us write

$$\frac{\partial^2 A}{\partial T_1 \partial X_1} = -\left(\sigma_k \frac{\partial^2 A}{\partial X_1^2} + \sigma_\ell \frac{\partial^2 A}{\partial X_1 \partial Y_1}\right),$$

and

$$\frac{\partial^2 A}{\partial T_1^2} = \sigma_k^2 \frac{\partial^2 A}{\partial X_1^2} + 2\sigma_k \sigma_\ell \frac{\partial^2 A}{\partial X_1 \partial Y_1} + \sigma_\ell^2 \frac{\partial^2 A}{\partial Y_1^2},$$

from (2.04), substitute into (2.02), and divide through by $-\mathrm{i}F_\sigma$. Then the coefficients of the second-derivative terms can be shown to be proportional to the second derivatives of $\sigma(k, \ell)$: σ_{kk}, $\sigma_{k\ell}$, and $\sigma_{\ell\ell}$. That is, (2.02) becomes

$$\frac{\partial A}{\partial T_2} + \sigma_k \frac{\partial A}{\partial X_2} + \sigma_\ell \frac{\partial A}{\partial Y_2} - \frac{\mathrm{i}}{2}\left[\sigma_{kk} \frac{\partial^2 A}{\partial X_1^2} + 2\sigma_{k\ell} \frac{\partial^2 A}{\partial X_1 \partial Y_1} + \sigma_{\ell\ell} \frac{\partial^2 A}{\partial Y_1^2}\right] = \frac{\mathrm{i}N}{F_\sigma} |A|^2 A. \tag{2.05}$$

(Δ has been set to zero.) The terms in X_2 and Y_2 could now be eliminated by transforming to the group velocity frame of reference.

This equation, in this general form, was first derived by Benney & Newell (1967). Since then it has been found to apply to many stable non-dissipative wave systems. In particular, it has been also found (without the Y-terms) by Nayfeh & Saric (1972) for two-dimensional waves in stable Kelvin–Helmholtz flow. It also applies to ordinary deep-water surface waves (cf. Benney & Roskes 1969; Chu & Mei 1970, 1971; Hasimoto & Ono 1972; Davey & Stewartson 1974). With the instability term (i.e. $\mathrm{i}(\Delta/\epsilon^2)(F_U/F_\sigma)A$) included, equations of this form, first order in time and second order in space (but with complex coefficients), govern 'viscous' instabilities (Newell & Whitehead 1969; Stewartson & Stuart 1971). (See Lange & Newell (1974) for a recent study of the properties of (2.05) with the instability term included.)

Note that in the present case of Kelvin–Helmholtz flow, in contrast to ordinary surface waves, the group velocity vector (σ_k, σ_ℓ) is not parallel to the wavenumber vector (k, ℓ). This is not simply due to advection by the mean flow. From (A 13),

$$\sigma_k = \rho_1 U \pm \frac{(\tilde{g} + \tilde{\tau}\kappa^2)\, k/\kappa}{2(\sigma - \rho_1 U k)} \mp \frac{\rho_1 \rho_2 U^2 k^2}{(\sigma - \rho_1 U k)}, \tag{2.06a}$$

and

$$\sigma_\ell = \pm \frac{(\tilde{g} + \tilde{\tau}\kappa^2)\, \ell/\kappa}{2(\sigma - \rho_1 U k)}. \tag{2.06b}$$

The first term in (2.06a) indicates a bodily advection of wave trains, but the third term is a result of the change in effective inertia of the system due to the (perturbation) pressure difference across the interface.

2.2. *The neutral surface away from the minimum*: $F_\sigma = 0$; $F_k, F_\ell \neq 0$

Everywhere on the neutral surface, $F_\sigma = 0$, as expected for any inviscid instability. Without dissipation, there must be both growing and decaying modes in the unstable region, modes with complex conjugate values for σ (assuming k real). On the neutral surface, the complex conjugate pair coalesces; the equation $F(\sigma, k, \ell, U) = 0$ must yield a double root for σ. This can only happen if $F_\sigma = 0$ (which is readily verified in the Kelvin–Helmholtz case from (A 23a), (A 13) and (A 14)).

With $F_\sigma = 0$, the first derivatives in time drop out of (2.01) and (2.02) and the system becomes second order in time:

$$F^{\mathrm{m}}_{k}\frac{\partial A}{\partial X_1}+F^{\mathrm{m}}_{\ell}\frac{\partial A}{\partial Y_1}=0, \tag{2.07}$$

$$\mathrm{i}\left(F^{\mathrm{m}}_{k}\frac{\partial A}{\partial X_2}+F^{\mathrm{m}}_{\ell}\frac{\partial A}{\partial Y_2}\right)+\tfrac{1}{2}F^{\mathrm{m}}_{\sigma\sigma}\frac{\partial^2 A}{\partial T_1^2}-F^{\mathrm{m}}_{\sigma k}\frac{\partial^2 A}{\partial T_1\,\partial X_1}+\tfrac{1}{2}F^{\mathrm{m}}_{kk}\frac{\partial^2 A}{\partial X_1^2}$$
$$+F^{\mathrm{m}}_{k\ell}\frac{\partial^2 A}{\partial X_1\,\partial Y_1}+\tfrac{1}{2}F^{\mathrm{m}}_{\ell\ell}\frac{\partial^2 A}{\partial Y_1^2}-F^{\mathrm{m}}_{\sigma\ell}\frac{\partial^2 A}{\partial T_1\,\partial Y_1}=(\Delta/\epsilon^2)\,F^{\mathrm{m}}_{U}A+N^{\mathrm{m}}|A|^2A, \tag{2.08}$$

where m indicates evaluation on the neutral ('marginal') surface. This may be compared to 'viscous' instabilities, for which coalescence of modes on the neutral curve does not occur†; $F_\sigma \neq 0$, and the system retains the first-order terms. Thus the order of the amplitude equation (after simplification) is not determined by the order of the original equations of motion, but only by the vanishing or non-vanishing of F_σ. (A good example of this is Rayleigh–Bénard convection, a 'viscous' instability. Although the original equations are second order in time, the resulting amplitude equation is first order (Newell & Whitehead 1969).)

Leaving aside the spatial dependence for the moment (i.e. $A = A(T_1)$ only), we find

$$\tfrac{1}{2}F^{\mathrm{m}}_{\sigma\sigma}\frac{\mathrm{d}^2A}{\mathrm{d}T_1^2}=\frac{\Delta}{\epsilon^2}F^{\mathrm{m}}_{U}A+N^{\mathrm{m}}|A|^2A \tag{2.09}$$

from (2.08). This is a familiar equation from other studies of the nonlinear development of inviscid instability (see, for example Drazin 1970; Pedlosky 1970), and it is well known that the exact solution can be found in terms of Jacobian elliptic functions (which we review in §3). Of primary concern is the sign of the nonlinear coefficient, for this indicates whether the nonlinearity will be stabilizing ($N < 0$) or destabilizing ($N > 0$).

For the Kelvin–Helmholtz instability, N can take either sign (Weissman 1972; Nayfeh & Saric 1972). Evaluating (A 24) on the neutral curve, we have

$$N=-\tfrac{1}{2}\kappa^3\left[4\tilde{g}+\tilde{\tau}\kappa^2+4\,\frac{(\rho_2-\rho_1)^2(\tilde{g}+\tilde{\tau}\kappa^2)^2}{\tilde{g}-2\tilde{\tau}\kappa^2}\right]=N^{\mathrm{m}}. \tag{2.10}$$

Thus, as long as

$$\kappa^2 < \tilde{g}/2\tilde{\tau},$$

then $N < 0$ and the nonlinear term is stabilizing. However, if

$$\kappa^2 > \tilde{g}/2\tilde{\tau} \tag{2.11a}$$

and

$$(\rho_2-\rho_1)^2 > \frac{(\tilde{g}+\tilde{\tau}\kappa^2/4)\,(2\tilde{\tau}\kappa^2-\tilde{g})}{(\tilde{g}+\tilde{\tau}\kappa^2)^2}, \tag{2.11b}$$

N is positive and the nonlinear effects are destabilizing.‡ The point $\kappa^2 = \tilde{g}/2\tilde{\tau}$ (and its neighbourhood; see appendix A) is of course excluded from the present theory. The singularity is due to the occurrence of second-harmonic resonance (see appendix B).

The important case at the minimum, $k^2 = \tilde{g}/\tilde{\tau}$, $\ell = 0$, is included in condition (2.11*a*). This is returned to in the next sub-section.

† At least, I know of no case where it does. But, since coalescence *per se* can occur in viscous flows (see, for example, Gallagher & Mercer 1962, 1964) perhaps the possibility should not be ruled out.

‡ In the first study on nonlinear Kelvin–Helmholtz instability, Drazin (1970) found that N is always negative. However, he only considered the two cases $\tilde{\tau} = 0$, ρ_1/ρ_2 arbitrary (i.e. $\kappa^2 \ll g/2\tau$) and $\tilde{\tau}$ arbitrary, $\rho_1/\rho_2 \approx 1$ (i.e. $(\rho_2-\rho_1)^2 \ll 1$).

Returning now to the situation with spatial dependence, (2.07) implies that A cannot vary on an ϵ^{-1} length scale in the direction defined by the vector $(F^{\mathrm{m}}_{k}, F^{\mathrm{m}}_{\ell})$. The variation in that direction must occur on the ϵ^{-2} scale as determined by equation (2.08). Appropriate new variables could be defined, or $\partial/\partial Y_1$ could simply be replaced by $[(F^{\mathrm{m}}_{k}/F^{\mathrm{m}}_{\ell})\,\partial/\partial X_1]$ – or vice versa – in (2.08), but the result of either procedure is not very enlightening. Let us simplify the discussion of this case by assuming strict two-dimensionality; that is, set $\partial/\partial Y_1 = \ell = 0$. The points under consideration now lie on the neutral curve,

$$U_{\mathrm{m}}(k, 0) = [(\tilde{g}/k + \tilde{\tau}k)/\rho_1\rho_2]^{\frac{1}{2}}.$$

Equation (2.07) becomes

$$F^{\mathrm{m}}_{k}\,\partial A/\partial X_1 = 0. \tag{2.12}$$

Therefore, whether A can be dependent on X_1 depends upon the vanishing of F_k. Evaluating F_k on the neutral curve, we find

$$F^{\mathrm{m}}_{k} = \tilde{g} - \tilde{\tau}k^2. \tag{2.13}$$

This vanishes only at the minimum of the neutral curve where $k = (\tilde{g}/\tilde{\tau})^{\frac{1}{2}}$. Hence, only there can A be a function of X_1; elsewhere, (2.12) implies $\partial A/\partial X_1 = 0$. Away from the minimum, then, and on the neutral curve, (2.08) gives A as a function of T_1 and X_2 alone:

$$\mathrm{i}\,F^{\mathrm{m}}_{k}\frac{\partial A}{\partial X_2} + \tfrac{1}{2}F^{\mathrm{m}}_{\sigma\sigma}\frac{\partial^2 A}{\partial T_1^2} = \frac{\Delta}{\epsilon^2}F^{\mathrm{m}}_{U}A + N^{\mathrm{m}}|A|^2 A \tag{2.14}$$

where the coefficients are evaluated with $\ell = 0$. This equation has also been found by Watanabe (1969) and Nayfeh & Saric (1972).

We shall not consider this case further because, as mentioned in the introduction, in this region of parameter space, near the neutral surface but away from the minimum, solutions involving slowly modulated wave trains are not very useful (because of the possibility of rapidly growing waves inside the unstable region).

2.3. *The critical point*: $F_\sigma = F_k = F_\ell = 0$

At the minimum of the neutral surface ($U = U_{\mathrm{c}}$, $k = k_{\mathrm{c}}$, $\ell = 0$), all three first derivatives of F vanish. Once more, this is a general result for inviscid flow. Consider the variation of F with respect to k along the neutral surface, $U_{\mathrm{m}}(k, \ell)$,

$$\frac{\partial}{\partial k}[F(\sigma_{\mathrm{m}}, k, \ell, U_{\mathrm{m}})] = F^{\mathrm{m}}_{\sigma}\frac{\partial \sigma_{\mathrm{m}}}{\partial k} + F^{\mathrm{m}}_{k} + F^{\mathrm{m}}_{U}\frac{\partial U_{\mathrm{m}}}{\partial k} \equiv 0, \tag{2.15}$$

where $\sigma_{\mathrm{m}}(k, \ell) = \sigma(k, \ell, U_{\mathrm{m}}(k, \ell))$. ($\partial\sigma_{\mathrm{m}}/\partial k$ is well behaved but $(\partial\sigma/\partial k)_{\mathrm{m}}$ is not; see below.) Then, since $F^{\mathrm{m}}_{\sigma} \equiv 0$ for all points on the neutral surface and $\partial U_{\mathrm{m}}/\partial k = 0$ at the minimum,

$$F_k = 0 \quad \text{at the critical point.}$$

Similarly, $F_\ell = 0$. Moreover, since F_ℓ vanishes everywhere on the plane $\ell = 0$ because of the symmetry of F, $F_{k\ell} = F_{\sigma\ell} = 0$ also at the critical point.

With F_σ, F_k, and F_ℓ all equal to zero, equation (2.01) disappears, and (2.02), the sole governing equation for the amplitude, becomes

$$\tfrac{1}{2}F_{\sigma\sigma}\frac{\partial^2 A}{\partial T^2} - F_{\sigma k}\frac{\partial^2 A}{\partial T\,\partial X} + \tfrac{1}{2}F_{kk}\frac{\partial^2 A}{\partial X^2} + \tfrac{1}{2}F_{\ell\ell}\frac{\partial^2 A}{\partial Y^2} = (\Delta/\epsilon^2)\,F_U A + N|A|^2 A, \tag{2.16}$$

where the subscript 1 on X, Y, and T has been dropped and the coefficients are evaluated at the critical point.

On the marginal surface away from the minimum, the components of group velocity,

$$\sigma_k = -F_k/F_\sigma, \quad \sigma_\ell = -F_\ell/F_\sigma, \tag{2.17}$$

are infinite and consequently do not have any physical meaning. However, at the minimum they again take on finite values. It can be shown that

$$\sigma_k = [-F_{\sigma k} \pm (F_{\sigma k}^2 - F_{\sigma\sigma}F_{kk})^{\frac{1}{2}}]/F_{\sigma\sigma} \tag{2.18a}$$

and

$$\sigma_\ell = \pm(-F_{\ell\ell}/F_{\sigma\sigma})^{\frac{1}{2}}. \tag{2.18b}$$

The physical significance of the group velocity in the present context will be seen below but at this point we note that it is multivalued. This is a consequence of the multiplicity of modes. On the neutral surface, the frequencies of two modes coalesce, but their derivatives remain distinct. It is also notable that, even though the wave is travelling in the x-direction, there is a component of group velocity in the y-direction.

The components of the group velocity can be brought into the amplitude equation. The first three terms of (2.16) are factored,

$$\tfrac{1}{2}F_{\sigma\sigma}\left(\frac{\partial}{\partial T}+c_1\frac{\partial}{\partial X}\right)\left(\frac{\partial}{\partial T}+c_2\frac{\partial}{\partial X}\right)A,$$

provided

$$c_1+c_2 = -2F_{\sigma k}/F_{\sigma\sigma}, \quad c_1 c_2 = F_{kk}/F_{\sigma\sigma}.$$

Comparison with (2.18a) shows that c_1 and c_2 are the two values of the group velocity in the x-direction, σ_k^+ and σ_k^-, say. Thus, dividing through by $\frac{1}{2}F_{\sigma\sigma}$ and using (2.18), we can write (2.16) as

$$\left(\frac{\partial}{\partial T}+\sigma_k^+\frac{\partial}{\partial X}\right)\left(\frac{\partial}{\partial T}+\sigma_k^-\frac{\partial}{\partial X}\right)A+\sigma_\ell^+\sigma_\ell^-\frac{\partial^2 A}{\partial Y^2} = GA+\bar{N}|A|^2A, \tag{2.19}$$

where

$$G = 2(\Delta/\epsilon^2)F_U/F_{\sigma\sigma}, \quad \bar{N} = 2N/F_{\sigma\sigma}. \tag{2.20}$$

In one dimension (set $\partial/\partial Y = 0$), this form of amplitude equation has been obtained by Pedlosky (1972) for baroclinic instability (however, his nonlinear term is different because of coupling with the mean flow). As he shows, for a linear wave (neglect the nonlinear term) that is marginally stable (set $\Delta = 0$), modulations of the amplitude can propagate with one of the two values of the group velocity. Thus, the situation at the critical point merges with that in the stable region, where modulations also propagate with the group velocity (cf. (2.04), the equation for behaviour on the same scales as (2.19)).

Equation (2.19) also shows that linear, marginally-stable packets can propagate in the y-direction with one of the two values of the corresponding components of group velocity. Without the X dependence or the terms on the right hand side, (2.19) becomes

$$(\partial/\partial T+\sigma_\ell^+\partial/\partial Y)(\partial/\partial T+\sigma_\ell^-\partial/\partial Y)A = 0$$

(since $\sigma_\ell^+ = -\sigma_\ell^-$). Thus amplitude modulations can propagate along the crests, for the wave itself is travelling in the x-direction. The situation is now quite different from that in the stable region where, for $\ell = 0$, the component of group velocity in the y-direction vanishes and modulations of stable waves do not propagate along the crests on these scales of motion. (However, modulations on the Y_1 scale do affect the behaviour on the T_2 time scale, as (2.05) shows.)

There is another instructive way of writing the amplitude equation. From (2.18), we can let

$$\sigma_k^\pm = c \pm \omega_k, \quad \sigma_\ell^\pm = \pm\omega_\ell,$$

where
$$c = -F_{\sigma k}/F_{\sigma\sigma}, \tag{2.21a}$$
$$\omega_k = +(F_{\sigma k}^2 - F_{\sigma\sigma}F_{kk})^{\frac{1}{2}}/F_{\sigma\sigma}, \tag{2.21b}$$
$$\omega_\ell = +(-F_{\ell\ell}/F_{\sigma\sigma})^{\frac{1}{2}}, \tag{2.21c}$$
and then (2.19) becomes
$$\left(\frac{\partial}{\partial T}+c\frac{\partial}{\partial X}\right)^2 A - \omega_k^2\frac{\partial^2 A}{\partial X^2} - \omega_\ell^2\frac{\partial^2 A}{\partial Y^2} = GA + \bar{N}|A|^2 A. \tag{2.22}$$

The coefficient c has two interpretations. As defined above, it is a 'mean' group velocity, the average of the multiple values of group velocity. We shall see, in fact, that c does have the role of a group velocity in the sense that it is a convection velocity for unstable wave packets (ω_k and ω_ℓ becoming spreading rates).

There is another general definition for c. Let
$$\sigma_{\mathrm{m}}(k,\ell) \equiv \sigma(k,\ell,U_{\mathrm{m}}(k,\ell))$$
be the frequency as evaluated on the neutral surface. Then, since $F_\sigma(\sigma_{\mathrm{m}}, k, \ell, U_{\mathrm{m}}(k,\ell)) \equiv 0$,
$$F^{\mathrm{m}}_{\sigma\sigma}\frac{\partial\sigma_{\mathrm{m}}}{\partial k} + F^{\mathrm{m}}_{\sigma k} + F^{\mathrm{m}}_{\sigma U}\frac{\partial U_{\mathrm{m}}}{\partial k} = 0,$$
after differentiation with respect to k. At the minimum $\partial U_{\mathrm{m}}/\partial k = 0$; therefore,
$$\partial\sigma_{\mathrm{m}}/\partial k = -F_{\sigma k}/F_{\sigma\sigma} = c$$
from (2.21a). Thus c is equal to a special derivative of the frequency with respect to the wavenumber, the derivative *after* evaluation on the neutral surface.

For Kelvin–Helmholtz flow, c also turns out to be equal to the phase speed. By evaluating $F_{\sigma k}$ and F_{kk} from (A 23),
$$c = \rho_1 U,$$
which is the phase speed of marginal and unstable waves (see (A 13)). However, this need not be the case for other flows, e.g. the baroclinic instability (Pedlosky 1972).

Consequently, in a frame of reference moving with this convection velocity (i.e. let $X' = X - cT$; $\pm\omega_k$ and $\pm\omega_\ell$ become the components of group velocity in this 'preferred' frame), the amplitude equation reduces to its simplest from,
$$\frac{\partial^2 A}{\partial T'^2} - \omega_k^2\frac{\partial^2 A}{\partial X'^2} - \omega_\ell^2\frac{\partial^2 A}{\partial Y^2} = GA + \bar{N}|A|^2 A, \tag{2.23}$$
a nonlinear Klein–Gordon equation. In addition to the (inviscid) baroclinic instability, it has been found to be applicable to the buckling of thin shells (Lange & Newell 1971). If the terms on the right hand side are again set to zero, a localized initial packet at the origin would propagate away in an ellipsoidal pattern with speeds $\pm\omega_k$, $\pm\omega_\ell$ on the axes, as governed by the anisotropic wave operator. Since $\omega_\ell > \omega_k$ (see the next paragraph), the major axis would lie in the y-direction.

These results concerning the convection and group velocities were obtained by simply using the fact that the coefficients of the amplitude equation are the various derivatives of the characteristic function. Thus these results are quite general, for the *form* of equations (2.01) and (2.02) can be shown to apply to many flows. However, returning to the Kelvin–Helmholtz model, we

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

may evaluate the coefficients in (2.23), but first let us choose velocity and length scales for non-dimensionalization given in (A 07). For this choice, $\tilde{g} = \tilde{\tau} = 1$. Then

$$\left.\begin{aligned} \mathscr{k}_c = 1, \quad \mathscr{l}_c = 0, \quad U_c = (2/\rho_1\rho_2)^{\frac{1}{2}}, \\ c = \rho_1 U_c = (2\rho_1/\rho_2)^{\frac{1}{2}}, \quad \omega_{\mathscr{k}} = 1, \quad \omega_{\mathscr{l}} = \sqrt{2}, \\ G = 2\frac{\Delta}{\epsilon^2}\frac{F_U}{F_{\sigma\sigma}} = 2(2\rho_1\rho_2)^{\frac{1}{2}}\frac{\Delta}{\epsilon^2}, \\ \bar{N} = 2N/F_{\sigma\sigma} = \tfrac{1}{2}[16(\rho_2-\rho_1)^2-5], \end{aligned}\right\} \tag{2.24}$$

and, from (A 23) and (2.21),

at the critical point.

Thus, the nonlinear coefficient can be positive or negative. If the density difference is large enough, i.e. if

$$(\rho_2-\rho_1) > (\tfrac{5}{16})^{\frac{1}{2}} = 0.559, \tag{2.25a}$$

the nonlinear term is destabilizing. By using the definitions of ρ_1 and ρ_2, (A 04a), this condition can be written in terms of the density ratio:

$$\rho_1/\rho_2 < 0.283 \quad \text{for nonlinear destabilization.} \tag{2.25b}$$

The change in sign of the nonlinear coefficient is due to the influence of the forced second-harmonic component of the solution. The last term in the coefficient resulted from products involving the second harmonic, whose amplitude is proportional to the density difference (cf. (A 19) with γ evaluated at the critical point).

3. Uniform, time-dependent wave trains

From now on we will study only the situation at the critical point, where (2.16), (2.19), (2.22) or (2.23) govern the amplitude development. When the amplitude is uniform, varying only in time, they reduce to

$$\mathrm{d}^2A/\mathrm{d}T^2 = GA + \bar{N}|A|^2A, \tag{3.01}$$

Although this equation and its solutions are familiar (Pedlosky 1970; Drazin 1970; Nayfeh & Saric 1972). Solutions of (3.01) will be needed later in order to form exact solutions of (2.23). Therefore, let us briefly review the various solutions of (3.01), allowing for arbitrary values of G and $\bar{N}$.

It is important to keep in mind here (and indeed throughout the rest of the paper) that amplitude equations such as (3.01) hold for linear waves as well as nonlinear ones. When the nonlinear term on the right hand side is negligible compared to the linear one, that is, when

$$A^2 \ll A_{\mathrm{eq}}^2 = |G/\bar{N}| \quad \text{or} \quad a^2 \ll a_{\mathrm{eq}}^2 = |\Delta F_U/N|,$$

where

$$a = \epsilon A,$$

we recover the linear problem for small growth rate.

The amplitude function A is complex in general and thus contains phase information as well as giving the actual amplitude of the wave. If $A = Re^{i\theta}$, where R is the modulus and θ the phase (both real), then substitution into (3.01) and separation into real and imaginary parts yield

$$\mathrm{d}\theta/\mathrm{d}T = \lambda/R^2, \tag{3.02}$$

and

$$\mathrm{d}^2R/\mathrm{d}T^2 = \lambda^2/R^3 + GR + \bar{N}R^3, \tag{3.03}$$

where λ is a constant determined by the initial conditions, i.e.

$$\lambda = R(0)^2 \,\mathrm{d}\theta/\mathrm{d}T(0).$$

For the sake of brevity, we will only consider the initial condition

$$\mathrm{d}\theta/\mathrm{d}T = 0 \quad \text{at} \quad T = 0. \tag{3.04}$$

This gives $\lambda = 0$ and, hence, $\theta =$ constant for $R \neq 0$. With $\lambda = 0$, (3.03) allows solutions which change sign (note this is not the case for $\lambda \neq 0$), and we may return to equation (3.01) under the assumption that A is purely real. The opposite case, $\mathrm{d}\theta/\mathrm{d}T(0) \neq 0$, cannot be ruled out as a possible initial condition for a uniform wave train (its physical meaning is that the initial frequency is slightly shifted from that given by the dispersion relation); however, we shall find in § 6 that the assumption that A is real is the correct one for a localized, impulsive disturbance.

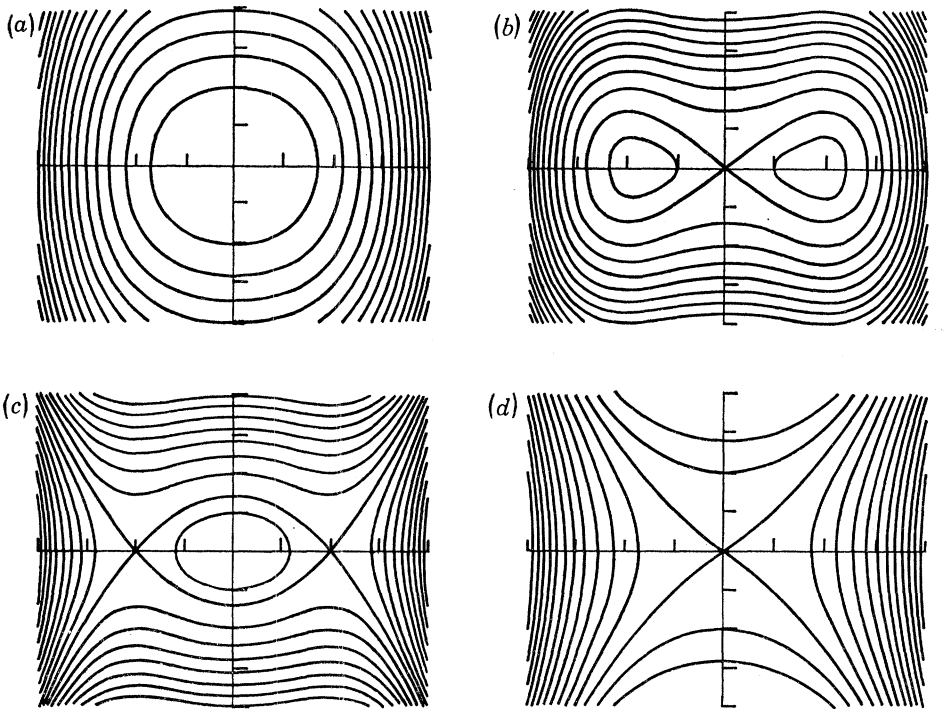

FIGURE 2. Phase plane trajectories for $\mathrm{d}^2A/\mathrm{d}T^2 = GA + \bar{N}A^3$. ($A$ on the horizontal axis, $\mathrm{d}A/\mathrm{d}T$ on the vertical axis.) (*a*) $G < 0$, $\bar{N} < 0$. (*b*) $G > 0$, $\bar{N} < 0$. (*c*) $G < 0$, $\bar{N} > 0$. (*d*) $G > 0$, $\bar{N} > 0$.

For A real, solutions to (3.01) may be found in terms of Jacobian elliptic functions (Milne–Thomson 1950). Bounded or unbounded solutions will result, depending on the signs of G and $\bar{N}$. The situation is summarized in figure 2, where the phase-plane trajectories are drawn for the various sign combinations. These curves were obtained by plotting contours of the first integral of (3.01):

$$E = (\mathrm{d}A/\mathrm{d}T)^2 + \phi(A) = \text{constant}, \tag{3.05}$$

where

$$\phi(A) = -GA^2 - \tfrac{1}{2}\bar{N}A^4. \tag{3.06}$$

The function $\phi(A)$ can be thought of as a 'background potential'; its gradient is a 'restoring force'. Thus, figure 3, in which $\phi(A)$ is sketched, provides a useful visualization of the dynamics, both for uniform wave trains and for the later discussion of spatial variation.

(*a*) $G < 0$, $\bar{N} < 0$. G negative corresponds to Δ negative; that is, the shear difference is slightly below the critical value. Since the nonlinear term is also stabilizing, all solutions are bounded, as the phase plane shows, and they take the form of the cn function.

(*b*) $G > 0$, $\bar{N} < 0$. The amplitude in this case is said to 'equilibrate'. The equation is linearly unstable but the nonlinear term is stabilizing, yielding bounded oscillating solutions of two different types. Inside the separatrix, near the stable equilibrium points ($A = \pm A_{eq}$, $\mathrm{d}A/\mathrm{d}T = 0$) the solutions do not change sign and are given in terms of the dn function. Outside the separatrix, on the trajectories where A does change sign, the solutions have the form of the cn function. On the separatrix, the solution is the hyperbolic secant.

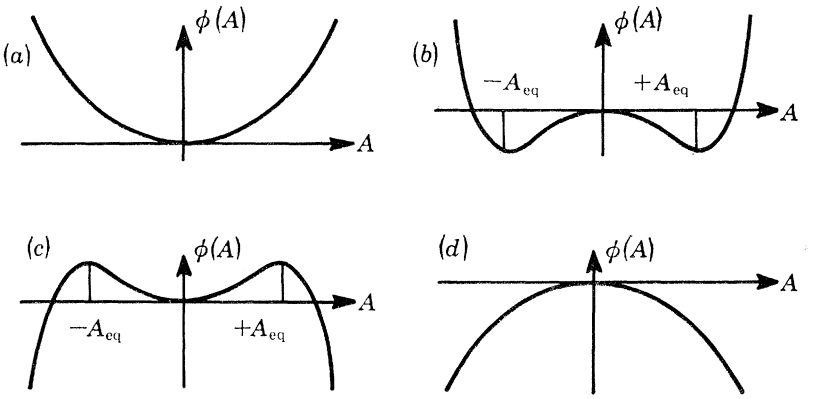

FIGURE 3. The 'background potential', $\phi(A) = -GA^2 - \frac{1}{2}\bar{N}A^4$. (*a*) $G < 0$, $\bar{N} < 0$. (*b*) $G > 0$, $\bar{N} < 0$. (*c*) $G < 0$, $\bar{N} > 0$. (*d*) $G > 0$, $\bar{N} > 0$.

(*c*) $G < 0$, $\bar{N} > 0$. This case yields 'nonlinear' or 'subcritical' instability. If the amplitude is small enough, the linear theory correctly predicts stability, as is shown by the closed trajectories near the origin of the phase plane. The solutions there are given in terms of the sn function (which becomes the hyperbolic tangent on the separatrix). However, if, for example, $\mathrm{d}A/\mathrm{d}T = 0$ and $|A| > A_{eq}$ initially, the flow will be unstable (cf. figure 3*a*). The solutions (e.g. the nc function) are unbounded but remain periodic; they become infinite in a finite amount of time.† (The amount of subcritical destabilization is not particularly large. For a 'typical' wave amplitude of 0.3 and $\bar{N}$ equal to its maximum value, 5.5, the critical velocity difference is reduced by about 10 %.)

(*d*) $G > 0$, $\bar{N} > 0$. There are no bounded solutions in this case. The nonlinear term, because it is destabilizing, cannot equilibrate the linear instability. The growth is 'super-exponential' in the sense that the growth rate is always greater than the linear growth rate. From (3.05), for a trajectory passing near the origin ($E \approx 0$), the growth rate is

$$\left|\frac{1}{A}\frac{\mathrm{d}A}{\mathrm{d}T}\right| \simeq G^{\frac{1}{2}}\left(1+\frac{1}{2}\frac{\bar{N}}{G}A^2\right)^{\frac{1}{2}}.$$

Of course, with unbounded solutions where $\bar{N} > 0$, the method of analysis used here would soon become invalid. When the unscaled amplitude $a = \epsilon A$, becomes $O(1)$, that is, when $A = O(\epsilon^{-1})$, the higher-order terms in the original expansions can no longer be neglected. The calculation could be taken to higher order to determine the motion as a function of slower time scales, but this dependence cannot prevent the solution becoming infinite on the ϵ^{-1} time scale. Only a

† Viscous flows behave similarly, cf. Stuart (1960).

fully nonlinear theory can determine whether the nonlinear instability in cases (*c*) and (*d*) will eventually equilibrate.

When the amplitude *equilibrates* on the ϵ^{-1} time scale, i.e. when the expansion remains valid on this time scale, then it would be useful to calculate the longer scale variations by going to higher order. The oscillations we have found on the ϵ^{-1} scale might be modulating, or even growing, on the ϵ^{-2} scale.

One case where it is actually *necessary* to go to higher order is when $\bar{N}$ becomes small or vanishes, as it can do in the Kelvin–Helmholtz model. If $\bar{N} = O(\epsilon)$, some rescaling is necessary, but an equation of the same form as (3.01) would emerge. However, if $\bar{N} = O(\epsilon^2)$, the cubic term would arise at fifth order, where another self-interaction secularity appears, a term proportional to A^5. The governing amplitude equation would be

$$\mathrm{d}^2A/\mathrm{d}T_2^2 = GA + \bar{N}A^3 + N_5 A^5, \tag{3.07}$$

where $T_2 = \epsilon^2 t$ and Δ is assumed to be $O(\epsilon^4)$. Now the sign of N_5 (assumed $O(1)$) would determine whether unbounded solutions are possible. (Of course this argument could be extended to include spatial dependence.)

4. Steady-state, space-dependent wave trains

With $\partial/\partial T = \partial/\partial Y = 0$, equation (2.22) becomes

$$(c^2 - \omega_k^2)\,\mathrm{d}^2A/\mathrm{d}X^2 = GA + \bar{N}A^3, \tag{4.01}$$

again by taking A to be real. The question of interest here is the occurrence of a steady-state (on the long time) spatially-growing wave driven by a steady source; can there be 'spatial' instability as opposed to the 'temporal' type of the previous section? This question is not easily answered, for a stability analysis is essentially an initial-value problem and might not have a steady-state solution. A solution of (4.01) which yields spatial growth may be unstable itself when time dependence is taken into account. However, (4.01) does tell us which solutions are possible *if* a steady state is achieved.

If we consider first the linear case ($\bar{N} = 0$) with $G > 0$, equation (4.01) shows that exponential growth with X occurs if

$$|c| > |\omega_k|. \tag{4.02}$$

In the stable case ($G < 0$) it would appear that exponential growth could also occur if $|c| < |\omega_k|$. However, by considering the initial value problem, Briggs (1964, ch. 2) has shown that temporal instability must be possible in order to have spatial instability. (The proper solution in this case would be the exponentially *decaying* one.) Thus (4.02) and $G > 0$ are necessary conditions for steady-state *linear* spatial instability. (We shall find that (4.02) is not necessary in the nonlinear case.)

Condition (4.02) can be understood physically by considering a steady-state disturbance (at $X = 0$, say) to be a continuous series of impulses. We will find later (in § 6) that the linear (time-dependent) response to a single impulse disturbance is a modulation envelope that is growing exponentially (approximately) between two 'fronts' which propagate with speeds $c + \omega_k$ and $c - \omega_k$. The packet as a whole is *convected* with speed c while it *spreads* from its centre with speed ω_k. Therefore, to attain a steady state, each packet produced by an impulse must be convected 'downstream' faster than it spreads, or else it will continue to grow near the origin. Both fronts

must propagate in the same direction; hence, $(c+\omega_\ell)\,(c-\omega_\ell) > 0$ or $c^2 > \omega_\ell^2$. (This explanation also shows why G must be positive for spatial instability. For G negative, the response to a single impulse would be a stable dispersing wave train.) The case when a steady-state can be achieved has been termed 'convective' instability; the opposite case, 'absolute' instability (Briggs 1964; Bers 1975).

The value of the 'convective' speed c depends on the frame of reference, but the value of ω_ℓ does not. In the original frame, where the mean velocity of the lower layer is zero, $c = (2\rho_1/\rho_2)^{\frac{1}{2}}$ and $\omega_\ell = 1$ (see (2.24)). Condition (4.02) becomes, in terms of the density ratio,

$$\rho_1/\rho_2 > \tfrac{1}{2}. \tag{4.03}$$

Thus for an air–water system ($\rho_1/\rho_2 \approx 0.0012$) with the water at rest, one would not expect to see steady-state spatially-growing (linear) waves resulting from Kelvin–Helmholtz instability. The same would hold for Thorpe's (1969) experimental set-up. Even though his fluids had $\rho_1/\rho_2 \approx 1$, in his frame of reference $c = 0$, so (4.02) would not be satisfied.

If a steady-state spatial instability is attained, the spatial growth rate, μ_s say, may be compared with the temporal growth rate, $\mu_t = G^{\frac{1}{2}}$. From the linear part of (4.01), $\mu_s^2 = G/(c^2-\omega_\ell^2)$; hence,

$$\mu_t/\mu_s = (c^2-\omega_\ell^2)^{\frac{1}{2}}, \tag{4.04a}$$

or, since $\sigma_\ell^+ = c+\omega_\ell$, $\sigma_\ell^- = c-\omega_\ell$,

$$\mu_t/\mu_s = (\sigma_\ell^+\sigma_\ell^-)^{\frac{1}{2}}. \tag{4.04b}$$

Therefore, as is the case with 'viscous' instabilities, the temporal and spatial growth rates are related by the group velocity, but in a very different way. For 'viscous' instabilities (i.e. when F_σ and F_ℓ are not zero), this ratio is

$$\mu_t/\mu_s = \sigma_\ell,$$

the (single-valued) group velocity itself (Gaster 1962). Equation (4.04b) of course reflects the multi-valued nature of the group velocity, which is a result of the coalescence of modes in the linear problem. To complete the analogy to the 'viscous' case, (4.04b) can also be found by the method used by Gaster, an expansion of the characteristic function for small growth rates (with $F_\sigma = F_\ell = 0$; $F_{\sigma\sigma}$, $F_{\ell\ell} \neq 0$). This yields (4.04b) in the form

$$\mu_t/\mu_s = (F_{\ell\ell}/F_{\sigma\sigma})^{\frac{1}{2}}. \tag{4.04c}$$

Inclusion of the nonlinear term changes the situation somewhat. There are many solutions possible; they fit into the four categories as discussed in §3. When $c^2 > \omega_\ell^2$, the spatial variable is 'time-like' and the solutions are exactly the same as the temporal ones with T replaced by $X/(c^2-\omega_\ell^2)^{\frac{1}{2}}$. For small initial amplitudes (at $X = 0$), the solutions for $G > 0$ will grow exponentially near the origin and then either equilibrate ($\bar{N} < 0$) or continue to grow ($\bar{N} > 0$).

However, when $c^2 < \omega_\ell^2$, it appears that the situations giving stability and instability have been reversed. For example, when $G > 0$ and $\bar{N} > 0$ (complete temporal instability), (4.01) gives a solution that remains bounded (cf. case (i) of the temporal example where $G < 0$ and $\bar{N} < 0$). Although this is a possible steady-state solution, we expect it to be unstable. The 'restoring force' (cf. figure 3d) is always destabilizing; if perturbed by a random disturbance, the solution of the full equation (i.e. with time dependence) must tend to large values.

Thus all the solutions of the nonlinear equation, (4.01), with $c^2 < \omega_\ell^2$, are not expected to be relevant, but there is one case at least when they are. When $G > 0$, $\bar{N} < 0$, solutions of the form

$$A = A_{eq}\tanh(\alpha(|X|-X_0)) \tag{4.05}$$

are possible (see (5.06); the shift in origin is a convenient way of allowing arbitrary amplitude (less than A_{eq}) at $X = 0$). This solution allows a specified amplitude at $X = 0$ (i.e. a fixed source), and, as $X \to \infty$, approaches the equilibrium solution $A_{eq} \equiv |G/\bar{N}|^{\frac{1}{2}}$, which can be shown to be stable (see §7.2). Thus (4.05) is expected to be a stable spatially-growing nonlinear steady-state solution.†

Solutions of the form (4.05) resolve a paradox raised (implicitly) in the earlier part of this section. It was established that if $|c| < |\omega_\ell|$, steady spatially-growing (linear) waves are not possible even if $G > 0$. But what does happen, especially in the case of $G > 0$? The instability must still occur even if A is held constant at $X = 0$. The linear solution (e.g. to an initial value problem) would simply predict that the amplitude would continue to grow everywhere except at $X = 0$.

The same would hold for a nonlinear solution having $\bar{N} > 0$. However, if the nonlinear effects were stabilizing ($\bar{N} < 0$), the solution would be expected to equilibriate while satisfying the boundary condition $A =$ constant at $X = 0$, as in (4.05). Thus there would be steady spatial growth near the origin, where the amplitude could be specified to be very small and where, therefore, the linear equation might be expected to hold: the linear equation which indicates that a steady solution is not possible! The resolution is that the linear equation does not apply at all. Since (4.01) is a second-order equation, both the unknown function and its slope must be small in order to linearize. Although the amplitude could be chosen to be very much less than A_{eq}, its derivative would be $O(A_{eq})$, assuming (4.05) is the steady-state solution. This is too large to linearize but of course it is within our overall scaling scheme. Note that in this essentially *nonlinear* solution, the growth near the origin is *linear* with X.

The same equation (4.01) appears in the next section where nonlinear solutions of permanent form, X and T dependent, are sought. The case $G > 0$, $\bar{N} < 0$ is discussed further there.

5. Nonlinear envelopes of permanent form

Let us consider wave trains with modulations in both time and space (the x-direction only; the two-dimensional case is returned to in §7). Equation (2.23) becomes

$$\frac{\partial^2 A}{\partial T^2} - \omega_\ell^2 \frac{\partial^2 A}{\partial X^2} = GA + \bar{N}A^3, \tag{5.01}$$

where A is again taken to be real and the prime on X has been dropped (X is now in the frame of reference moving with speed c). This equation is very similar to the equation found by Pedlosky (1972) for the baroclinic instability (the difference is in the nonlinear term). Following his lead, we look for travelling solutions of permanent form.‡ Let

$$A = A(X - VT);$$

then (5.01) becomes

$$(V^2 - \omega_\ell^2)\, \mathrm{d}^2A/\mathrm{d}\chi^2 = GA + \bar{N}A^3, \tag{5.02}$$

where $\chi = X - VT$. Note that (4.01), considered in the previous section, is a special case of (5.02): the one where $V = -c$.

† This was confirmed to a certain extent by a numerical experiment which used the method described in §7. At $T = 0$ the amplitude was zero everywhere except at $X = 0$, where it was held constant (and small) as time proceeded. Since there was no damping, a steady state was not achieved; however the solution for long time did appear to be oscillating about a solution of the form (4.05).

‡ Fleishman (1959) has also considered solutions of permanent form of (5.01). However, he did not discuss a solution of interest in the present context: the solitary 'sech' packet.

Many different solutions are possible depending on the signs of $(V^2-\omega_\ell^2)$, G, and N. They fall into four groups as discussed in §3. However, as discussed in §4, not all are expected to be relevant, because they may be unstable. Let us restrict consideration here to the case that is perhaps the most interesting physically and mathematically, the case where the linear instability is equilibrated by nonlinear effects.

With $G>0$, $N<0$, (5.02) is normalized by taking $V'=V/\omega_\ell$, $X'=(G^{\frac{1}{2}}/\omega_\ell)X$, $T'=G^{\frac{1}{2}}T$, $\chi'=(G^{\frac{1}{2}}/\omega_\ell)\chi$, and $A'=(-\bar{N}/G)^{\frac{1}{2}}A$. It becomes

$$(V^2-1)\,\mathrm{d}^2A/\mathrm{d}\chi^2=A-A^3, \tag{5.03}$$

where we drop the primes.

For $V^2<1$ the situation is like that of figure 3*c*, the phase plane for the uniform case where $G<0$, $N>0$. The bounded solutions can be given in terms of the sn and tanh functions. For the sn function (a periodic function),

$$A=A_0\,\mathrm{sn}\{\alpha\chi|m\}=A_0\,\mathrm{sn}\{\alpha(X-VT)|m\}, \tag{5.04}$$

(using the notation of Milne–Thomson 1950; m is the 'parameter') provided

$$V^2=1-(2-A_0^2)/2\alpha^2, \tag{5.05a}$$

$$m=A_0^2/(2-A_0^2), \tag{5.05b}$$

and $A_0^2<1$. Thus the shape (a function of m) depends only on the maximum amplitude A_0; the speed depends upon A_0 and the scale of the oscillation, $1/\alpha$. As the scale decreases, the speed increases but it is always less than 1, the value of the group velocity, ω_ℓ, in the normalized coordinates. As $A_0\to 1$, $m\to 1$, and the solution becomes

$$A=\tanh[\alpha(X-VT)], \tag{5.06}$$

where

$$V^2=1-1/2\alpha^2. \tag{5.07}$$

This solution goes from an equilibrium solution of $A=-1$ to an equilibrium of $+1$, representing a single change of phase of 180° that propagates with the speed V. As the 'sharpness' of the change increases, i.e. $1/\alpha\to 0$, the speed increases to the limit of unity. (This limit cannot actually be reached without violating our scaling assumptions.)

For $V^2>1$, solutions are found in terms of the dn, cn and sech functions; the phase plane is given in figure 2*b*. The dn solution is

$$A=A_0\,\mathrm{dn}[\alpha(X-VT)|m], \tag{5.08}$$

where

$$V^2=1+A_0^2/2\alpha^2, \tag{5.09a}$$

$$m=2(1-1/A_0^2). \tag{5.09b}$$

In this case, the speed is always greater than the group velocity and the larger the scale, $1/\alpha$, the faster the modulation propagates. As $A_0\to\sqrt{2}$, $m\to 1$ and the dn function becomes the hyperbolic secant:

$$A=2^{\frac{1}{2}}\,\mathrm{sech}[\alpha(X-VT)], \tag{5.10}$$

where

$$V^2=1+1/\alpha^2. \tag{5.11}$$

Pedlosky found a similar 'solitary packet' but the speed depended upon the amplitude. Here, the amplitude must be $\sqrt{2}$ and the speed depends only on $1/\alpha$, the width of the packet. The wider the

packet, the faster it goes; as the packet becomes narrower, the speed approaches the lower limit of 1. For $A_0 > \sqrt{2}$, the third type of solution is

$$A = A_0 \operatorname{cn}[\alpha(X - VT)|m], \tag{5.12}$$

where

$$V^2 = 1 + (A_0^2 - 1)/\alpha^2, \tag{5.13a}$$

$$m = A_0^2/[2(A_0^2 - 1)]. \tag{5.13b}$$

We will see in §7 that these solutions that propagate faster than the group velocity play an important part in the development of localized disturbances, particularly when the initial condition is a hyperbolic secant. We will also find that the solitary packet is unstable because any small disturbance in its 'tails' can grow.

6. The linear development of an initial impulse

Again in this section, we consider one-dimensional modulation, i.e. equation (5.01). In order to solve this equation, appropriate initial conditions must be specified. The case of a small, localized disturbance is of special interest.

There are two approaches to the question of initial conditions for A. In one, the disturbance is allowed to develop under linear theory for a long period if time; the asymptotic long-time solution is then taken as the starting point for the nonlinear solution. For example, for an 'infinitesimal' localized disturbance of plane Poiseuille flow, Stewartson & Stuart (1971) find the asymptotic solution to be a Gaussian packet centred on the critical wavenumber and propagating with the group velocity. They find the modulation of the critical mode to be of the form

$$a = \frac{\bar{\epsilon}}{t^{\frac{1}{2}}} \exp\left\{\Delta t - \frac{(x')^2}{t}\right\}, \tag{6.01}$$

to lowest order, for $t \gg 1$ and $(x'/t) \ll 1$, where a is the unscaled amplitude ($= \epsilon A$), $\bar{\epsilon}$ is the characteristic amplitude of the initial disturbance, and x' is the x-coordinate in a frame of reference moving with the group velocity. Δ is the same as we have been using, the perturbation of the stability parameter (Reynolds number in their case) above the critical value. (Note that the growth rate is proportional to Δ, this being a 'viscous' type of instability.) For Δ very small, there is region in time, $t \gg 1$, $\Delta t \ll 1$, in which they 'match' this asymptotic linear solution to the solution of the nonlinear amplitude equation for $\tau \ll 1$, $\tau = \Delta t$ being the long-time variable in their case. This amounts to letting (6.01) – evaluated at some $t = t_0$ – be the initial condition for the amplitude equation.

In using this approach, one must be careful as to what the amplitude actually is at (and before) the time of the matching, for whether the linear solution is still a good representation depends upon its amplitude, not upon how long it has developed. We have seen that there is a parameter available to determine when nonlinear effects can be neglected. This parameter is

$$A_{\mathrm{eq}} = \left|\frac{G}{\bar{N}}\right|^{\frac{1}{2}}, \quad \text{or} \quad a_{\mathrm{eq}} = \left|\frac{F_U \Delta}{N}\right|^{\frac{1}{2}},$$

in terms of the unscaled amplitude. If $a(x, t) \ll a_{\mathrm{eq}}$ at the time of the 'matching' then the linear solution will still be valid.

The other approach is more straight forward. Equation (5.01) describes the evolution of the packet of unstable waves centred on ℓ_c whether the motion is linear or nonlinear. To find $A(X, 0)$,

the amplitude of this packet at $T = t = 0$, we need only to transform to physical space that portion of the initial spectrum corresponding to these waves. Suppose the initial disturbance of the interface is given by

$$\zeta'(x, 0) = \int_0^\infty z(k)\, e^{ikx}\, dk + \text{c.c.},$$

where $\zeta' = \epsilon\zeta$ is the unscaled elevation and c.c. means the complex conjugate of all proceeding terms, and suppose the unstable waves lie in the band $k_1 < k < k_2$. Then the contribution to the initial elevation from the band of unstable waves is

$$\begin{aligned}\zeta'_{\text{un}}(x, 0) &= \int_{k_1}^{k_2} z(k)\, e^{ikx}\, dk + \text{c.c.} \\ &= a(x, 0)\, e^{ik_c x} + \text{c.c.},\end{aligned}$$

where $a(x, 0)$, the initial modulation of the critical mode, is simply

$$a(x, 0) = \int_{k_1 - k_c}^{k_2 - k_c} z(k_c + k')\, e^{ik'x}\, dk', \tag{6.02}$$

letting $k' = k - k_c$.

Near the minimum, the neutral curve is given by

$$U_m(k, 0) = U_c + \tfrac{1}{2}\partial^2 U_m/\partial k^2|_c\, (k - k_c)^2,$$

to lowest order in $(k - k_c)$, or, since it can be shown in general that

$$\left.\frac{\partial^2 U_m}{\partial k^2}\right|_c = \frac{F_{\sigma k}^2 - F_{\sigma\sigma} F_{kk}}{F_{\sigma\sigma} F_U} = \frac{F_{\sigma\sigma}\omega_k^2}{F_U}$$

(by differentiating $F_k^m + F_U^m\, \partial U_m/\partial k \equiv 0$ (cf. (2.15)) with respect to k, evaluating at the critical point, and making use of (2.21 b)),

$$(k - k_c)^2 = \left(2\frac{F_U}{F_{\sigma\sigma}\omega_k^2}\right)(U_m - U_c).$$

Hence, for a given $\Delta = U_m - U_c > 0$, the band of unstable waves lies between

$$k_1 = k_c - \left(2\frac{F_U \Delta}{F_{\sigma\sigma}\omega_k^2}\right)^{\frac{1}{2}} \quad \text{and} \quad k_2 = k_c + \left(2\frac{F_U \Delta}{F_{\sigma\sigma}\omega_k^2}\right)^{\frac{1}{2}}.$$

If we let $k' = k - k_c = \epsilon K$ and $\hat{G} = 2F_U \Delta/F_{\sigma\sigma}\omega_k^2 \epsilon^2$, (6.02) becomes

$$a(X, 0) = \epsilon \int_{-\hat{G}^{\frac{1}{2}}}^{+\hat{G}^{\frac{1}{2}}} z(k_c + \epsilon K)\, e^{iKX}\, dK. \tag{6.03}$$

For example, consider the impulse

$$\zeta'(x, 0) = 2\pi z_c\, \delta(x) = z_c \int_0^\infty e^{ikx}\, dk + \text{c.c.},$$

where $\delta(x)$ is the delta function and z_c is a measure of the strength of the disturbance; (6.03) becomes

$$a(X, 0) = \epsilon z_c \int_{-\hat{G}^{\frac{1}{2}}}^{+\hat{G}^{\frac{1}{2}}} e^{iKX}\, dK = 2\epsilon z_c \frac{\sin(\hat{G}^{\frac{1}{2}}X)}{X}. \tag{6.04}$$

(The same result would follow if z were constant, or nearly constant, only in the neighbourhood of k_c.) We note that $a(X, 0)$ is real, supporting our assumption of real A in (5.01). Furthermore, note that if $z_c = O(1)$, the amplitude is already large enough for nonlinear effects to be important

(i.e. if $z_c = O(1)$, $\epsilon z_c = O(|A|^{\frac{1}{2}}) = O(a_{eq})$). To get linear development for a period of time, one must choose $z_c \ll 1$.

This second approach illustrates the defect in the theory, mentioned in the introduction and worth repeating here: we simply ignore the stable waves that are also present in the initial condition. In a linear problem, they would certainly be negligible compared to the unstable modes, but through nonlinear interactions, these waves can affect the nonlinear development of the unstable packet. This is not a serious handicap in Stewartson & Stuart's viscous example because the stable modes are decaying on the $O(1)$ time scale. Any nonlinear exchange of energy on a longer time scale can not prevent their eventual disappearance. However, in a completely inviscid model such as ours, the stable waves are neutral on the $O(1)$ time scale, neither growing nor decaying. They are always present and, if they satisfy the resonance conditions (see, for example, McGoldrick 1970), are able to partake in an interaction with the unstable modes.

Let us now consider the linear development of the unstable packet. By assuming $A \ll A_{eq}$, (5.01) becomes the linear, but unstable (for $G > 0$), Klein–Gordon equation,

$$\frac{\partial^2 A}{\partial T^2} - \omega_\ell^2 \frac{\partial^2 A}{\partial X^2} = GA,$$

or, if we let $T' = G^{\frac{1}{2}} T$ and $X' = G^{\frac{1}{2}} X/\omega_\ell = \hat{G}^{\frac{1}{2}} X$ and then drop the primes, we have

$$\frac{\partial^2 A}{\partial T^2} - \frac{\partial^2 A}{\partial X^2} = A \tag{6.05}$$

in normalized form. For the delta function (a 'typical' localized disturbance), the initial conditions on A are taken as, cf. (6.04),

$$A(X, 0) = 2z_c' \frac{\sin X}{X}, \quad \frac{\partial A}{\partial T}(X, 0) = 0,$$

when we use the new scaling, where $z_c' = \hat{G}^{\frac{1}{2}} z_c \ll A_{eq}$. For boundary conditions we assume

$$A(X, T) \to 0 \quad \text{as} \quad |X| \to \infty.$$

To solve this it has been found necessary to separate the solution into two parts, $A = A_1 + A_2$, and to write the initial condition as

$$A(X, 0) = z_c' \int_0^1 e^{iKX}\, dK + \text{c.c.} = A_1(X, 0) + A_2(X, 0),$$

where

$$A_1(X, 0) = z_c' \int_0^\infty e^{iKX}\, dK + \text{c.c.} = 2\pi z_c' \delta(X),$$

$$A_2(X, 0) = -z_c' \int_1^\infty e^{iKX}\, dK + \text{c.c.}$$

For the second part of this initial condition, a formal solution can be found by using Fourier transforms:

$$A_2(X, T) = -z_c' \int_1^\infty \cos[\Sigma(K)\, T]\, e^{iKX}\, dK + \text{c.c.},$$

where

$$\Sigma(K) = (K^2 - 1)^{\frac{1}{2}}.$$

Since $K^2 > 1$ in this integral, it forms a stable part of the solution and, after a period of time, will be negligible compared with the other part, which is growing due to the instability. The neglect of this part of the solution is consistent with the neglect of the stable waves that were present in the initial condition on ζ.

A_1 is found by using a Laplace transform; let

$$\alpha(X,p) = \int_0^\infty A_1(X,T)\,\mathrm{e}^{-pT}\,\mathrm{d}T.$$

Then, from (6.05), $\alpha(X,p)$ must satisfy

$$\partial^2\alpha/\partial X^2 - (p^2-1)\,\alpha = -2\pi z_c' p\delta(X),$$

with $\alpha(X,p)\to 0$ as $|X|\to\infty$.

The solution is

$$\alpha = \frac{\pi z_c' p}{(p^2-1)^{\frac{1}{2}}}\exp\left[-|X|(p^2-1)^{\frac{1}{2}}\right] = \pi z_c' p\beta(X,p),$$

if we take $\mathrm{Re}\{p\} > 1$ to ensure that $\mathrm{Re}\{(p^2-1)^{\frac{1}{2}}\} > 0$. From a table of Laplace transforms (see, for example, Abramowitz & Stegun 1964, p. 1027), the inverse of

$$\beta(X,p) = \frac{1}{(p^2-1)^{\frac{1}{2}}}\exp\left[-|X|\,(p^2-1)^{\frac{1}{2}}\right]$$

is

$$B(X,T) = I_0[(T^2-X^2)^{\frac{1}{2}}]\,U(T-|X|),$$

where I_0 is the zero-order modified Bessel function of the first kind and U is the unit step function. Therefore,

$$\begin{aligned} A_1(X,T) &= \pi z_c'\,\partial B(X,T)/\partial T \\ &= \pi z_c' I_0(S)\,\delta(T-|X|) \\ &\quad + \pi z_c'\cdot\frac{T}{(T^2-X^2)^{\frac{1}{2}}} I_1(S)\,U(T-|X|), \end{aligned} \tag{6.06}$$

where

$$S = (T^2-X^2)^{\frac{1}{2}},$$

and I_1 is the first-order modified Bessel function. Thus two 'fronts' propagate away at speeds ± 1, the (normalized) values of the group velocity of the unstable packet. Preceding the fronts, there is no disturbance (except for the stable part that we have neglected). For a point behind the front moving along with some constant velocity, say $X = VT$ where $V < 1$, the growth becomes exponential after a period of time because, for large $S = T(1-V^2)^{\frac{1}{2}}$,

$$I_1(S) = \frac{\mathrm{e}^S}{\sqrt{(2\pi S)}}[1+O(S^{-\frac{3}{2}})]$$

(cf. Abramowitz & Stegun 1964, p. 377). Surrounding the origin, where $X^2 \ll T^2$,

$$S \approx T(1-\tfrac{1}{2}X^2/T^2);$$

therefore,

$$A \approx z_c'\left(\frac{\pi}{2T}\right)^{\frac{1}{2}}\exp\left(T-\frac{X^2}{2T}\right), \tag{6.07}$$

for $T \gg 1$ and $T^2 \gg X^2$.

This long-time approximation is the same form as that obtained by Stewartson & Stuart (1971), cf. (6.01), but it is not as useful. In both cases it is necessary that $x/t \ll 1$, in suitable unscaled variables. In the scaling appropriate to their 'viscous' problem, i.e. $X = \epsilon x$, $T = \epsilon^2 t$, this condition becomes $\epsilon X/T \ll 1$, and thus is satisfied even when X, $T = O(1)$ (as pointed out by Hocking, Stewartson & Stuart 1972). However, in our case, $X = \epsilon x$ and $T = \epsilon t$ and the approximation (6.07) breaks down for X, $T = O(1)$. It is only valid near the origin; nevertheless, this is where the amplitude is largest, and thus (6.07) represents the most important part of the solution, the part that will be first influenced by nonlinearity.

7. Numerical solutions of the amplitude equation

For the most part, linearity unstable, but equilibrating, solutions were studied. That is, G was taken to be positive and $\bar{N}$, negative. The equation used was (5.01); normalized, as in (5.03), it is

$$\frac{\partial^2 A}{\partial T^2} - \frac{\partial^2 A}{\partial X^2} = A - A^3. \tag{7.01}$$

Small localized packets were allowed to develop, at first growing due to the instability and then equilibrating due to nonlinearity. The results are presented in §7.1. Use of (7.01) also enabled an investigation of the stability of uniform solutions to spatial modulation, in the manner shown in §7.2. Two-dimensional, equilibrating packets were studied by using equation (2.23) and assuming axial symmetry; their development is discussed in §7.3.

A few words about the numerical method: It was found necessary to use finite-difference equations with truncation errors $O(\Delta T^4)$ and $O(\Delta X^4)$, where ΔT is the time step and ΔX the grid spacing. That is, the second derivative in X was calculated from

$$\left.\frac{\partial^2 A}{\partial X^2}\right|_j^n = \frac{1}{12(\Delta X)^2}[-A_{j+2}^n + 16A_{j+1}^n - 30A_j^n + 16A_{j-1}^n - A_{j-2}^n], \tag{7.02}$$

where n indicates the time step and j the grid point. By using this representation, (7.01) (or the appropriate equation for the two-dimensional case, see below) gave the value of the second derivative in T,

$$\left.\frac{\partial^2 A}{\partial T^2}\right|_j^n = \left.\frac{\partial^2 A}{\partial X^2}\right|_j^n + A_j^n - [A_j^n]^3. \tag{7.03}$$

The time-stepping was then accomplished by using

$$A_j^{n+1} = 2A_j^n - A_j^{n-1} + \frac{(\Delta T)^2}{12}\left[\left.\frac{\partial^2 A}{\partial T^2}\right|_j^{n+1} + 10\left.\frac{\partial^2 A}{\partial T^2}\right|_j^n + \left.\frac{\partial^2 A}{\partial T^2}\right|_j^{n-1}\right], \tag{7.04}$$

which is also a fourth-order approximation. To handle the implicit nature of this equation, a one-step predictor-corrector method was used. The first approximation to A_j^{n+1} was taken as

$$A_j^{n+1} = 2A_j^n - A_j^{n-1} + (\Delta T)^2 \partial^2 A/\partial T^2|_j^n.$$

This was used in (7.03) (with n replaced by $n+1$) to obtain a first approximation to $\partial^2 A/\partial T^2|_j^{n+1}$, which, in turn, was used in (7.04) to calculate the 'corrected' value of A_j^{n+1}. Finally $\partial^2 A/\partial T^2|_j^{n+1}$ was calculated again to prepare for the next time step. (One cycle was found to be sufficient.) Usually, ΔT and ΔX were taken to be equal and of the value 0.025, 0.05, or 0.1. The resulting solution was graphically displayed on microfilm, making its study much easier.†

7.1. *One-dimensional solutions*

The first initial condition tried was

$$A(X, 0) = A_0 \frac{\sin X}{X}, \quad A_0 = \text{constant}, \tag{7.05}$$

which is the initial amplitude of the unstable packet when the initial spectrum (for ζ) is uniform (or nearly uniform); cf. (6.04). However, this decays rather slowly with X and caused some problems at the boundaries. The alternative approach was then used, taking the distribution obtained

† These data and a 16 mm film of a few of the cases studied are available from the author.

from the long-time asymptotic approximation to the linear solution. This is a Gaussian packet, cf. equation (6.07),

$$A(X, 0) = A_0 \exp(-\alpha^2 X^2), \tag{7.06}$$

where α is a scaling factor. It was found to give similar solutions as $\sin X/X$, without causing disturbances at the far boundaries. The initial condition on the time derivative was taken to be

$$\frac{\partial A}{\partial T}(X, 0) = 0, \tag{7.07a}$$

and symmetry boundary conditions were used, enabling the calculation to be carried out for $X > 0$ only. This is equivalent to specifying

$$\frac{\partial A}{\partial X}(0, T) = \frac{\partial A}{\partial X}(X_L, T) = 0, \tag{7.07b}$$

where X_L is the far boundary.

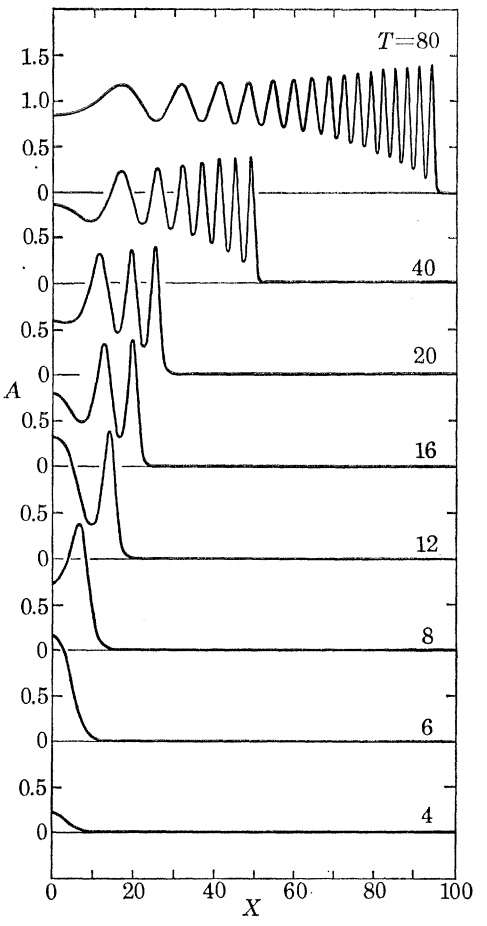

FIGURE 4. Development of a Gaussian packet; $A_0 = 0.01$, $\alpha = 0.25$. (The solution is symmetric about $X = 0$.)

The result of the calculation for a typical case, $A_0 = 0.01$ and $\alpha = 0.25$, is shown in figure 4 for selected values of T. At the origin, the amplitude quickly reaches the equilibrium value, $A = 1$, about which it oscillates. The packet starts spreading with a speed greater than unity, but as the 'front' steepens, it slows down and approaches a speed of 1. Undulations are continually transmitted toward the front, propagating much faster than 1. This causes a concentration of peaks near the front; as the undulations crowd together, they grow in peak-to-trough amplitude,

steepen, and slow down, reflecting the behaviour of the exact nonlinear solutions. In fact, the waveform behind the front resembles a slowly-varying dn function, the solution for $V > 1$. Eventually, they pack so closely together and become so steep that the calculation is no longer accurate.

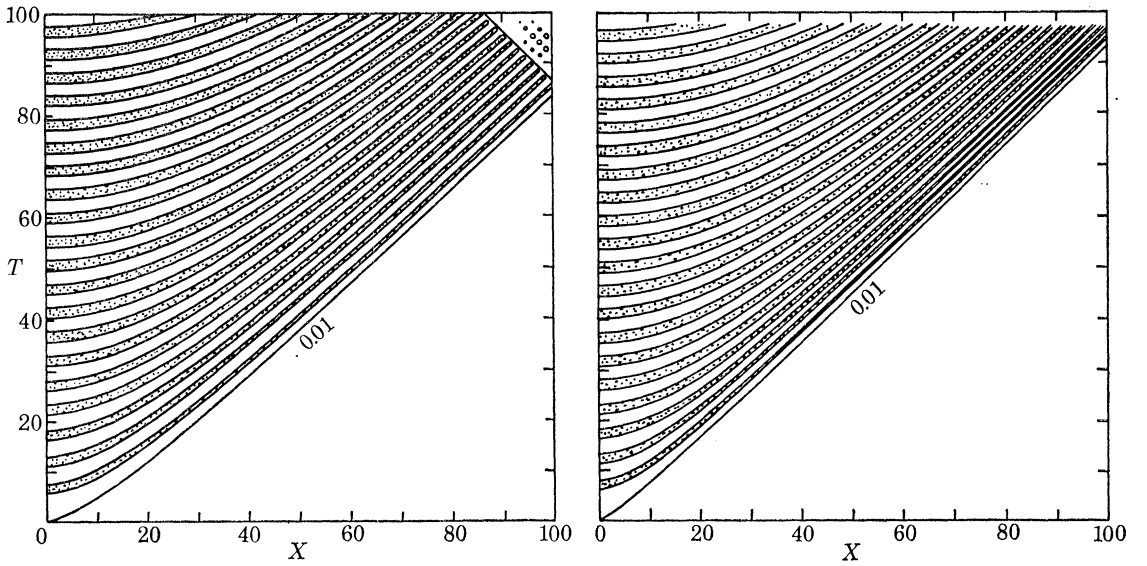

FIGURE 5. X–T trajectories for two Gaussian packets; contours of $A = 0.01$ (marked) and $A = 1.0$. The shaded area indicates the peaks. (There is symmetry about $X = 0$.) (*a*) $A_0 = 0.01, \alpha = 0.25$. (*b*) $A_0 = 0.01, \alpha = 0.5$.

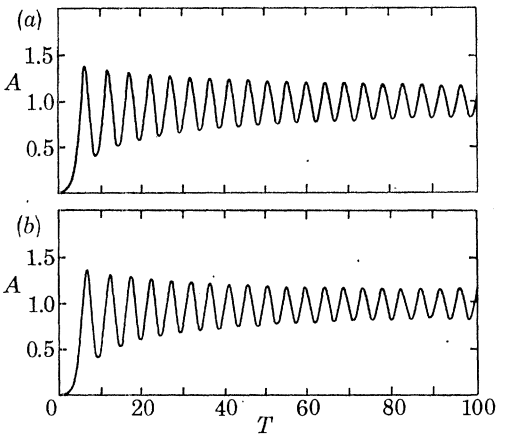

FIGURE 6. The amplitude at the centre of the packet ($X = 0$) against time. (*a*) Gaussian initial packet; $A_0 = 0.01$, $\alpha = 0.25$. (*b*) 'Sech' initial packet; $A_0 = 0.01$, $\alpha = 0.25$.

Other views of the solution are offered in figure 6*a*, where the amplitude at $X = 0$ is plotted against time (the period and amplitude of the oscillation around $A = 1$ slowly decrease), and in figure 5*a*, which is an X–T contour plot of the function $A(X, T)$. The line marked 0.01 is the contour line for $A = 0.01$ and indicates the front of the disturbance. The other lines are all contours for $A = 1.0$; thus the shaded area corresponds to the trajectory of a peak ($A > 1.0$) and the speed can be easily determined as the inverse of the slope of a trajectory. In the upper right-hand corner, the wave train has reflected from the far boundary.

Other values of A_0 (0.1, 1.0, $\sqrt{2}$, 2, $2\sqrt{2}$) and α(0.5, 1.0) were tried and the same type of wave train evolved (unless A_0 was large (above 2), in which case a portion of the solution would be negative). One such case is summarized in figure 5*b*: $A_0 = 0.01, \alpha = 0.5$. Being a narrower packet to begin with, it steepens up faster and consequently propagates more slowly than the $\alpha = 0.25$ example. However, the oscillation at the origin is practically the same, with a period of about 5.

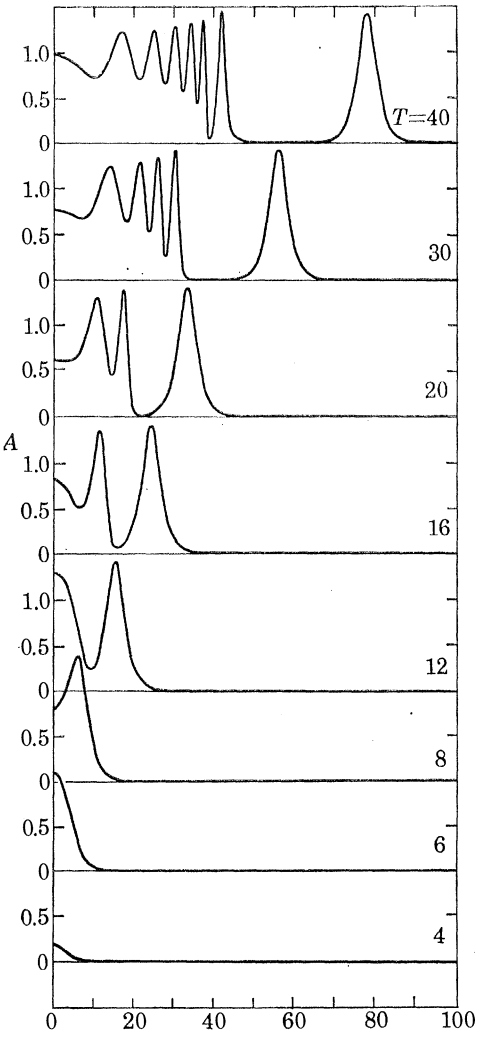

FIGURE 7. Development of a 'sech' packet; $A_0 = 0.01$, $\alpha = 0.5$.

A radically different solution developed when the initial condition for A was

$$A(X, 0) = A_0 \operatorname{sech}(\alpha X), \tag{7.08}$$

the other conditions remaining the same. This distribution was tried to test the relation, if any, to the exact solution

$$A(X, T) = 2^{\frac{1}{2}} \operatorname{sech}[\alpha(X - VT)], \tag{7.09}$$

where $V^2 = 1 + 1/\alpha^2$. It was found that (7.08) does give rise to (7.09) for all the values of A (0.01, 1.0, $\sqrt{2}$, $2\sqrt{2}$) and α (0.25, 0.5, 1.0) that were used. As shown in figure 7 for $A_0 = 0.01$ and $\alpha = 0.5$, the initial development is very much like the Gaussian, but at $T \approx 15$, a solitary packet starts to form. Fully developed by $T = 25$, it moves away from the rest of the wave train, which resembles

the solution for the Gaussian. The amplitude of this solitary packet is always $\sqrt{2}$ and its width and speed always correspond to the α chosen for the initial condition (within the accuracy of measuring these quantities from the graphical display). Figure 6*b* shows that the solution at $X = 0$ is very similar to that for the Gaussian.

The separation of the solitary packet can also be seen in the X–T trajectories of figure 8, in which the solutions for the three values of α are compared in parts (*a*), (*b*), and (*c*). The speed of the packet decreases as α increases, in agreement with the exact solution. We also see in figure 8 the development of another solitary packet behind the first one. Unfortunately we can not be confident of this result, for it was found that the solution behind the solitary packet depends upon the accuracy of the numerical calculation. This is demonstrated in figure 8*d*, which is the same case as in figure 8*b* but the time step and grid size have been decreased from $\Delta T = 0.1$ to $\Delta T = 0.025$

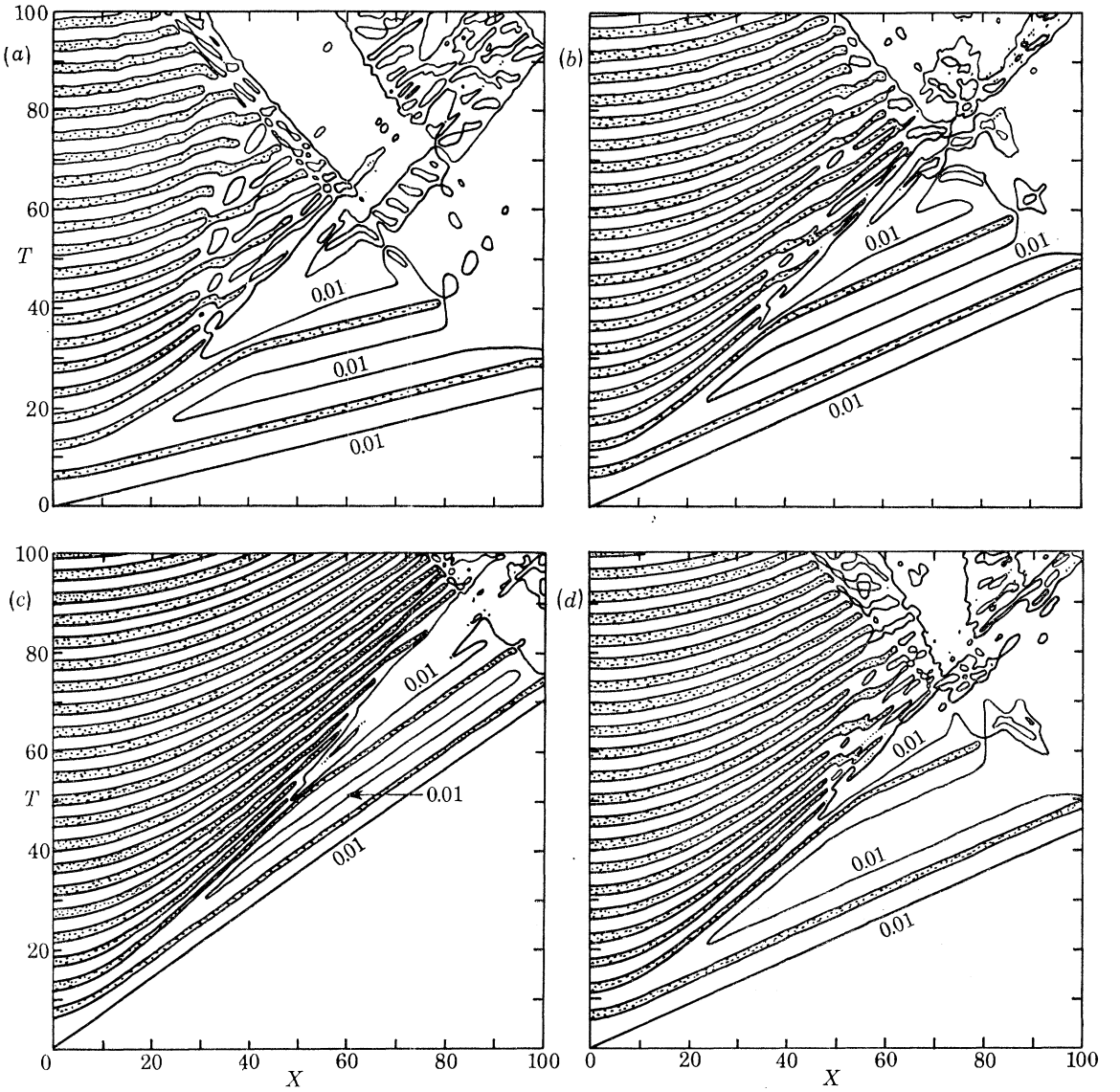

FIGURE 8. X–T trajectories for three 'sech' packets; contours of $A = 0.01$ (marked) and $A = 1.0$. The shaded area indicates the peaks. (*a*) $A_0 = 0.01$, $\alpha = 0.25$. (*b*) $A_0 = 0.01$, $\alpha = 0.5$, $\Delta X = 0.1$, $\Delta T = 0.1$. (*c*) $A_0 = 0.01$, $\alpha = 1.0$. (*d*) $A_0 = 0.01$, $\alpha = 0.5$, $\Delta X = 0.05$, $\Delta T = 0.025$.

and from $\Delta X = 0.1$ to $\Delta X = 0.05$. A second solitary packet still forms, but it follows a further distance behind.

A combination of two factors causes the solution to depend on the accuracy: the hyperbolic secant solution is an exact analytical result, one that starts at $A = 0$ and returns to $A = 0$. The numerical solution, on the other hand, is approximate; it can not reproduce the 'sech' exactly. As the solitary packet propagates forward, it leaves behind a 'residue', a small error remaining from the approximate calculation of the 'sech'. Now the second factor comes into play: $A = 0$ is a position of unstable equilibrium (cf. figure 3*b*); the 'residue' in the tail of the packet goes unstable!

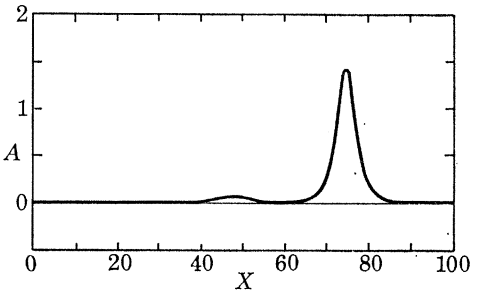

FIGURE 9. A solitary wave packet showing instability in its 'tail'; $\alpha = 0.5$, $T = 11$.

This sequence of events is supported by an experiment which produced the solitary packet by itself. By using initial conditions appropriate for the exact 'sech' solution (7.09), the calculation gave a solitary packet of correct shape and speed for the value of α chosen. However, after a period of time, an instability developed in the tail of the packet, as shown in figure 9. It was found that with an increase in accuracy, the appearance of the instability could be delayed, but not eliminated completely. (In the 'best' example, for $\Delta T = \Delta X = 0.025$, the instability appeared at $T = 15$.)

Figure 8 shows that the increase in accuracy between cases (*b*) and (*d*) delayed the development of the second solitary packet, but it did not affect the first packet or the solution near the origin. Hence, it appears that if we could calculate the exact solution, it would be composed of a single solitary packet propagating away from a trailing wave train similar to that found for the Gaussian. The numerical calculation can not reproduce this because it can not calculate the 'sech' exactly. The small errors in the calculation, which one expects in any numerical experiment, do not remain small.

Numerous types of localized distributions were tried; most of them produced a wave train like that for the Gaussian. For example, the initial development of a 'delta function' (the initial amplitude was zero everywhere except at $X = 0$) was very similar to the Gaussian, with the solution growing to and oscillating about $A = 1$ at the origin, but the speed of its front was very nearly a constant value of unity, and the trailing peaks converged much sooner.

However, whenever the initial condition contained the hyperbolic secant in some manner, the solitary packet would appear. A 'sech' times a Gaussian, a 'sech' plus a Gaussian, a 'sech' squared – all produced the solitary packet separating out and propagating ahead. Even when random noise was 'shaped' by the 'sech' (a random function was multiplied by the hyperbolic secant), the packet appeared. (This last example illustrates the stability of the *major* portion of the 'sech' solution; only in its tails is there danger of instability. This technique was also used to test the stability of the wave train produced by the Gaussian; it was also found to be stable.)

7.2. *Instability of uniform solutions*

As discussed in § 3, exact uniform solutions of (7.01) can be found in terms of the dn function. They are

$$A = \bar{A}(T) = A_0 \,\mathrm{dn}\,(\alpha T|m), \tag{7.10}$$

where $\alpha^2 = \frac{1}{2}A_0^2$, $m = 2(1-1/A_0^2)$, and A_0 is the initial amplitude. The numerical scheme reproduced these solutions very well; the initial conditions were simply taken to be

$$A(X,0) = A_0, \quad \partial A(X,0)/\partial T = 0.$$

The stability of these solutions was tested first by taking as initial conditions

$$A(X,0) = A_0 + 0.01\,R(X), \quad \partial A(X,0)/\partial T = 0,$$

where R is a random function ranging from 0 to 1.0. The result for $A_0 = 0.3$ is shown in figure 10. The period of the basic oscillation is about 6 so the sequence shows the solution after the 2nd, 3rd, 4th, and 5th cycles. The pattern in the last picture continues to grow, the solution becoming quite irregular with positive and negative values.

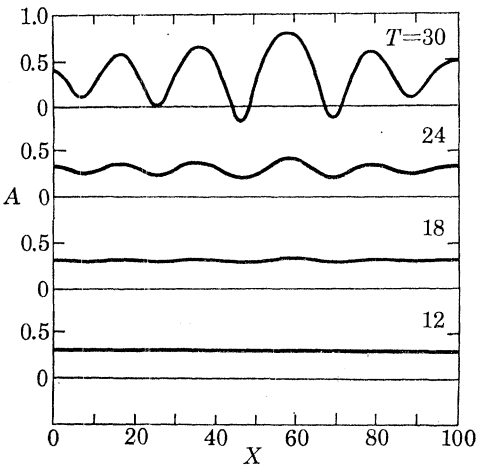

FIGURE 10. Instability of an oscillating uniform solution to a random disturbance; $A_0 = 0.3$, period ≈ 6.

In order to ensure that the discontinuous nature of the initial condition had no influence on the occurrence of this instability, the initial conditions were next taken as

$$A = A_0 + 0.01 \sin(\alpha X), \quad \partial A(X,0)/\partial T = 0,$$

i.e., uniform plus a small sinusoidal disturbance. The boundary conditions were periodic with X_L taken to be 20π so that α could assume the values 0.2, 0.4, ..., 2.0. (For the largest α, there were fifty grid points per wavelength.) A_0 was chosen as 1.1, 1.2, ..., 1.7; for $1.1 \leqslant A \leqslant 1.4$ the basic solution is the dn function, for $A_0 \geqslant 1.5$ it is the cn function. ($A < 1$ need not be considered since that would also give the dn function, cf. figure 2*b*.) The basic uniform solution was first determined for each value of A_0. Then, for a particular value of α, the sinusoidal perturbation was included, yielding the 'total' solution. Subtraction of the two gave the perturbation solution. The stability diagram, figure 11, was then determined on the basis of whether the perturbation grew beyond its initial value over a period of the basic (uniform) oscillation.

Although rather crude, the diagram is sufficient for our purposes; it shows the existence of the instability and of a well-defined stability boundary. We note that the larger the amplitude of the basic oscillation is about $A = 1$, the more unstable it becomes. When $A \equiv 1$, the steady-state uniform solution, it can be shown (by a linear perturbation analysis) that all disturbances are stable (this holds for the two-dimensional case as well).† The limit $\alpha \to 0$ could not, of course, be considered by this method, but along the line $\alpha = 0$, the case of a uniform perturbation, there must also be stability. This can be seen from § 3, where exact (total) solutions are discussed for the uniform case. In particular, figure 3*b* shows that there are closed trajectories in the phase plane (for $G = 1 > 0$ and $\bar{N} = -1 < 0$). A small (uniform) perturbation to one of these solutions merely puts the solution onto a neighbouring trajectory. (Of course, in the neighbourhood of the separatrix, the character of the solution may change substantially, but it would still be 'stable' in the sense that the solution remains periodic and returns to its original value after one cycle.)

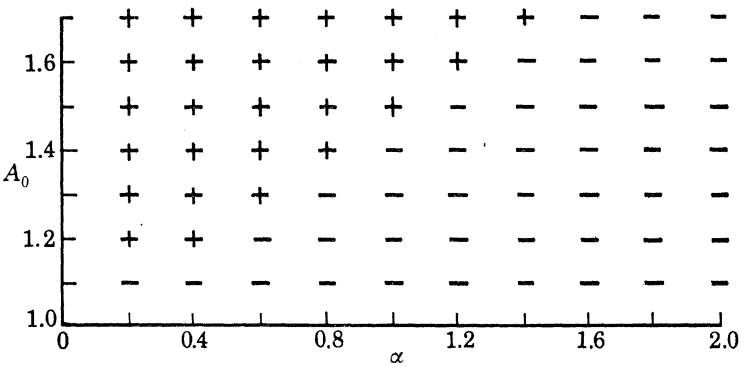

FIGURE 11. Diagram for the stability of uniform solutions to sinusoidal disturbances: + indicates instability; − stability.

7.3. *Two-dimensional solutions*

In two dimensions, we have equation (2.23)

$$\frac{\partial^2 A}{\partial T^2} - \omega_k^2 \frac{\partial^2 A}{\partial X^2} - \omega_\ell^2 \frac{\partial^2 A}{\partial Y^2} = GA + \bar{N}A^3, \tag{7.11}$$

again taking A to be real. This is normalized by letting $T' = G^{\frac{1}{2}}T$, $X' = (G^{\frac{1}{2}}/\omega_k)\,X$, $Y' = (G^{\frac{1}{2}}/\omega_\ell)\,Y$ and $A' = (-\bar{N}/G)^{\frac{1}{2}}A$, with $G > 0$ and $N < 0$. (7.11) becomes

$$\partial^2 A/\partial T'^2 - \nabla^2 A = A - A^3, \tag{7.12}$$

again dropping the primes, where ∇^2 is the Laplacian operator.

Equation (7.12) was studied by assuming axial symmetry, in which case

$$\nabla^2 A = \frac{\partial^2 A}{\partial R^2} + \frac{1}{R}\frac{\partial A}{\partial R},$$

where R is the radial coordinate. Thus, in the numerical scheme, equation (7.03) needed only to be changed by replacing X with R and adding the term $(\partial A/\partial R|_j^n)/R_j$ to the right hand side. Fourth-order finite differencing was also used for this term:

$$\left.\frac{\partial A}{\partial R}\right|_j^n = \frac{1}{12(\Delta R)}\left(-A_{j+2}^n + 8A_{j+1}^n - 8A_{j-1}^n + A_{j-2}^n\right).$$

† Of course, a linear perturbation analysis could also be attempted for the time-dependent solutions considered above. This yields a linear equation for the perturbation, but the coefficients are time-dependent, being proportional to $\mathrm{dn}(T)$ or $\mathrm{cn}(T)$ squared. Thus one may be forced to a numerical solution in any case.

The initial and boundary conditions were the same as the one-dimensional case: Gaussian and hyperbolic secant distributions, as in (7.06) and (7.08) with X replaced by R, and symmetry at $R = 0$ and $R = R_L$. For the Gaussian, the development is similar to that for the one-dimensional case; at the origin the solution oscillates about $A = 1$, sending forward undulations that steepen as they approach the 'front', which is well defined. However, for the 'sech' initial condition, the solitary packet no longer appears. A different type of characteristic solution forms, a wave train that appears to be linearly modulated dn function, as shown in figure 12*b*. (For comparison, figure 12*a* gives the Gaussian type of solution at the same time.) The speed of the train is greater than unity and increases when the width of the initial distribution is increased, i.e. when α decreases. Since this solution does not attempt to return to $A = 0$, it is not affected by the instability found in the one-dimensional case. The series of peaks are still evolving and may eventually separate into individual, solitary 'sech' packets: a possible exact solution when $R \to \infty$.

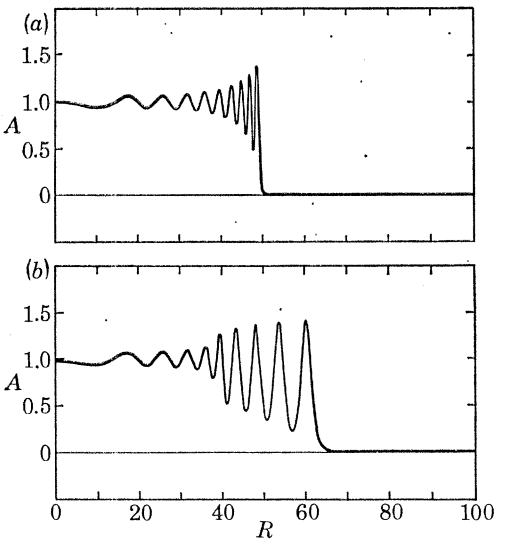

FIGURE 12. Two-dimensional, axisymmetric solutions; $T = 45$. (*a*) Gaussian initial packet; $A_0 = 0.1$, $\alpha = 0.5$. (*b*) 'Sech' initial packet; $A_0 = 0.1$, $\alpha = 1.0$.

8. SUMMARY AND FURTHER DISCUSSION

We have found weakly nonlinear solutions that describe the growth of wave packets and wave trains in the Kelvin–Helmholtz instability. Of course, the theory is limited in various ways:

(1) In parameter space, it does not apply inside the unstable region except near the neutral surface.

(2) In terms of the amplitude, it ceases to be valid if the amplitude becomes $O(1)$, that is, if $A = O(\epsilon^{-1})$. (Physically this means the wave height is the order of the wavelength and the perturbation velocities are the order of the mean flow or the phase speed.) Therefore if the nonlinear effects do not limit the growth, the theory only indicates the initial stage of the nonlinear development.

(3) In time and space, the theory only holds for scales $O(\epsilon^{-1})$ (in the marginally unstable case). Beyond these times and distances, other nonlinear effects, which we have not calculated, might become important.†

† For a further discussion of the validity of multiple scaling techniques, see Mahony (1972).

We have derived the governing equations for the amplitude, in such a way that they take the elegant form of (2.01) and (2.02). That is, the coefficient of an nth-order derivative (with respect to a slow scale) is multiplied by the corresponding nth-order derivative of the characteristic function.† This fundamental form of the (linear) operators arises because the method of multiple scaling is basically an expansion of the (linear) operators of the original system. As shown by Newell (1972, 1974) and Weissman (1973), one may consider an arbitrary system of the form

$$\mathscr{L}(\partial/\partial \boldsymbol{x}, S)\,\phi = \text{nonlinear terms } (O(\epsilon)),$$

where $\mathscr{L}$ is a partial differential operator with constant coefficients; $\mathscr{L}$ can be considered to be a *function* of $\partial/\partial \boldsymbol{x} = (\partial/\partial x, \partial/\partial y, \partial/\partial t)$ and S, a stability parameter. The lowest order solution is

$$\phi = A \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}) + \text{c.c.},$$

where $\boldsymbol{k} = (k, \ell, -\sigma)$, provided

$$F(\boldsymbol{k}, S) = \mathscr{L}(\mathrm{i}\boldsymbol{k}, S) = 0.$$

When multiple scaling and a perturbation in S is introduced,

$$\mathscr{L} \to \hat{\mathscr{L}} = \mathscr{L}(\partial/\partial \boldsymbol{X}_0 + \mu\,\partial/\partial \boldsymbol{X}_1 + \mu^2 \partial/\partial \boldsymbol{X}_2 + \ldots, S + \Delta),$$

where $\boldsymbol{X}_n = \mu^n \boldsymbol{x}$, $|\mu| \ll 1$. This can be expanded in a multiple Taylor's expansion, i.e.

$$\hat{\mathscr{L}} = \mathscr{L}(\partial/\partial X_{0i}, S) + \mu \mathscr{L}_{,i}\,\partial/\partial X_{1i} + \mu^2(\tfrac{1}{2}\mathscr{L}_{,ij}\,\partial^2/\partial X_{1i}\,\partial X_{1j} + \mathscr{L}_{,i}\,\partial/\partial X_{2i}) + \ldots + \Delta \mathscr{L}_s,$$

(summation over $i, j = 1, 2, 3$) where the $\mathscr{L}_{,i}$, $\mathscr{L}_{,ij}$, etc., are simply the derivatives of $\mathscr{L}$ with respect to its ith, jth, etc., arguments and $\mathscr{L}_s$ is the derivative with respect to S. Therefore, when these operate on the lowest-order solution (at higher order),

$$\mathscr{L}_{,i}(\mathrm{i}\boldsymbol{k}, S) = \frac{\partial \mathscr{L}(\mathrm{i}\boldsymbol{k}, S)}{\partial(\mathrm{i}k_i)} = -\mathrm{i}\frac{\partial F(\boldsymbol{k}, S)}{\partial k_i} = -\mathrm{i}F_{,i},$$

$$\mathscr{L}_{,ij}(\mathrm{i}\boldsymbol{k}, S) = \frac{\partial^2 \mathscr{L}(\mathrm{i}\boldsymbol{k}, S)}{\partial(\mathrm{i}k_i)\,\partial(\mathrm{i}k_j)} = -\frac{\partial^2 F(\boldsymbol{k}, S)}{\partial k_i\,\partial k_j} = -F_{,ij} \text{ etc.},$$

showing that the various derivatives of F will always multiply the long-scale derivatives. This argument can easily be extended to systems of equations, to multiwave solutions, etc. Note that it is independent of the balance between μ, Δ and ϵ. As noted by Newell (1974) (and verified independently by the author), it can also be extended (not so easily) to systems having non-constant coefficients which yield 'modal' (and non-singular) solutions in one or two directions. (In this case, F can be defined as a particular integral involving the eigenfunction.)

By using the general system (2.01) and (2.02), the proper amplitude equations were found for the different regions of parameter space. The main determining factor was the vanishing or non-vanishing of the first derivative of F. In particular, at the critical point for the instability, i.e. at the minimum of the neutral surface, a second-order, hyperbolic equation was found. Since this is basically due to the coalescence of modes on the neutral surface, a necessary feature of inviscid flow, it must be typical of other 'inviscid' instabilities. However, the nonlinear terms may be different. For example, in Pedlosky's (1972) baroclinic flow, there is a second-order correction to the mean flow – a feature that is not present in the Kelvin–Helmholtz model considered here. Even though the resulting nonlinear terms are different, the solutions are quite similar. Solitary

† This was first pointed out to the author by L. F. McGoldrick in 1970.

wave packets of hyperbolic-secant shape were also found, but their speed and width depended on the amplitude in the centre of the packet. (The packets found here have fixed amplitude; the speed depends only on the width.) Thus, it is reasonable to expect that the other solutions found here are also typical of inviscid instabilities in general, although perhaps differing in detail.

The solutions to the unstable initial value problem, both linear and nonlinear, show the development of sharp fronts. These propagate at the multiple values of the group velocity (multiple values, two in each direction, arise because of the coalescence of modes). The wave packet as a whole travels with a 'convection velocity', the mean of the multiple group velocities. When the nonlinear effects are stabilizing, the numerical calculations show that the amplitude undergoes decaying oscillations in the centre of the packet about the characteristic amplitude

$$a_{\mathrm{eq}} = |F_U \Delta / N|^{\frac{1}{2}}.$$

Undulations continually propagate from the centre to the fronts, where they pack closer and closer together.

However, a dramatic change was found in this general pattern whenever the initial condition contained the hyperbolic secant. Even when the initial amplitude was very small, solitary wave packets formed and separated themselves from the main part of the solution, travelling much faster than the group velocities. This behaviour must be related to the fact that the solitary wave packet, of the form of the hyperbolic secant, is an exact solution.

These solitary solutions are reminiscent of 'solitons', the exact nonlinear solutions which emerge from rather general initial conditions in certain *conservative* wave systems (such as the Korteweg–de Vries equation; see, for example, Scott *et al.* 1973; Whitham 1974). Our solitary solutions emerge in the same way and travel with the predicted speed, but they must not be termed 'solitons' because our unstable flow does not admit the conservation principles found in the other systems. The other attributes of solitons, such as passing through each other, could not be tested because of the instability that developed in the 'tail' (see § 7.1).

Numerical experimentation also uncovered another important result. The uniform time-dependent oscillatory solutions, which have been applied by Drazin (1970) and Nayfeh & Saric (1971, 1972) to the Kelvin–Helmholtz instability and by Pedlosky (1970) to the baroclinic instability, are in fact unstable to spatial modulation. When the oscillation about a_{eq} was large enough, small (spatially dependent) perturbations were found to grow and to eventually dominate the solution.

Other new results have come from the application of the amplitude equation to the case of spatial instability. In order to have steady-state exponential growth with distance away from an oscillating source ('convective' instability), the convective velocity, which depends on one's frame of reference, must be greater than the spreading rate of wave packets. For the case of the lower layer at rest, (in the mean), we found that this condition amounts to

$$\rho_1/\rho_2 > \tfrac{1}{2}.$$

When convective instability is possible, the spatial growth rate is related to the temporal growth rate via the product of the multiple values of the group velocity:

$$\mu_{\mathrm{t}}/\mu_{\mathrm{s}} = (\sigma_k^+ \sigma_k^-)^{\frac{1}{2}}.$$

When convective instability is not possible, linear theory indicates that there is no steady-state solution ('absolute' instability). However, we have found that if nonlinear effects are considered and are stabilizing, they prevent unlimited growth and thus allow steady-state solutions. These

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

take the form of the hyperbolic tangent, which, for small initial amplitudes, grows linearly with distance from the source.

Some of the other results obtained here can be summarized, and illustrated, by considering the application of the theory to the case of air blowing over water. Kelvin's original aim in proposing the model appears to have been to explain the generation of waves by the wind, but as he himself pointed out, the critical wind velocity far exceeds that for which waves are observed. Nevertheless, the 'Kelvin–Helmholtz mechanism', a pressure distribution on the water surface in phase with the wave crests, is always present, even when waves are growing due to other effects (Miles 1959). This mechanism is most effective for waves of the order of the critical wavelength, 1.7 cm, and one would think that it must play a part in their generation.

If waves were to be generated on the air–water interface by Kelvin–Helmholtz instability, they would be highly transitory in nature. As mentioned above, in order to have a steady-state, small-amplitude (linear), spatially-growing solution in a frame of reference with the lower fluid at rest (in the mean), the density ratio must be greater than 0.5, a condition far from satisfied by air over water, for which the density ratio is 0.0012. In addition, we found in §2 that for a density ratio less than 0.283, the nonlinear effects are destabilizing. Thus, if an instability appears, for either sub- or super-critical conditions, it will not remain small. The flow will reach some highly nonlinear state that is indescribable by perturbation techniques.

It is possible that Kelvin–Helmholtz instability is responsible for 'cat's-paws', the extremely wrinkled pattern with scales of the order of 1–2 cm that appears when a gust of wind strikes the water surface. This phenomenon seems to satisfy the requirements of being transitory and highly nonlinear. The notion that spindraft, a spray of droplets over the sea surface, is caused by Kelvin–Helmholtz instability (Kelvin 1884; Phillips 1966) is also consistent with the nonlinear theory.

However, some recent works by Weissman (1976) and Valenzuela (1976) raise doubts on the suitability of the inviscid, discontinuous model to the flow of air over water. Weissman (1976) studied a continuous model with constant shear layers on each side of the interface. The flow was taken to be inviscid; however, the shear on the two sides was chosen to satisfy the viscous shear stress condition. It was found that the inclusion of boundary layers greatly destabilizes the flow. Even when the thicknesses of the boundary layers tend to zero, the flow is unstable for lower values of velocity difference than in the Kelvin–Helmholtz model (for reasonable values of surface drift velocity). This is due to the presence of a mode of instability that is not present in the original model, a coalescence of surface waves and boundary layers waves. The mode of instability found in the discontinuous model (a coalescence of surface waves) is still present, but it appears at wind speeds higher than those predicted by the original model (as also found by Miles (1959)).

Valenzuela (1976) considered a completely viscous quasi-laminar model, chosen to conform to Larson & Wright's (1975) experiment. The boundary layers on each side were log-linear profiles, suitable for turbulent flow. Since Valenzuela's results (for waves in the range 0.7 to 7.0 cm wavelength) agree very well with the experiment, it may be that shear layers and/or viscous effects are essential in predicting the instability of the air–water interface on these scales of motion.

Most of this work was completed when the author was a student in the Department of the Geophysical Sciences at the University of Chicago, as partial fulfilment of the requirements for a doctorate. I am indebted to Professor L. F. McGoldrick and Professor J. Pedlosky for many

helpful discussions and suggestions at that time. Financial support was provided by the Office of Naval Research under contract no. N00014-67-A-0285-0002, task no. NR 083-214, and computer time by the National Center for Atmospheric Research, which is sponsored by the National Science Foundation. I would also like to thank Professor J. T. Stuart, F.R.S., and Dr M. Gaster for further support and encouragement throughout the intervening years, and the Director, the National Maritime Institute, for the use of institute facilities which enabled the completion of this work.

Appendix A. Derivation of the amplitude equations

The model consists of two layers of inviscid incompressible immiscible fluids of semi-infinite extent, separated by a horizontal interface and flowing relative to each other with uniform and constant velocities. In a frame of reference imbedded in the lower fluid, the velocity of the upper fluid is U^{D}, the magnitude of the velocity difference between the layers, and is taken to be in the x-direction. (The superscript D refers to dimensional quantities.)

For irrotational disturbances, the velocity potentials of the perturbation satisfy Laplace's equation:

$$\nabla^2\phi_n = \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\phi_n = 0 \quad (n = 1, 2), \tag{A 01}$$

where $n = 1$ indicates the upper layer, $z > 0$; and $n = 2$, the lower layer, $z < 0$. The spatial variables have been non-dimensionalized by L, a typical scale for the wavelength, and the velocity potential by ϵVL, where V is the scale for U^{D} and ϵV is the magnitude of the perturbation velocities $(0 < \epsilon \ll 1)$.

The interfacial boundary conditions are evaluated at $z^{\mathrm{D}} = \zeta^{\mathrm{D}}(x, y, t)$, the position of the interface. If H is a measure of the magnitude of ζ^{D}, this becomes $z = (H/L)\,\zeta(x, y, t)$.

The kinematic condition at the interface is

$$\frac{\partial\phi_n}{\partial z} = \frac{H}{\epsilon L}\left(\frac{\partial}{\partial t}+U_n\frac{\partial}{\partial x}\right)\zeta+\epsilon\nabla\phi\cdot\nabla\zeta \quad (n = 1, 2), \tag{A 02}$$

where $U_1 = U$, $U_2 = 0$ and time has been non-dimensionalized by L/V. For the linear balance, we set $H/L = \epsilon$. Using Taylor expansions about $z = 0$, such as

$$\left.\frac{\partial\phi_n}{\partial x}\right|_{z=\epsilon\zeta} = \left.\frac{\partial\phi_n}{\partial x}\right|_0+\epsilon\zeta\left.\frac{\partial^2\phi_n}{\partial x\,\partial z}\right|_0+\tfrac{1}{2}\epsilon^2\zeta^2\left.\frac{\partial^3\phi_n}{\partial x\,\partial z^2}\right|_0+\cdots,$$

where $|_0$ indicates evaluation at $z = 0$, and using (A 01), we can write (A 02) as

$$\left.\frac{\partial\phi_n}{\partial z}\right|_0 = \left(\frac{\partial}{\partial t}+U_n\frac{\partial}{\partial x}\right)\zeta+\epsilon\{\nabla\cdot[\zeta\nabla(\phi_n|_0)]\}+\tfrac{1}{2}\epsilon^2\left\{\nabla\cdot\left[\zeta^2\nabla\left(\left.\frac{\partial\phi_n}{\partial z}\right|_0\right)\right]\right\} \quad (n = 1, 2), \tag{A 03}$$

correct to $O(\epsilon^2)$.

The dynamic interfacial condition is that any pressure difference across the interface is due to the surface tension; that is,

$$p_2^{\mathrm{D}}-p_1^{\mathrm{D}} = -\tau\left(\frac{\partial^2\zeta^{\mathrm{D}}}{\partial x^{\mathrm{D}2}}+\frac{\partial^2\zeta^{\mathrm{D}}}{\partial y^{\mathrm{D}2}}\right)\left[1+\left(\frac{\partial\zeta^{\mathrm{D}}}{\partial x^{\mathrm{D}}}\right)^2+\left(\frac{\partial\zeta^{\mathrm{D}}}{\partial y^{\mathrm{D}}}\right)^2\right]^{-\frac{3}{2}},$$

evaluated at $z^{\mathrm{D}} = \zeta^{\mathrm{D}}$, where τ is the surface tension. The total pressure can be expressed as a mean hydrostatic part plus a (non-dimensional) perturbation pressure p_n; that is,

$$p_n^{\mathrm{D}} = -\rho_n^{\mathrm{D}}gz^{\mathrm{D}}+\epsilon\rho_n^{\mathrm{D}}V^2p_n(x, y, z, t) \quad (n = 1, 2),$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

where ρ_n^{D} is the density of the layer and g is the acceleration of gravity. Then, with non-dimensionalization of the other terms, the pressure condition becomes

$$\rho_2 p_2 - \rho_1 p_1 = \tilde{g}\zeta - \tilde{\tau}(\nabla^2\zeta)\,(1+\epsilon^2\nabla\zeta\cdot\nabla\zeta)^{-\frac{3}{2}}$$

at $z = \epsilon\zeta$, where

$$\rho_n = \rho_n^{\mathrm{D}}/(\rho_1^{\mathrm{D}}+\rho_2^{\mathrm{D}}) \quad (n=1,2), \tag{A 04a}$$

$$\tilde{g} = (\rho_2^{\mathrm{D}}-\rho_1^{\mathrm{D}})\,gL/(\rho_1^{\mathrm{D}}+\rho_2^{\mathrm{D}})\,V^2, \tag{A 04b}$$

$$\tilde{\tau} = \tau/(\rho_1^{\mathrm{D}}+\rho_2^{\mathrm{D}})\,V^2L. \tag{A 04c}$$

Now expanding the surface tension term and the pressure terms, we obtain

$$\rho_2 p_2|_0 - \rho_1 p_1|_0 = \tilde{g}\zeta - \tilde{\tau}\nabla^2\zeta - \epsilon\zeta\left[\rho_2\frac{\partial p_2}{\partial z}\bigg|_0 - \rho_1\frac{\partial p_1}{\partial z}\bigg|_0\right] - \tfrac{1}{2}\epsilon^2\zeta^2\left[\rho_2\frac{\partial^2 p_2}{\partial z^2}\bigg|_0 - \rho_1\frac{\partial^2 p_1}{\partial z^2}\bigg|_0\right] + \tfrac{3}{2}\epsilon^2\tilde{\tau}(\nabla^2\zeta)\,(\nabla\zeta\cdot\nabla\zeta) \tag{A 05}$$

correct to $O(\epsilon^2)$. The kinematic pressure p_n is then related to the velocity potential by the integral of the perturbation momentum equations (i.e. Bernoulli's equation for the perturbation):

$$p_n = -\left(\frac{\partial}{\partial t}+U_n\frac{\partial}{\partial x}\right)\phi_n - \tfrac{1}{2}\epsilon(\nabla\phi_n.\nabla\phi_n) \quad (n=1,2). \tag{A 06}$$

Note that 'natural' space and velocity scales would result if we chose to make $\tilde{g} = \tilde{\tau} = 1$. For then,

$$\left.\begin{aligned} L &= [\tau/(\rho_2^{\mathrm{D}}-\rho_1^{\mathrm{D}})\,g]^{\frac{1}{2}}, \\ \text{and} \quad V &= \{[(\rho_2^{\mathrm{D}}-\rho_1^{\mathrm{D}})\,g\tau]^{\frac{1}{2}}/(\rho_1^{\mathrm{D}}+\rho_2^{\mathrm{D}})\}^{\frac{1}{2}}. \end{aligned}\right\} \tag{A 07}$$

Nevertheless, we leave L and V arbitrary for now, and thus leave $\tilde{g}$ and $\tilde{\tau}$ in the equations, in order that the individual influences of gravity and surface tension can be traced through the development.

Equations (A 01), (A 03), (A 05) and (A 06) form our basic set of equations, with the usual boundary conditions that

$$\nabla\phi_n \to 0, \quad \text{as} \quad |z|\to\infty, \quad n=1,2. \tag{A 08}$$

At this stage we introduce the two sets of slow variables as discussed in the introduction. That is, we assume that

$$\zeta = \zeta(X_0, Y_0, T_0, X_1, Y_1, T_1, X_2, Y_2, T_2),$$

$$\phi_n = \phi_n(X_0, Y_0, T_0, X_1, Y_1, T_1, X_2, Y_2, T_2, z),$$

$$p_n = p_n(X_0, Y_0, T_0, X_1, Y_1, T_1, X_2, Y_2, T_2, z),$$

where

$$(X_r, Y_r, T_r) = \epsilon^r(x, y, t) \quad (r=0,1,2).$$

Then the derivatives must be transformed; for example,

$$\frac{\partial}{\partial x} \to \frac{\partial}{\partial X_0}+\epsilon\frac{\partial}{\partial X_1}+\epsilon^2\frac{\partial}{\partial X_2} = \sum_r \epsilon^r\frac{\partial}{\partial X_r},$$

$$\frac{\partial^2}{\partial x^2} \to \left(\sum_r \epsilon^r\frac{\partial}{\partial X_r}\right)\left(\sum_s \epsilon^s\frac{\partial}{\partial X_s}\right) = \sum_{r,s}\epsilon^{r+s}\frac{\partial^2}{\partial X_r\,\partial X_s},$$

where the summation is over $r, s = 0, 1, 2$. (The reader should not be dismayed at the appearance of all these variables. Once the point of interest in parameter space is determined, some will either drop out of the final equation, or be eliminated by suitable transformations.)

A small perturbation into the unstable region could be taken by letting $U = U_m + \Delta$, $|\Delta| \ll 1$, where U_m corresponds to a point of marginal stability. However, we want the following development to be applicable to all regions of parameter space. Therefore, let us merely replace U by $U+\Delta$ in the equations. If U is marginal, $\Delta > 0$ will cause the wave (of fixed wavenumber) to be linearly unstable, but if U corresponds to a point in the stable region, Δ may be set to zero without loss of generality, for it would only give rise to a small shift in frequency or wavenumber (this can be seen from the dispersion relation, (A 13), or from the amplitude equation appropriate for the stable region, (2.05), with the Δ term retained).

With these preliminaries aside, we now introduce perturbation expansions for the dependent variables,

$$\left.\begin{aligned} \zeta &= \zeta^{(1)} + \epsilon\zeta^{(2)} + \epsilon^2\zeta^{(3)} + \ldots, \\ \phi_n &= \phi_n^{(1)} + \ldots, \\ p_n &= p_n^{(1)} + \ldots, \end{aligned}\right\} \tag{A 09}$$

and the expansions of the derivatives into (A 01), (A 03), (A 05), (A 06) and (A 08). The equations separate into successive systems of linear equations as follows:

First order, $O(1)$ (*the linear problem*).

$$\nabla^2\phi_n^{(1)} = 0, \tag{A 10a}$$

$$\partial\phi_n^{(1)}/\partial z|_0 - D_n\zeta^{(1)} = 0, \tag{A 10b}$$

$$p_n^{(1)} + D_n\phi_n^{(1)} = 0, \tag{A 10c}$$

$$\rho_2 p_2^{(1)}|_0 - \rho_1 p_1^{(1)}|_0 - \tilde{g}\zeta^{(1)} + \tilde{\tau}\nabla^2\zeta^{(1)} = 0, \tag{A 10d}$$

$$\nabla\phi_n^{(1)} \to 0 \quad \text{as} \quad |z| \to \infty, \tag{A 10e}$$

where

$$D_n = \partial/\partial t + U_n\,\partial/\partial x,$$

x, y, t has replaced X_0, Y_0, T_0, and it is understood that n takes on the values 1 and 2. A single-wave solution satisfying (A 10*a*), (A 10*b*), (A 10*c*) and (A 10*e*) is

$$\zeta^{(1)} = A\exp(\mathrm{i}\theta) + \text{c.c.}, \tag{A 11a}$$

$$\phi_n^{(1)} = B_n\exp(\mathrm{i}\theta + m_n z) + \text{c.c.}, \tag{A 11b}$$

$$p_n^{(1)} = C_n\exp(\mathrm{i}\theta + m_n z) + \text{c.c.}, \tag{A 11c}$$

where

$$\theta = kx + \ell y - \sigma t, \quad m_n = (-1)^n\kappa, \quad \kappa = +\sqrt{(k^2+\ell^2)},$$

$$B_n = -\mathrm{i}\frac{(\sigma - U_n k)}{m_n}A, \quad C_n = \frac{(\sigma - U_n k)^2}{m_n}A,$$

and c.c. indicates the complex conjugate of all preceding terms. A, B_n and C_n are functions of the slow variables. Substitution into (A 10*d*) gives the characteristic equation,

$$F(\sigma, k, \ell, U) = \rho_2\sigma^2 + \rho_1(\sigma - Uk)^2 - (\tilde{g}\kappa + \tilde{\tau}\kappa^3) = 0, \tag{A 12}$$

and from this, the dispersion relation,

$$\sigma = \rho_1 Uk \pm (\tilde{g}\kappa + \tilde{\tau}\kappa^3 - \rho_1\rho_2 U^2k^2)^{\frac{1}{2}} \tag{A 13}$$

(where use has been made of the identity $\rho_1 + \rho_2 = 1$).

The points of marginal stability lie on the 'neutral surface',

$$U = U_m(k, \ell) = [(\tilde{g}\kappa + \tilde{\tau}\kappa^3)/\rho_1\rho_2 k^2]^{\frac{1}{2}}. \tag{A 14}$$

The cross-section of this surface in the $U-k$ plane is the 'neutral curve', as depicted in figure 1. U_m has a minimum at

$$k = (\tilde{g}/\tilde{\tau})^{\frac{1}{2}} = k_c, \quad \ell = 0 = \ell_c;$$

thus, the critical velocity difference for instability is

$$U = [2(\tilde{g}\tilde{\tau})^{\frac{1}{2}}/\rho_1\rho_2]^{\frac{1}{2}} = U_c.$$

In dimensional terms,

$$k_c^D = k_c/L = [(\rho_2^D - \rho_1^D)g/\tau]^{\frac{1}{2}},$$

$$U_c^D = U_c V = \{2[(\rho_2^D - \rho_1^D)\, g\tau]^{\frac{1}{2}} (\rho_1^D + \rho_2^D)/\rho_1^D\rho_2^D\}^{\frac{1}{2}}.$$

Second order, $O(\epsilon)$

$$\nabla^2\phi_n^{(2)} = -2(\nabla\cdot\nabla_1\phi_n^{(1)}), \tag{A 15a}$$

$$\partial\phi_n^{(2)}/\partial z|_0 - D_n\zeta^{(2)} = D_{1n}\zeta^{(1)} + \nabla.[\zeta^{(1)}\nabla(\phi_n^{(1)}|_0)], \tag{A 15b}$$

$$p_n^{(2)} + D_n\phi_n^{(2)} = -D_{1n}\phi_n^{(1)} - \tfrac{1}{2}(\nabla\phi_n^{(1)}\cdot\nabla\phi_n^{(1)}), \tag{A 15c}$$

$$\rho_2 p_2^{(2)}|_0 - \rho_1 p_1^{(2)}|_0 - \tilde{g}\zeta^{(2)} + \tilde{\tau}\nabla^2\zeta^{(2)} = -2\tilde{\tau}\nabla\cdot\nabla_1\zeta^{(1)} - \zeta^{(1)}\left[\rho_2\frac{\partial p_2^{(1)}}{\partial z}\bigg|_0 - \rho_1\frac{\partial p_1^{(1)}}{\partial z}\bigg|_0\right], \tag{A 15d}$$

$$\nabla\phi_n^{(2)} \to -\nabla_1\phi_n^{(1)}, \quad \text{as} \quad |z| \to \infty, \tag{A 15e}$$

where

$$D_{1n} = \frac{\partial}{\partial T_1} + U_n\frac{\partial}{\partial X_1}, \quad \nabla_1 = \left(\frac{\partial}{\partial X_1}, \frac{\partial}{\partial Y_1}, 0\right).$$

Evaluating the right hand sides using (A 11),

$$\nabla^2\phi_n^{(2)} = -2i\boldsymbol{k}\cdot\nabla_1 B_n \exp(i\theta + m_n z) + \text{c.c.}, \tag{A 16a}$$

$$\frac{\partial\phi_n^{(2)}}{\partial z}\bigg|_0 - D_n\zeta^{(2)} = D_{1n}A\exp(i\theta) - 2\kappa^2 A B_n \exp(2i\theta) + \text{c.c.}, \tag{A 16b}$$

$$p_n^{(2)} + D_n\phi_n^{(2)} = -[D_{1n}B_n\exp(i\theta + m_n z) + \text{c.c.}] - 2\kappa^2|B_n|^2\exp(2m_n z), \tag{A 16c}$$

$$\begin{aligned}\rho_2 p_2^{(2)}|_0 - \rho_1 p_1^{(2)}|_0 - \tilde{g}\zeta^{(2)} + \tilde{\tau}\nabla^2\zeta^{(2)} \\ = -2\tilde{\tau}i\boldsymbol{k}\cdot\nabla_1 A\exp(i\theta) - \kappa(\rho_2 C_2 + \rho_1 C_1)[A\exp(2i\theta) + A^*] + \text{c.c.},\end{aligned} \tag{A 16d}$$

$$\nabla\phi_n^{(2)} \to 0, \quad \text{as} \quad |z| \to \infty, \tag{A 16e}$$

where

$$\boldsymbol{k} = (k, \ell)$$

and A^* is the complex conjugate of A.

The forcing terms on the right hand sides are of three types: mean, harmonic (or fundamental) and second harmonic. The harmonic terms could produce resonance, and thus a secularity, in our perturbation expansion. However, in seeking the harmonic part of the solution, we will find an equation that the amplitude must satisfy in order to 'remove' the secularity. To find the harmonic part, let

$$\zeta^{(2)} = A_1^{(2)}\exp(i\theta) + \text{c.c.},$$

$$\phi_n^{(2)} = (B_{n1}^{(2)} + F_n^{(2)}z)\exp(i\theta + m_n z) + \text{c.c.}$$

The term linear in z in the coefficient of $\phi_n^{(2)}$ is necessary to satisfy (A 16a). Although this term arises because of resonant forcing in (A 16a), it is not secular because $\phi_n^{(2)}$ satisfies the boundary conditions (A 16e). (An alternative approach would be to eliminate the term linear in z by allowing B_n to be a function of a long vertical variable, $Z_1 = \epsilon z$.)

Considering only the harmonic terms, (A 16*b*) gives $B_{n1}^{(2)}$ in terms of $A_1^{(2)}$ and the forcing terms, and (A 16*c*) in turn determines the pressure. Substitution of the pressure into (A 16*d*) yields (after a bit of algebra)

$$\begin{aligned}[\rho_2\sigma^2+\rho_1(\sigma-U\ell)^2-(\tilde{g}\kappa+\tilde{\tau}\kappa^3)]A_1^{(2)} &\\ = -\mathrm{i}\{&[2\rho_2\sigma+2\rho_1(\sigma-U\ell)]\,\partial A/\partial T_1\\ &+[2\rho_1(\sigma-U\ell)\,U+(\tilde{g}+3\tilde{\tau}\kappa^2)\,\ell/\kappa]\,\partial A/\partial X_1\\ &+[(\tilde{g}+3\tilde{\tau}\kappa^2)\,\ell/\kappa]\,\partial A/\partial Y_1\}. \end{aligned} \tag{A 17}$$

The coefficient of $A_1^{(2)}$ vanishes since it is the characteristic function. Therefore the right hand side of (A 17) must also vanish, producing a condition to be satisfied by A with respect to X_1, Y_1 and T_1. It can be seen that the coefficients here are the various derivatives of the characteristic function; that is, this condition can be written

$$\frac{\partial F}{\partial\sigma}\frac{\partial A}{\partial T_1}-\frac{\partial F}{\partial\ell}\frac{\partial A}{\partial X_1}-\frac{\partial F}{\partial\ell}\frac{\partial A}{\partial Y_1}=0. \tag{A 18}$$

$A_1^{(2)}$ is arbitrary to this order, but since it merely repeats the lowest-order solution, we may set it to zero. (This may be thought of as a normalizing condition.) The rest of the harmonic solution then represents a phase shift of the velocity potentials and pressures with respect to $\zeta^{(1)}$ when the amplitude is changing with time or space. Completing the solution for the mean and second-harmonic forcing, we find the total solution at this order is

$$\zeta^{(2)}=\gamma\kappa A^2\exp(2\mathrm{i}\theta)+\text{c.c.}, \tag{A 19a}$$

$$\phi_n^{(2)}=(B_{n1}^{(2)}+F_n^{(2)}z)\exp(\mathrm{i}\theta+m_n z)+B_{n2}^{(2)}\exp(2\mathrm{i}\theta+2m_n z)+\text{c.c.}, \tag{A 19b}$$

$$\begin{aligned}p_n^{(2)}=&-2\kappa^2|B_n|^2\exp(2m_n z)+[(C_{n1}^{(2)}+G_n^{(2)}z)\\ &\times\exp(\mathrm{i}\theta+m_n z)+C_{n2}^{(2)}\exp(2\mathrm{i}\theta+2m_n z)+\text{c.c.}],\end{aligned} \tag{A 19c}$$

where

$$\gamma=[\rho_1(\sigma-U\ell)^2-\rho_2\sigma^2]/(2\tilde{\tau}\kappa^3-\tilde{g}\kappa),$$

$$B_{n1}^{(2)}=[\mathrm{D}_{1n}A+(\sigma-U_n\ell)\,(\boldsymbol{k}\cdot\nabla_1 A)/\kappa^2]/m_n,$$

$$F_n^{(2)}=-(\sigma-U_n\ell)\,(\boldsymbol{k}\cdot\nabla_1 A)/\kappa^2,$$

$$B_{n2}^{(2)}=\mathrm{i}(\sigma-U_n\ell)\,(1-\gamma m_n/\kappa)\,A^2,$$

$$C_{n1}^{(2)}=\mathrm{i}\frac{(\sigma-U_n\ell)}{m_n}\left[2\mathrm{D}_{1n}A+\frac{(\sigma-U_n\ell)}{\kappa^2}(\boldsymbol{k}\cdot\nabla_1 A)\right],$$

$$G_n^{(2)}=-\mathrm{i}\frac{(\sigma-U_n\ell)^2}{\kappa^2}(\boldsymbol{k}\cdot\nabla_1 A),$$

$$C_{n2}^{(2)}=-2(\sigma-U_n\ell)^2\,(1-\gamma m_n/\kappa)\,A^2.$$

Note that the second-harmonic coefficient is singular when

$$\kappa=(\tilde{g}/2\tilde{\tau})^{\frac{1}{2}}=\kappa_{\mathrm{res}}.$$

This indicates that the second-harmonic forcing terms are secular for this value of κ, or in other words, that the second-harmonic is *also* a solution of the homogeneous equations. This secularity, called 'second-harmonic resonance', can be handled by allowing both the fundamental and the

second harmonic to appear in the lowest-order solution, and then their products appear in the second-order equations. For stable waves (i.e. below the neutral surface), the situation is very similar to that for surface gravity-capillary waves and has been well studied (see McGoldrick 1970 for a review). However, since $\kappa_{\rm res}$ is independent of U, second-harmonic resonance can also occur for marginal and unstable waves and this has not been studied.† We will not consider this case in detail here as our main concern is the effect of nonlinear self-interaction, but in appendix B we consider the form the amplitude equations would take. For the most part, then, and particularly for the analysis of this section, we assume that

$$\kappa \neq \kappa_{\rm res}.$$

To be more precise, since $\epsilon\zeta^{(2)}$ must be very much less than 1, we must have $(\kappa-\kappa_{\rm res}) \gg \epsilon$.

Third order, $O(\epsilon^2)$

$$\nabla^2\phi_n^{(3)} = -2\nabla\cdot\nabla_2\phi_n^{(1)} - 2\nabla\cdot\nabla_1\phi_n^{(2)} - \nabla_1^2\phi_n^{(1)}, \tag{A 20a}$$

$$\left.\frac{\partial\phi_n^{(3)}}{\partial z}\right|_0 - \mathrm{D}_n\zeta^{(3)} = \mathrm{D}_{2n}\zeta^{(1)} + \mathrm{D}_{1n}\zeta^{(2)} + \Delta_n\frac{\partial\zeta^{(1)}}{\partial x} + \nabla\cdot[\zeta^{(1)}\nabla(\phi_n^{(2)}|_0)] + \nabla\cdot[\zeta^{(2)}\nabla(\phi_n^{(1)}|_0)]$$
$$+ \nabla\cdot[\zeta^{(1)}\nabla_1(\phi_n^{(1)}|_0)] + \nabla_1\cdot[\zeta^{(1)}\nabla(\phi_n^{(1)}|_0)] + \tfrac{1}{2}\nabla\cdot\left[(\zeta^{(1)})^2\nabla\left(\left.\frac{\partial\phi_n^{(1)}}{\partial z}\right|_0\right)\right], \tag{A 20b}$$

$$p_n^{(3)} + \mathrm{D}_n\phi_n^{(3)} = -\mathrm{D}_{2n}\phi_n^{(1)} - \mathrm{D}_{1n}\phi_n^{(2)} - \Delta_n\frac{\partial\phi_n^{(1)}}{\partial x} - \nabla\phi_n^{(1)}\cdot\nabla\phi_n^{(2)} - \nabla\phi_n^{(1)}\cdot\nabla_1\phi_n^{(1)}, \tag{A 20c}$$

$$\rho_2 p_2^{(3)}|_0 - \rho_1 p_1^{(3)}|_0 - \tilde{g}\zeta^{(3)} + \tilde{\tau}\nabla^2\zeta^{(3)}$$
$$= -2\tilde{\tau}\nabla\cdot\nabla_2\zeta^{(1)} - 2\tilde{\tau}\nabla\cdot\nabla_1\zeta^{(2)} - \tilde{\tau}\nabla_1^2\zeta^{(1)} - \zeta^{(1)}\left[\rho_2\left.\frac{\partial p_2^{(2)}}{\partial z}\right|_0 - \rho_1\left.\frac{\partial p_1^{(2)}}{\partial z}\right|_0\right]$$
$$- \zeta^{(2)}\left[\rho_2\left.\frac{\partial p_2^{(1)}}{\partial z}\right|_0 - \rho_1\left.\frac{\partial p_1^{(1)}}{\partial z}\right|_0\right] - \tfrac{1}{2}(\zeta^{(1)})^2\left[\rho_2\left.\frac{\partial^2 p_2^{(1)}}{\partial z^2}\right|_0 - \rho_1\left.\frac{\partial^2 p_1^{(1)}}{\partial z^2}\right|_0\right]$$
$$+ \tfrac{3}{2}\tilde{\tau}(\nabla^2\zeta^{(1)})\,(\nabla\zeta^{(1)}\cdot\nabla\zeta^{(1)}), \tag{A 20d}$$

$$\nabla\phi_n^{(3)} \to -\nabla_1\phi_n^{(2)} - \nabla_2\phi_n^{(1)}, \quad \text{as} \quad |z|\to\infty, \tag{A 20e}$$

where $$\mathrm{D}_{2n} = \frac{\partial}{\partial T_2} + U_n\frac{\partial}{\partial X_2}, \quad \nabla_2 = \left(\frac{\partial}{\partial X_2}, \frac{\partial}{\partial Y_2}, 0\right), \quad \Delta_n = \begin{cases}\Delta/\epsilon^2, & n=1,\\ 0, & n=2.\end{cases}$$

The right hand sides are now evaluated by using the lower-order solutions. Fortunately we need not write out all of the terms. The harmonic forcing terms will again produce a secularity, the removal of which gives the amplitude equation we are seeking. Therefore, we write out only the harmonic terms, using '…' to indicate mean, second-harmonic and third-harmonic terms:

$$\nabla^2\phi_n^{(3)} = \{-2\mathrm{i}\boldsymbol{k}\cdot\nabla_2 B_n - 2\mathrm{i}\boldsymbol{k}\cdot\nabla_1 B_{n1}^{(2)} - \nabla_1^2 B_n - 2\mathrm{i}(\boldsymbol{k}\cdot\nabla_1 F_n^{(2)})z\}\exp(\mathrm{i}\theta + m_n z) + \text{c.c.} + \ldots, \tag{A 21a}$$

$$\partial\phi_n^{(3)}/\partial z|_0 - D_n\zeta^{(3)} = \{\mathrm{D}_{2n}A + \mathrm{i}\ell\Delta_n A + \kappa^2(\gamma\kappa + \tfrac{1}{2}m_n)\,B_n^* A^2$$
$$- 2\kappa^2 B_{n2}^{(2)} A^* - \kappa^2 m_n B_n|A|^2\}\exp(\mathrm{i}\theta) + \text{c.c.} + \ldots, \tag{A 21b}$$

$$p_n^{(3)} + \mathrm{D}_n\phi_n^{(3)} = \{-\mathrm{D}_{2n}B_n - \mathrm{D}_{1n}B_{n1}^{(2)} - (\mathrm{D}_{1n}F_n^{(2)})\,z - \mathrm{i}\ell\Delta_n B_n\}$$
$$\times\exp(\mathrm{i}\theta + m_n z) - 4\kappa^2 B_n^* B_{n2}^{(2)}\exp(\mathrm{i}\theta + 3m_n z) + \text{c.c.} + \ldots, \tag{A 21c}$$

† Nayfeh & Saric (1972) consider second-harmonic resonance, but their analysis only holds in the stable region (see appendix B). Kelly (1967) has considered second-harmonic resonance of marginal waves in other unstable shear flows.

$$\rho_2 p_2^{(3)}|_0 - \rho_1 p_1^{(3)}|_0 - \tilde{g}\zeta^{(3)} + \tilde{\tau}\nabla^2\zeta^{(3)} = \{ -2\tilde{\tau}\mathrm{i}\boldsymbol{k}\cdot\nabla_2 A$$
$$-\tilde{\tau}\nabla_1^2 A + 4\kappa^3(\rho_2|B_2|^2 + \rho_1|B_1|^2)\, A$$
$$-2\kappa(\rho_2 C_{22}^{(2)} + \rho_1 C_{12}^{(2)})\, A^* - \gamma\kappa^2(\rho_2 C_2^* + \rho_1 C_1^*)\, A^2$$
$$-\kappa^2(\rho_2 C_2 - \rho_1 C_1)|A|^2 - \tfrac{1}{2}\kappa^2(\rho_2 C_2^* - \rho_1 C_1^*)\, A^2$$
$$-\tfrac{3}{2}\tilde{\tau}\kappa^4|A|^2 A\}\exp(\mathrm{i}\theta) + \text{c.c.} + \ldots, \qquad \text{(A 21}d\text{)}$$

$$\nabla\phi_n^{(3)} \to 0, \quad |z| \to \infty. \qquad \text{(A 21}e\text{)}$$

To satisfy (A 21a), $\phi_n^{(3)}$ must take the form

$$\phi_n^{(3)} = [B_{n1}^{(3)} + F_n^{(3)} z + H_n^{(3)} z^2]\exp(\mathrm{i}\theta + m_n z) + \text{c.c.}$$

$F_n^{(3)}$ and $H_n^{(3)}$ are determined from (A 21a) and $B_{n1}^{(3)}$ from (A 21b). Then, proceeding as before, the pressure is calculated from (A 21c) and substituted into (A 21d). Finally (after *quite* a bit of algebra; the coefficients must all be replaced by their definitions in terms of A), the amplitude equation appears, but it is quite complicated. Use of (A 18) clears it up somewhat, and the following form is found:

$$-\mathrm{i}\left(F_\sigma\frac{\partial A}{\partial T_2} - F_k\frac{\partial A}{\partial X_2} - F_\ell\frac{\partial A}{\partial Y_2}\right) + \tfrac{1}{2}F_{\sigma\sigma}\frac{\partial^2 A}{\partial T_1^2} - F_{\sigma k}\frac{\partial^2 A}{\partial X_1\,\partial T_1} + \tfrac{1}{2}F_{kk}\frac{\partial^2 A}{\partial X_1^2} + F_{k\ell}\frac{\partial^2 A}{\partial X_1\,\partial Y_1}$$
$$+\tfrac{1}{2}F_{\ell\ell}\frac{\partial^2 A}{\partial Y_1^2} - F_{\sigma\ell}\frac{\partial^2 A}{\partial Y_1\,\partial T_1} = (\Delta/\epsilon^2)\, F_U\, A + N|A|^2 A \qquad \text{(A 22)}$$

The coefficients of the linear terms are simply derivatives of the characteristic function,

$$F = \rho_2\sigma^2 + \rho_1(\sigma - Uk)^2 - (gk + \tau k^3),$$

and are as follows:

$$\left.\begin{aligned}
F_\sigma &= 2(\sigma - \rho_1 Uk),\\
F_k &= -2\rho_1 U(\sigma - Uk) - (\tilde{g} + 3\tilde{\tau}\kappa^2)\, k/\kappa,\\
F_\ell &= -(\tilde{g} + 3\tilde{\tau}\kappa^2)\, \ell/\kappa,\\
F_{\sigma\sigma} &= 2, \quad F_{\sigma k} = -2\rho_1 U,\\
F_{kk} &= 2\rho_1 U^2 + (\tilde{g} - 3\tilde{\tau}\kappa^2)\, k^2/\kappa^3 - (\tilde{g} + 3\tilde{\tau}\kappa^2)/\kappa,\\
F_{k\ell} &= (\tilde{g} - 3\tilde{\tau}\kappa^2)\, \ell k/\kappa^3,\\
F_{\ell\ell} &= (\tilde{g} - 3\tilde{\tau}\kappa^2)\, \ell^2/\kappa^3 - (\tilde{g} + 3\tilde{\tau}\kappa^2)/\kappa,\\
F_{\sigma\ell} &= 0, \quad F_U = -2\rho_1 k(\sigma - Uk),
\end{aligned}\right\} \qquad \text{(A 23)}$$

where use has been made of $\rho_1 + \rho_2 = 1$. (The $F_{\sigma\ell}$ term is included in (A 22) for the sake of illustration. In the Kelvin–Helmholtz model it does not appear ($F_{\sigma\ell} \equiv 0$), but, by the arguments given in the summary, it might be present in other problems.) The nonlinear coefficient is

$$N = \tfrac{1}{2}\kappa^2\left(-4\tilde{g}\kappa - \tilde{\tau}\kappa^3 + 4\frac{[\rho_2\sigma^2 - \rho_1(\sigma - Uk)^2]^2}{2\tilde{\tau}\kappa^3 - \tilde{g}\kappa}\right). \qquad \text{(A 24)}$$

Equation (A 22) must be used in conjunction with (A 18).

Appendix B. Second-harmonic resonance

Equation (2.09), the time-dependent-only amplitude equation for waves anywhere on the neutral surface, fails if $\kappa^2 = \tilde{g}/2\tilde{\tau} = \kappa_{\rm res}^2$. This is due to 'second-harmonic resonance', the situation whereby both the fundamental wave and its second-harmonic are free wave solutions (i.e. both satisfy the dispersion relation). The perturbation expansion fails at second-order because of resonant forcing by the nonlinear terms. (See McGoldrick (1970) for a review in the context of ordinary surface gravity-capillary waves.)

Nayfeh & Saric (1972) have considered second-harmonic resonance for the Kelvin–Helmholtz flow. Their equations – first order in time and essentially the same as McGoldrick's – hold in the stable region, but as the neutral curve is approached, their nonlinear coefficients become singular. This is due to the vanishing of F_σ on the neutral surface (F_σ has been implicitly divided through on their equations). We have seen in our more general approach (e.g. equations (2.01), (2.02)) that a first time derivative on a long scale must be multiplied by the first derivative of the characteristic function with respect to the frequency, etc. Thus, on the neutral surface, we must have a second-order equation, even for second-harmonic resonance.

But the second-harmonic resonance does imply that a rescaling is needed between ϵ and Δ. The secularity now arises at second order, and to achieve a balance between the instability and the nonlinearity, we must bring in the terms proportional to Δ at this order. Thus we must have $\Delta = O(\epsilon)$. The long time scale must also appear at the same order and the operator must be the second derivative; the choice $T_1 = \epsilon^{\frac{1}{2}} t$ provides for this. (Note that this implies that the expansions for ζ and the other functions will be in powers of $\epsilon^{\frac{1}{2}}$.) Thus, in terms of ϵ, the growth rate is much larger; however, in terms of Δ, it is the same, $O(\Delta^{\frac{1}{2}})$. The nonlinear terms will be of the same form as those for second-harmonic resonance of surface waves, but the coefficients will be different. It follows that the equations for the amplitudes of the fundamental, $A_1(T_1)$, and of the second harmonic, $A_2(T_1)$, when they are resonant and slightly unstable will be of the form

$$\tfrac{1}{2} F_{\sigma\sigma 1}\, \mathrm{d}^2 A_1/\mathrm{d}\, T_1^2 = (\Delta/\epsilon)\, F_{U1}\, A_1 + N_1 A_1^* A_2,$$

$$\tfrac{1}{2} F_{\sigma\sigma 2}\, \mathrm{d}^2 A_2/\mathrm{d}\, T_1^2 = (\Delta/\epsilon)\, F_{U2}\, A_2 + N_2 A_1^2.$$

The linear coefficients are the derivatives of the characteristic equation, as before, but those with subscript 1 are evaluated at $\boldsymbol{k} = (k_1, \ell_1)$, where $k_1^2 + \ell_1^2 = \kappa_{\rm res}^2$, and those with subscript 2 are evaluated at $\boldsymbol{k} = (2k_1, 2\ell_1)$. The level of shear is given by

$$U = U_{\rm m}(k_1, \ell_1) = U_{\rm m}(2k_1, 2\ell_1) = \left[\frac{g\kappa_{\rm res} + \tau\kappa_{\rm res}^3}{\rho_1 \rho_2\, k_1^2}\right]^{\frac{1}{2}}.$$

The nonlinear coefficients must be determined by a separate analysis.

References

Abramowitz, M. & Stegun, I. A. 1964 *Handbook of mathematical functions*. Washington: National Bureau of Standards.

Benney, D. J. & Newell, A. C. 1967 The propagation of nonlinear wave envelopes. *J. Math. Phys.* **46**, 133.

Benney, D. J. & Roskes, G. J. 1969 Wave instabilities. *Stud. appl. Math.* **48**, 377.

Bers, A. 1975 Linear waves and instabilities. In *Physique de Plasmas* (ed. C. De Witt & J. Peyraud). New York: Gordon and Breach.

Briggs, R. J. 1964 *Electron–stream interaction with plasmas*, Research Monograph No. 29. Cambridge, Mass: The M.I.T. Press.

Chu, V. H. & Mei C. C. 1970 On slowly-varying Stokes waves. *J. Fluid Mech.* **41**, 873.

Chu, V. H. & Mei, C. C. 1971 The non-linear evolution of Stokes waves in deep water. *J. Fluid Mech.* **47**, 337.
Cole, J. D. 1968 *Perturbation methods in applied mathematics*. Blaisdell–Ginn.
Davey, A. & Stewartson, K. 1974 On three-dimensional packets of surface waves. *Proc. R. Soc. Lond.* A **338**, 101.
Drazin, P. G. 1970 Kelvin–Helmholtz instability of finite amplitude. *J. Fluid Mech.* **42**, 321.
Fleishman, B. A. 1959 Progressing waves in an infinite nonlinear string. *Proc. A.M.S.* **10**, 329.
Gallagher, A. P. & Mercer, A.McD. 1962 On the behaviour of small disturbances in plane Couette flow. *J. Fluid Mech.* **13**, 91.
Gallagher, A. P. & Mercer, A.McD. 1964 On the behaviour of small disturbances in plane Couette flow. Part 2. The higher eigenvalues. *J. Fluid Mech.* **18**, 350.
Gaster, M. 1962 A note on the relationship between temporally-increasing and spatially-increasing disturbances in hydrodynamic stability. *J. Fluid Mech.* **14**, 222.
Hasimoto, H. & Ono, H. 1972 Nonlinear modulation of gravity waves. *J. Phys. Soc. Japan* **33**, 805.
Hocking, L. M., Stewartson, K. & Stuart, J. T. 1972 A nonlinear instability burst in plane parallel flow. *J. Fluid Mech.* **51**, 705.
Kelly, R. E. 1967 On the resonant interaction of neutral disturbances in two inviscid shear flows. *J. Fluid Mech.* **31**, 789.
Kelvin, Lord 1871 The influence of wind on waves in water supposed frictionless. *Phil. Mag.* (4) **42**, 368. Also in *Mathematical and physical papers* vol. 4, p. 76. Cambridge University Press.
Kelvin, Lord 1884 *Baltimore Lectures*, p. 592. Cambridge University Press (1904).
Lange, C. G. & Newell, A. C. 1971 The post-buckling problem for thin elastic shells. *SIAM J. appl. Math.* **21**, 605.
Lange, C. G. & Newell, A. C. 1974 A stability criterion for envelope equations. *SIAM J. appl. Math.* **27**, 441.
Larson, T. R. & Wright, J. W. 1975 Wind-generated gravity-capillary waves: laboratory measurements of temporal growth rates using microwave backscatter. *J. Fluid Mech.* **70**, 417.
Mahony, J. J. 1972 Validity of averaging methods for certain systems with periodic solutions. *Proc. R. Soc. Lond.* A **330**, 349.
McGoldrick, L. F. 1970 An experiment on second-order capillary-gravity resonant wave interactions. *J. Fluid Mech.* **40**, 251.
Miles, J. W. 1957 On the generation of surface waves by shear flows. *J. Fluid Mech.* **3**, 185.
Miles, J. W. 1959 On the generation of surface waves by shear flows. Part 3. *J. Fluid Mech.* **6**, 583.
Milne-Thomson, L. M. 1950 *Jacobian elliptic function tables*, New York: Dover.
Nayfeh, A. H. & Saric, W. S. 1971 Nonlinear Kelvin–Helmholtz instability. *J. Fluid Mech.* **46**, 209.
Nayfeh, A. H. & Saric, W. S. 1972 Nonlinear waves in a Kelvin–Helmholtz flow. *J. Fluid Mech.* **55**, 311.
Newell, A. C. 1972 The post bifurcation state of baroclinic instability. *J. atmos. Sci.* **29**, 64.
Newell, A. C. 1974 Envelope equations. *Lect. appl. Math.* **15**, 157.
Newell, A. C. & Whitehead, J. A. 1969 Finite amplitude, finite bandwidth convection. *J. Fluid Mech.* **38**, 279.
Pedlosky, J. 1970 Finite-amplitude baroclinic waves. *J. atmos. Sci.* **27**, 15.
Pedlosky, J. 1972 Finite-amplitude baroclinic wave packets. *J. atmos. Sci.* **29**, 680.
Phillips, O. M. 1966 *The dynamics of the upper ocean*, Cambridge University Press.
Rayleigh, Lord 1880 On the stability, or instability, of certain fluid motions. *Proc. Lond. Math. Soc.* **2**, 57.
Scott, A. C., Chu, F. Y. & McLaughlin, D. W. 1973 The soliton: a new concept in applied science. *Proc. I.E.E.E.* **61**, 1443.
Stewartson, K. & Stuart, J. T. 1971 A nonlinear instability theory for a wave system in plane Pouseuille flow. *J. Fluid Mech.* **48**, 529.
Stuart, J. T. 1960 On the non-linear mechanics of wave disturbances in stable and unstable flows. Part 1. *J. Fluid Mech.* **9**, 353.
Thorpe, S. A. 1969 Experiments on the instability of stratified shear flows: immiscible fluids. *J. Fluid Mech.* **39**, 25.
Valenzuela, G. R. 1976 The growth of gravity-capillary waves in a coupled shear flow. *J. Fluid Mech.* **76**, 229.
Watanabe, T. 1969 A nonlinear theory of two-stream instability. *J. Phys. Soc. Japan* **27**, 1341.
Weissman, M. A. 1972 Nonlinear development of the Kelvin–Helmholtz instability. *Proceedings of Conference on Mathematical Topics in Stability Theory*. The Washington State University.
Weissman, M. A. 1973 Nonlinear wave packets in Kelvin–Helmholtz flow. Ph.D. thesis, The University of Chicago.
Weissman, M. A. 1976 A simple model for the generation of ripples. *Mem. Soc. R. Sci. Liège* **10**, 287. Also available as Report R135, Division of Maritime Science, The National Physical Laboratory, November 1975.
Whitham, G. B. 1974 *Linear and nonlinear waves*. Englewood Cliffs, N.J.: Prentice–Hall.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES