J. Fluid Mech. (2005), *vol.* 539, *pp.* 229–271.
doi:10.1017/S002211200500563X Printed in the United Kingdom

Stabilizing the Benjamin–Feir instability

By HARVEY SEGUR[1], DIANE HENDERSON[2], JOHN CARTER[3], JOE HAMMACK[2]†, CONG-MING LI[1], DANA PHEIFF[2] AND KATHERINE SOCHA[4]

[1]Department of Applied Mathematics, University of Colorado, Boulder, CO 80309-0526, USA

[2]William G. Pritchard Fluid Mechanics Laboratory, Department of Mathematics, Penn State University, University Park, PA 16802, USA

[3]Department of Mathematics, Seattle University, Seattle, WA 98122-4340, USA

[4]Mathematics Department, Lyman Briggs School, Michigan State University, East Lansing, MI 48825-1107, USA

(Received 19 February 2004 and in revised form 6 April 2005)

The Benjamin–Feir instability is a modulational instability in which a uniform train of oscillatory waves of moderate amplitude loses energy to a small perturbation of other waves with nearly the same frequency and direction. The concept is well established in water waves, in plasmas and in optics. In each of these applications, the nonlinear Schrödinger equation is also well established as an approximate model based on the same assumptions as required for the derivation of the Benjamin–Feir theory: a narrow-banded spectrum of waves of moderate amplitude, propagating primarily in one direction in a dispersive medium with little or no dissipation. In this paper, we show that for waves with narrow bandwidth and moderate amplitude, any amount of dissipation (of a certain type) stabilizes the instability. We arrive at this stability result first by proving it rigorously for a damped version of the nonlinear Schrödinger equation, and then by confirming our theoretical predictions with laboratory experiments on waves of moderate amplitude in deep water. The Benjamin–Feir instability is often cited as the first step in a nonlinear process that spreads energy from an initially narrow bandwidth to a broader bandwidth. In this process, sidebands grow exponentially until nonlinear interactions eventually bound their growth. In the presence of damping, this process might still occur, but our work identifies another possibility: damping can stop the growth of perturbations before nonlinear interactions become important. In this case, if the perturbations are small enough initially, then they never grow large enough for nonlinear interactions to become important.

1. Introduction

In a landmark paper, Benjamin & Feir (1967) showed that a uniform train of plane waves of moderate amplitude in deep water without dissipation is unstable to a small perturbation of other waves travelling in the same direction with nearly the same frequency. The instability is a finite-amplitude effect, in the sense that the unperturbed wavetrain (which we call the 'carrier wave') must have finite amplitude, and the growth rate of the instability is proportional to the square of that amplitude, at least for small amplitudes. Corroborating theoretical results were obtained about the same

† Professor Hammack died during the publication process of this paper.

time by Lighthill (1965), Ostrovsky (1967), Whitham (1967), Benney & Newell (1967), and Zakharov (1967, 1968). We note particularly that Zakharov (1968) derived the nonlinear Schrödinger (NLS) equation,

$$i\partial_t \psi + \alpha \partial_x^2 \psi + \beta \partial_y^2 \psi + \gamma |\psi|^2 \psi = 0, \tag{1.1}$$

where $\{\alpha, \beta, \gamma\}$ are real-valued constants, to describe approximately the slow evolution of the complex envelope of a nearly monochromatic train of plane waves of moderate amplitude in deep water. (Zakharov actually considered the one-dimensional version of (1.1), with $\beta = 0$.) Equation (1.1) is more general than the work of Benjamin & Feir (B–F) in the following sense: B–F showed that an infinitesimal perturbation of a uniform carrier wave must grow, while (1.1) predicts approximately both the initial instability and the subsequent evolution of the perturbed wavetrain.

Much of the original work on this subject focused on waves in deep water, but this kind of instability occurs in many physical systems. Necessary ingredients are that waves of infinitesimal amplitude should be dispersive (i.e. waves of different frequencies have different group velocities in the linearized limit), that these waves should have finite amplitude (but not too large), and that dissipation should be weak enough that it can be ignored at this order of approximation. Other physical systems in which the details have been worked out explicitly include optics (Ostrovsky 1967; Zakharov 1967; Anderson & Lisak 1984; Tai, Tomita & Hasegawa 1986; Hasegawa & Kodama 1995), and plasmas (Hasegawa 1971; McKinstrie & Bingham 1989).

The issue of what happens to a wavetrain once its perturbations begin to grow seems to be less clear than is commonly believed. Benjamin & Feir entitled their paper 'The disintegration of wavetrains in deep water', so their belief is clear. Benjamin (1967) shows two photographs of a train of waves of finite amplitude in deep water. One shows a wavetrain near the wavemaker, where the wavetrain is nearly uniform; the other shows the wavetrain further downstream, where it has largely disintegrated. However, Benjamin never actually applied the B–F theory to the photographed waves. To our knowledge, the most complete comparison of B–F theory with experiments was made by Lake *et al.* (1977) and Lake & Yuen (1977). They ran into some difficulties, which inspired further work by Crawford *et al.* (1981), Tulin & Waseda (1999), and Trulsen *et al.* (2000). We reconsider some of these comparisons later in this paper.

In experiments, Hammack & Henderson (2003) and Hammack, Henderson & Segur (2005) observed patterns of waves of moderate amplitude in deep water, some of which showed no apparent instability. Many of the experiments that we present in this paper show some growth of energy in sideband frequencies; none of our experiments show actual instability. This conflict between the conventional wisdom and our own experimental evidence led us to reconsider this entire issue.

2. Main results

Our main result is that dissipation plays a crucial role in a modulational instability like that of Benjamin & Feir, even when the dissipation is weak. In particular, it is known that with no dissipation, a uniform wavetrain is unstable to small perturbations of waves with nearly the same frequency and direction. These perturbations grow exponentially until their amplitudes become large enough for nonlinearity to impede their growth and to determine their subsequent evolution. Our result is that in the presence of dissipation, growth of the perturbations might be bounded by dissipation before nonlinearity comes into play. Further, dissipation reduces the set of unstable wavenumbers in time so that every wavenumber becomes stable eventually. In other

words, dissipation stabilizes the instability. (We define 'stable' precisely, below.) We prove stability for the spatially uniform solution of a dissipative version of (1.1), given below in (2.1). Additionally, we compare predictions of this model with experiments and find good agreement. We provide a measure of when this model accurately describes water waves and when it does not. We also compare our experimental results with a separate model that does not have dissipation, but uses higher-order terms in spectral bandwidth to bound the growth of perturbations. This model does not agree well with the experimental results. Below, we fill out the outline of this summary of our results.

To examine the effect of weak dissipation, we consider the following generalization of (1.1):

$$\mathrm{i}\partial_t\psi + \alpha\partial_x^2\psi + \beta\partial_y^2\psi + \gamma|\psi|^2\psi + \mathrm{i}\delta\psi = 0, \tag{2.1}$$

where $\{\alpha, \beta, \gamma, \delta\}$ are real-valued constants, and $\delta \geqslant 0$. Here, δ represents dissipation, and $\delta = 0$ reproduces (1.1). This approximate model has been used in water waves by Lake *et al.* (1977) and Mei & Hancock (2003); and in optics by Luther & McKinstrie (1990), Hasegawa & Kodama (1995), and Karlsson (1995). Miles (1967) reviewed and derived analytic formulae for δ, based on various kinds of physical dissipation that affect waves in deep water. In this paper, we consider δ to be an empirical parameter, which we measure experimentally as follows.

For definiteness, we seek solutions of either (1.1) or (2.1) on a rectangular domain D, with periodic conditions on the boundaries of D. Then (1.1) admits integral constants of the motion, including

$$M = \frac{1}{A_D}\iint_D |\psi(x, y, t)|^2 \,\mathrm{d}x\,\mathrm{d}y, \quad \boldsymbol{P} = \frac{\mathrm{i}}{A_D}\iint_D [\psi\nabla\psi^* - \psi^*\nabla\psi]\,\mathrm{d}x\,\mathrm{d}y, \tag{2.2}$$

where $()^*$ denotes complex conjugate, $\nabla\psi = (\partial_x\psi, \partial_y\psi)$, and A_D is the area of the domain. Sometimes M is called 'mass' or 'wave energy', and the two components of $\boldsymbol{P}$ are called 'linear momentum' (cf. Sulem & Sulem 1999). For (2.1) with $\delta > 0$, these quantities vary in time, but in a simple way:

$$M(t) = M(0)\exp(-2\delta t), \quad \boldsymbol{P}(t) = \boldsymbol{P}(0)\exp(-2\delta t). \tag{2.3}$$

We use these relations to validate (2.1) as a model of our experiments, by measuring $M(t)$ and $\boldsymbol{P}(t)$. For experiments in which these quantities decay exponentially in time, all with the *same* decay rate, we infer δ from the observed decay rate for that experiment and we use (2.1) to predict the detailed results of the experiment. As we discuss in § 6, some of our experiments involved either waves of large amplitude or waves with relatively large perturbations; in these experiments $\boldsymbol{P}(t)$ did not decay exponentially. For this reason, these experiments lay outside the range of validity of either (2.1) or (1.1).

In addition, waves of large amplitude often exhibited downshifting (Lake *et al.* 1977). To our knowledge, this effect also lies outside the range of (2.1) or (1.1). We plan to discuss the effects of large amplitude in a separate paper. We regard (2.1) as a reasonably accurate model of the evolution of nearly monochromatic waves of moderate amplitude.

The forms in (2.3) suggest a change of variables:

$$\psi(x, y, t) = \mu(x, y, t)\exp(-\delta t). \tag{2.4}$$

After this change, (2.1) becomes

$$\mathrm{i}\partial_t\mu + \alpha\partial_x^2\mu + \beta\partial_y^2\mu + \gamma\mathrm{e}^{-2\delta t}|\mu|^2\mu = 0. \tag{2.5}$$

For $t > 0$, (2.3) shows that $\psi \to 0$ in L_2-norm, but (2.4) factors out this overall decay. With periodic boundary conditions on D, (2.5) admits constants of the motion:

$$M_\mu = \frac{1}{A_D} \iint_D |\mu(x, y, t)|^2 \, dx \, dy = \text{const}, \tag{2.6a}$$

$$\boldsymbol{P}_\mu = \frac{i}{A_D} \iint_D [\mu \nabla \mu^* - \mu^* \nabla \mu] \, dx \, dy = \text{const}. \tag{2.6b}$$

In addition, (2.5) is a Hamiltonian system, with conjugate variables $\{\mu, \mu^*\}$ and Hamiltonian

$$H_\mu = i \iint_D \left[\alpha |\partial_x \mu|^2 + \beta |\partial_y \mu|^2 - \tfrac{1}{2} \gamma e^{-2\delta t} |\mu|^4\right] dx \, dy. \tag{2.6c}$$

This Hamiltonian is not a constant of the motion unless $\delta = 0$, but it is noteworthy that (2.5) is a Hamiltonian equation that describes a naturally occurring dissipative process.

Once the overall decay has been factored out, we can define 'stability'. As we discuss in § 3, a spatially uniform train of plane waves of moderate amplitude in deep water without dissipation is represented by a spatially constant solution of (1.1):

$$\psi_0(t) = A \exp\{i\gamma |A|^2 t\}, \tag{2.7}$$

where $\{A\}$ is a complex constant. In the dissipative situation, the corresponding solution of (2.1) is:

$$\psi_0(t) = A e^{-\delta t} \exp\left\{i\gamma |A|^2 \left(\frac{1 - e^{-2\delta t}}{2\delta}\right)\right\}, \tag{2.8a}$$

or for (2.5),

$$\mu_0(t) = A \exp\left\{i\gamma |A|^2 \left(\frac{1 - e^{-2\delta t}}{2\delta}\right)\right\}. \tag{2.8b}$$

We say that $\mu_0(t)$ is a stable solution of (2.5) if every solution of (2.5) that starts close to $\mu_0(t)$ at $t = 0$ remains close to it for all $t > 0$; otherwise, $\mu_0(t)$ is unstable. To make a precise definition of stability, define a perturbation $\sigma(x, y, t)$ by

$$\mu(x, y, t) = \mu_0(t) + \sigma(x, y, t), \tag{2.9}$$

and linearize (2.5) about $\mu_0(t)$, keeping only terms linear in σ. We say that μ_0 is *linearly stable* if for every $\varepsilon > 0$ there is a $\Delta > 0$ such that if

$$\iint_D |\sigma(x, y, 0)|^2 \, dx \, dy < \Delta \quad \text{at} \quad t = 0, \tag{2.10a}$$

then

$$\iint_D |\sigma(x, y, t)|^2 \, dx \, dy < \varepsilon \quad \text{for all} \quad t > 0, \tag{2.10b}$$

where $\sigma(x, y, t)$ satisfies the version of (2.5) that is linearized around $\mu_0(t)$. Nonlinear stability is similar, except that $\mu(x, y, t)$ evolves according to (2.5) instead of any linearization of it, and we need a more complicated norm. Precise details of the norms required are discussed in § 4 and Appendix B. For now, denote the norm in question by $\|\bullet\|$. We say that $\mu_0(t)$ is a *stable* (or nonlinearly stable) solution of (2.5) if for every $\varepsilon > 0$ there is a $\Delta > 0$ such that if

$$\|\mu(x, y, 0) - \mu_0(0)\|^2 < \Delta \qquad \text{at } t = 0, \tag{2.11a}$$

then

$$\|\mu(x, y, t) - \mu_0(t)\|^2 < \varepsilon \qquad \text{for all } t > 0, \tag{2.11b}$$

where $\mu(x, y, t)$ satisfies (2.5).

In the literature, we find many kinds of 'stability', all with slightly different meanings. The definitions in (2.10) and (2.11) are sometimes called 'stability in the sense of Lyapunov' (cf. Nemytskii & Stepanov 1960). This kind of stability is appropriate for a Hamiltonian system, like (2.5). Note that both (2.10) and (2.11) allow ε to be larger than Δ by a finite factor. This flexibility is essential for the results in this paper. Note also that this definition of stability expresses the familiar conceptual meaning of that term: any solution of (2.1) that starts close enough to $\mu_0(t)$ at $t = 0$ stays close to it for all $t \geqslant 0$, even after the overall decay of both solutions has been factored out. Thus, we can guarantee that the two solutions remain arbitrarily close to each other forever, by assuring that they start close enough to each other.

Here is the outline of this paper, and our main results. In § 3, we review briefly the derivation of (1.1) as an approximate model of the evolution of a nearly monochromatic wavetrain of moderate amplitude in deep water without dissipation, and we show that (2.7) is a linearly unstable solution of (1.1) for most choices of $\{\alpha, \beta, \gamma\}$. This is the Benjamin–Feir instability, as it appears in (1.1). The information in § 3 is not new, but it sets the stage for what follows.

In § 4 we prove two results.

THEOREM 1. *For any $\delta > 0$ and for any choice of $\{\alpha, \beta, \gamma\}$, $\mu_0(t)$ is a linearly stable solution of* (2.5)*, with periodic boundary conditions.*

Linear stability need not imply nonlinear stability, so we also prove

THEOREM 2. *For any $\delta > 0$ and for any choice of $\{\alpha, \beta, \gamma\}$, $\mu_0(t)$ is a nonlinearly stable solution of* (2.5)*, with periodic boundary conditions. Equivalently, $\psi_0(t)$ is a nonlinearly stable solution of* (2.1)*, with periodic boundary conditions.*

The proof of theorem 2 is straightforward in the one-dimensional problem ($\beta = 0$ or $\partial_y \equiv 0$ in (2.1) or (2.5)), and more involved in two dimensions. As a result, we prove the one-dimensional result in § 4, and place the proof for the two-dimensional case in Appendix B.

These two theorems establish our main theoretical result: *A uniform train of plane waves of finite amplitude is stable for any physical system modelled by* (2.1)*, including waves of moderate amplitude in deep water, with any $\delta > 0$. Dissipation, no matter how small, stabilizes the Benjamin–Feir instability in* (2.1) *or* (2.5).

The physical interpretation of this result is the following. With no dissipation, a uniform wavetrain (i.e. a Stokes wave) is unstable to small perturbations, and these perturbations can grow at an exponential rate until their amplitudes become large enough that nonlinear interactions impede further growth. This process might also occur in the presence of dissipation, but we will show that dissipation itself can quench the unstable growth well before nonlinear effects become relevant. If the perturbations are small enough initially (possibly very small), then they can grow by a finite amount and still be small, so nonlinear interactions never become important. In this way, dissipation controls the growth of small enough initial perturbations. What may be surprising is that dissipation can dominate nonlinear effects in this way for any non-zero dissipation, no matter how small, provided only that the initial perturbations are also small enough. Initial perturbation size is the only parameter that must be controlled.

Note that our assertion that dissipation controls the growth of small enough perturbations does not preclude the possibility that a large perturbation might lead to strong nonlinear interactions, in the same system. All definitions of 'stability' share the concept that small perturbations must remain small; 'stability' does not necessarily constrain large perturbations.

Even where dissipation is unimportant, (1.1) is known to be inaccurate for waves of large amplitude in deep water (see Lo & Mei 1985, for example). Dysthe (1979) and others have generalized (1.1) to improve this accuracy. Thus, Dysthe's model corrects (1.1) for one effect (waves with broader bandwidth), while (2.1) corrects (1.1) for another effect (dissipation). We will show that the stability of nearly monochromatic waves of moderate amplitude in deep water is controlled by the dissipative effect, not by the broader bandwidth effect.

In § 5, we obtain from (2.5) explicit predictions of growth and decay of various modes, in order to compare with the experimental results presented in § 6.

In § 6, we show that (2.1) accurately describes the evolution of nearly monochromatic waves of moderate amplitude in deep water, by comparing the predictions of (2.1) with laboratory experiments on waves propagating in deep water. Equation (2.1) predicts the results of these experiments, with good accuracy. In these experiments, the energy in sideband frequencies grows, but the total growth is bounded so it does not destroy the stability of the unperturbed solution. This establishes our main point: there is no modulational instability for nearly monochromatic waves of moderate amplitude in deep water.

Our experiments also demonstrate that in deep water, waves with small or moderate amplitude behave somewhat differently from waves with larger amplitudes. In § 6, we show that for waves of moderate amplitude, (2.1) predicts the experimental data significantly better than either (1.1) or Dysthe's (1979) generalization of (1.1). We also show that (2.1) breaks down as an accurate model for waves with large enough amplitudes in deep water, because $\boldsymbol{P}_\mu(t)$ is conserved by (2.1), but is not conserved in the experiments. We propose that an appropriate criterion for 'moderate amplitude' *vs.* 'large amplitude' is whether $\boldsymbol{P}_\mu(t)$ is conserved.

Finally, in § 7, we discuss briefly the existing literature, in which the results of Benjamin–Feir were compared with experiments.

The parameter δ plays a crucial role in our results, so we conducted additional experiments to learn how δ varies with experimental conditions. Appendix A describes these supplemental experiments. Appendix B contains the detailed proof of nonlinear stability of the Stokes wave in two dimensions.

3. Review of NLS as a model of waves in deep water without dissipation

The derivation of (1.1) as an approximate model of the evolution of nearly monochromatic waves of moderate amplitude in deep water without dissipation can be found in many places (e.g. Zakharov 1968; Ablowitz & Segur 1981), so we simply state results here. Let $\{X, Y\}$ represent coordinates on a horizontal plane, let T represent time, and let $\varepsilon > 0$ be a formal small parameter. For nearly monochromatic waves of small amplitude, we can represent the vertical displacement of the water's free surface in the form

$$\eta(X, Y, T; \varepsilon) = \varepsilon[\psi(x, y, t)\mathrm{e}^{\mathrm{i}\theta} + \psi^*\mathrm{e}^{-\mathrm{i}\theta}] + \varepsilon^2[\psi_1(x, y, t)\mathrm{e}^{\mathrm{i}\theta} + \psi_1^*\mathrm{e}^{-\mathrm{i}\theta}] + \varepsilon^2[\psi_2(x, y, t)\mathrm{e}^{2\mathrm{i}\theta} + \psi_2^*\mathrm{e}^{-2\mathrm{i}\theta}] + O(\varepsilon^3), \tag{3.1}$$

where,

$\theta = k_0 X - \omega(k_0) T$ represents the fast oscillation of a carrier wave,

$\omega(k)$ is the linearized dispersion relation, so $\omega^2 = gk + \tau k^3$ for inviscid waves in deep water, under the influence of both gravity (g) and surface tension (τ),

$x = \varepsilon\omega(k_0)(T - X/c_g)$ and $y = \varepsilon k_0 Y$ are coordinates to describe the slow modulation of the wave envelope in a coordinate system moving with speed $c_g = \mathrm{d}\omega/\mathrm{d}k$,

$t = \varepsilon^2 k_0 X$ is the time-like variable in which to observe the slow evolution of the envelope as it propagates down the tank.

For irrotational motion, the velocity potential has a corresponding expansion. We substitute these expansions into the governing equations for the motion of an incompressible, inviscid fluid with a free surface, under the influence of a constant gravitational field (g) and surface tension (τ), and solves order-by-order in ε. At $O(\varepsilon)$, the linearized dispersion relation appears. At $O(\varepsilon^2)$, we find that

$$\psi_2(x, y, t) = \left(\frac{1 + \tau k_0^2/g}{1 - 2\tau k_0^2/g}\right) k_0 [\psi(x, y, t)]^2. \tag{3.2}$$

For $\tau = 0$, this result is due to Stokes (1847). There is some flexibility in defining $\psi_1(x, y, t)$. We choose to set

$$\psi_1(x, y, t) = 0, \tag{3.3}$$

at the cost of making the corresponding term in the expansion for the velocity potential more complicated. The advantage of this choice is that we can identify $\psi(x, y, t)$ directly with measured values of the surface displacement, $\eta(X, Y, T; \varepsilon)$.

At $O(\varepsilon^3)$, we find that $\psi(x, y, t)$ approximately satisfies (1.1), with

$$\left.\begin{aligned} \alpha &= \frac{2k_0\omega''(k_0)}{\omega'(k_0)}\left(\frac{1 + \tau k_0^2/g}{1 + 3\tau k_0^2/g}\right)^2, \quad \beta = \tfrac{1}{2}, \\ \gamma &= -4k_0^2\left(\frac{1 + \tau k_0^2/g}{1 + 3\tau k_0^2/g}\right)\left\{1 + \frac{3}{4}\frac{\tau k_0^2/g}{1 - 2\tau k_0^2/g} - \frac{3}{8}\frac{\tau k_0^2/g}{1 + \tau k_0^2/g}\right\}. \end{aligned}\right\} \tag{3.4}$$

Comments

(i) These coefficients simplify without surface tension ($\tau = 0$), to

$$\alpha = -1, \quad \beta = \tfrac{1}{2}, \quad \gamma = -4k_0^2.$$

(ii) The nonlinear Schrödinger equation can be written in many ways. The coordinate system used here and the coefficients in (3.4) are appropriate for an experiment in which a time-periodic wave is generated at a fixed spatial location, and the wave evolves as it propagates down the tank.

(iii) The coefficients in (3.4) are valid for waves in deep water, under the influence of gravity and surface tension. Other combinations of coefficients occur in optics and in plasmas. We show below that for any choice of $\{\alpha, \beta, \gamma\}$, the B–F instability is stabilized in (2.1) for any $\delta > 0$.

A solution of (1.1) corresponding to a spatially uniform plane wave was given in (2.7). We call this a 'Stokes solution', because the phase information in (2.7), when used in (3.1), reproduces Stokes' (1847) correction to wave frequency for a wave of finite amplitude.

Zakharov (1968) calculated the linear stability of a Stokes solution of (1.1), for one-dimensional perturbations. For future reference, we now recall that calculation.

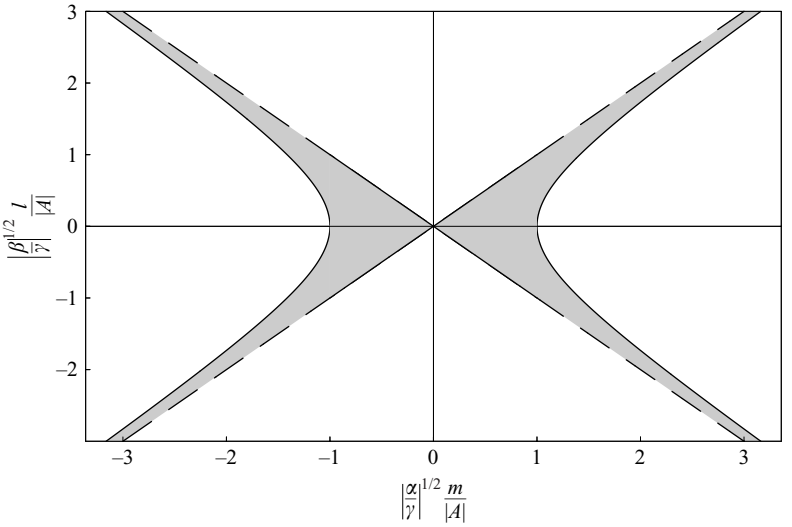

FIGURE 1. Stability space as defined by (3.6) in which $\delta = 0$. If $\alpha\beta < 0$, $\alpha\gamma > 0$, then any $\{m, l\}$ in the shaded region corresponds to a linearly unstable mode. If $\alpha\beta < 0, \alpha\gamma < 0$, then the unstable region has a different hyperbola as its boundary. If $\alpha\beta > 0, \alpha\gamma > 0$, then the unstable region lies within an ellipse. If $\alpha\beta > 0, \alpha\gamma < 0$, then there are no linearly unstable modes.

We linearize (1.1) around the solution in (2.7),

$$\psi(x, y, t) = \exp(\mathrm{i}\gamma|A|^2 t + \mathrm{i}\arg(A))\{|A| + \zeta u(x, y, t) + \mathrm{i}\zeta v(x, y, t) + O(\zeta^2)\} \quad (\zeta \ll 1), \tag{3.5}$$

where ζ is a formal small parameter, retaining only terms linear in ζ. The resulting linear differential equations for $\{u, v\}$ have constant coefficients, so we may seek solutions of the form

$$u(x, y, t) = U(m, l, \Omega)\mathrm{e}^{\mathrm{i}mx + \mathrm{i}ly + \Omega t} + (\text{c.c.}),$$
$$v(x, y, t) = V(m, l, \Omega)\mathrm{e}^{\mathrm{i}mx + \mathrm{i}ly + \Omega t} + (\text{c.c.}),$$

where (c.c.) denotes complex conjugate. The Stokes solution is linearly unstable if there is an allowable (m, l) for which $\mathrm{Re}\{\Omega(m, l)\} > 0$. We find that $\Omega(m, l)$ satisfies

$$\Omega^2 + (\alpha m^2 + \beta l^2)(\alpha m^2 + \beta l^2 - 2\gamma|A|^2) = 0. \tag{3.6}$$

If $\alpha\beta < 0$, then the first factor in the product in (3.6) vanishes on two straight lines in the (m, l) plane, while the second factor vanishes on a hyperbola with these straight lines as asymptotes (see figure 1). Any choice of (m, l) that lies between these two sets of curves corresponds to a mode that is linearly unstable. The maximum growth rate of these unstable modes is

$$\Omega_{max} = |\gamma|\,|A|^2, \tag{3.7a}$$

confirming that the growth rate of the instability is proportional to the square of the amplitude of the carrier wave. The maximal growth rate occurs along a curve in the (m, l) plane on which

$$\alpha m^2 + \beta l^2 = \gamma|A|^2. \tag{3.7b}$$

For perturbations that are purely one-dimensional, $l = 0$; no other results change.

4. Stability of a Stokes solution of (2.5)

The solution of (2.5) corresponding to a spatially uniform plane wave was given in (2.8). Its stability is the main theoretical result in this paper.

4.1. *Linearized stability*

Set

$$\mu(x,y,t) = \exp\left\{ \mathrm{i}\gamma|A|^2\left(\frac{1-\mathrm{e}^{-2\delta t}}{2\delta}\right) + \mathrm{i}\arg(A)\right\} \times \{|A| + \zeta u(x,y,t) + \mathrm{i}\zeta v(x,y,t) + O(\zeta^2)\}, \tag{4.1}$$

where ζ is a formal small parameter, substitute into (2.5), and retain only terms that are linear in ζ. The result is

$$\left.\begin{aligned} \partial_t v &= \alpha\partial_x^2 u + \beta\partial_y^2 u + 2\gamma\mathrm{e}^{-2\delta t}|A|^2 u, \\ -\partial_t u &= \alpha\partial_x^2 v + \beta\partial_y^2 v. \end{aligned}\right\} \tag{4.2}$$

According to (2.3), all square-integrable solutions of (2.1) decay to zero as $t\to\infty$. This overall decay has been factored out of (2.5), so the Stokes solution is unstable if the small perturbations (u, v) grow without bound relative to the constant $|A|$.

Without loss of generality, we seek solutions of (4.2) in the form

$$u(x,y,t) = U(t;m,l)\,\mathrm{e}^{\mathrm{i}mx+\mathrm{i}ly} + U^*\mathrm{e}^{-\mathrm{i}mx-\mathrm{i}ly},$$
$$v(x,y,t) = V(t;m,l)\,\mathrm{e}^{\mathrm{i}mx+\mathrm{i}ly} + V^*\mathrm{e}^{-\mathrm{i}mx-\mathrm{i}ly}.$$

Then (4.2) becomes

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -(\alpha m^2 + \beta l^2 - 2\gamma\mathrm{e}^{-2\delta t}|A|^2)U, \tag{4.3a}$$

$$\frac{\mathrm{d}U}{\mathrm{d}t} = (\alpha m^2 + \beta l^2)V, \tag{4.3b}$$

or

$$\frac{\mathrm{d}^2U}{\mathrm{d}t^2} + [(\alpha m^2 + \beta l^2)(\alpha m^2 + \beta l^2 - 2\gamma\mathrm{e}^{-2\delta t}|A|^2)]U = 0. \tag{4.3c}$$

Equation (4.3) is a Sturm–Liouville problem (cf. Ince 1956, chap. 10), so Sturmian theory implies the following results:

(i) if $[(\alpha m^2 + \beta l^2)(\alpha m^2 + \beta l^2 - 2\gamma\mathrm{e}^{-2\delta t}|A|^2)] > 0$, all solutions of (4.3) oscillate in time;

(ii) if $[(\alpha m^2+\beta l^2)(\alpha m^2+\beta l^2-2\gamma\mathrm{e}^{-2\delta t}|A|^2)] < 0$, then (4.3) admits a growing solution. As long as $[(\alpha m^2 + \beta l^2)(\alpha m^2 + \beta l^2 - 2\gamma\mathrm{e}^{-2\delta t}|A|^2)] \leqslant -C^2 < 0$, this solution grows at least as fast as $\mathrm{e}^{|C|t}$.

Here is the fundamental reason for the difference in stability between the appropriate Stokes solutions of (1.1) and (2.1). According to (4.3), at any fixed time, the set of (m, l) modes that can grow exponentially lies in a region like that shown in figure 1. However, the bounding hyperbola for (2.1) is defined by

$$\alpha m^2 + \beta l^2 - 2\gamma\mathrm{e}^{-2\delta t}|A|^2 = 0, \tag{4.4}$$

so the region of instability shrinks as t increases, and any specific (m, l) mode remains in this shrinking region only for a limited time. After that time, every solution of (4.3) for this (m, l) oscillates in time. Therefore, no (m, l) mode grows forever. This point was recognized by Luther & McKinstrie (1990), Hasegawa & Kodama (1995), Karlsson (1995) and Mei & Hancock (2003).

Using standard iterative methods from the theory of ordinary differential equations (cf. Ince 1956, chap. 3), we can show that for any (m, l), any $\delta > 0$ and any $t \geqslant 0$, the solution of (4.3) satisfies

$$\left.\begin{aligned} |U(t;m,l)| &\leqslant \sqrt{|U(0;m,l)|^2 + |V(0;m,l)|^2}\exp\left\{\frac{|\gamma||A|^2}{\delta}(1-\exp(-2\delta t))\right\}, \\ |V(t;m,l)| &\leqslant \sqrt{|U(0;m,l)|^2 + |V(0;m,l)|^2}\exp\left\{\frac{|\gamma||A|^2}{\delta}(1-\exp(-2\delta t))\right\}. \end{aligned}\right\} \tag{4.5}$$

Therefore, if $\delta > 0$ then for all $t \geqslant 0$, every (m, l) mode satisfies

$$|U(t)|^2 + |V(t)|^2 \leqslant 2[|U(0)|^2 + |V(0)|^2]\exp\left\{\frac{2|\gamma||A|^2}{\delta}(1-\exp(-2\delta t))\right\}. \tag{4.6}$$

Finally, we may solve (4.2) in a rectangular domain, D, in the (x, y)-plane, with periodic boundary conditions. We restrict our attention to initial conditions of (4.2) that satisfy

$$\iint_D [u^2(x, y, 0) + v^2(x, y, 0)]\,\mathrm{d}x\,\mathrm{d}y < \Delta, \tag{4.7a}$$

for some Δ to be determined. Using Parseval's relation (cf. Guenther & Lee 1988) and (4.6), it follows that for all $t \geqslant 0$,

$$\iint_D [u^2(x, y, t) + v^2(x, y, t)]\,\mathrm{d}x\,\mathrm{d}y < 2\Delta\exp\left\{\frac{2|\gamma||A|^2}{\delta}(1-\exp(-2\delta t))\right\}. \tag{4.7b}$$

However, $\{u, v\}$ are the real and imaginary parts of the linearized perturbation of the Stokes solution of (2.1). Therefore, for each $\varepsilon > 0$, if we choose

$$0 < \Delta \leqslant \tfrac{1}{2}\varepsilon\exp\left\{-\frac{2|\gamma|\,|A|^2}{\delta}\right\} \tag{4.8}$$

then it follows from (4.7) and (2.10) that the Stokes solution of (2.1) is linearly stable. This completes the proof of theorem 1.

Comments

(i) This proof applies for any choice of real constants $\{\alpha, \beta, \gamma\}$ in (2.1) or (2.5), provided $\delta > 0$. Figure 1 requires $\{\alpha\beta < 0, \alpha\gamma > 0\}$, but that restriction is not used in the proof.

(ii) A similar proof applies for square-integrable perturbations in the whole (x, y) plane, if we impose on (2.1) or (2.5) vanishing boundary conditions at infinity.

(iii) Stability of the Stokes solution of (2.1) or (2.5) does not mean that sideband modes gain no energy from the carrier wave. Instead, it means that a sideband mode can gain at most a finite amount of energy. Typically, sidebands grow for a limited time, but (4.5) and (4.6) provide bounds on their total growth. If $|\gamma|\,|A|^2/\delta \gg 1$, this growth can be substantial; but even with substantial growth, the Stokes solution of (2.5) is still linearly stable: (4.8) requires only that Δ be that much smaller.

(iv) Without dissipation, the Benjamin–Feir instability supplies energy from the carrier wave to the sidebands until the sidebands have grown so large that their own nonlinear interactions have become important. This process might still occur in the presence of dissipation, but the analysis presented here shows that dissipation provides a second option. Namely, a sideband stops growing once dissipation has removed enough energy from the carrier wave to quench the instability for that sideband, according to (4.3). Thus, if the sideband has a small enough initial amplitude, then it

stops growing before its own nonlinear interactions become important. If the amplitudes of all sidebands are small enough initially, then none of them grows very large, so the Stokes wave solution of (2.1) is linearly stable. Then nonlinear interactions among sidebands play no significant role. This is the physical significance of theorem 1.

4.2. *Nonlinear stability*

There are known problems in which linearized stability does not guarantee nonlinear stability (cf. Cherry 1926), so next we prove theorem 2, that the Stokes solution of (2.5) with periodic boundary conditions is nonlinearly stable. To do so, replace (4.1) with

$$\mu(x, y, t) = \exp\left\{\mathrm{i}\gamma|A|^2\left(\frac{1-\mathrm{e}^{-2\delta t}}{2\delta}\right)+\mathrm{i}\arg(A)\right\}[|A|+u(x, y, t)+\mathrm{i}v(x, y, t)],$$

substitute into (2.5), and retain all terms. Instead of (4.2), the result is

$$\partial_t v = \alpha\partial_x^2 u + \beta\partial_y^2 u + \dot{\phi}\left\{u + \frac{3u^2+v^2}{2|A|} + \frac{(u^2+v^2)u}{2|A|^2}\right\}, \tag{4.9a}$$

$$-\partial_t u = \alpha\partial_x^2 v + \beta\partial_y^2 v + \dot{\phi}\left\{\frac{uv}{|A|} + \frac{(u^2+v^2)v}{2|A|^2}\right\}, \tag{4.9b}$$

where

$$\dot{\phi} = 2\gamma\mathrm{e}^{-2\delta t}|A|^2. \tag{4.9c}$$

It follows from (4.9) that

$$\partial_t(u^2+v^2) = 2\alpha\partial_x(v\partial_x u - u\partial_x v) + 2\beta\partial_y(v\partial_y u - u\partial_y v) + 2\dot{\phi}uv + \dot{\phi}\left\{\frac{(u^2+v^2)v}{|A|}\right\}. \tag{4.10}$$

Define $\lambda(x, y, t) = u(x, y, t) + \mathrm{i}v(x, y, t)$, so $\partial_x\lambda(x, y, t) = \partial_x u(x, y, t) + \mathrm{i}\partial_x v(x, y, t)$, etc. Then denote

$$\|\lambda(t)\|_2^2 = \iint_D [u^2+v^2]\,\mathrm{d}x\,\mathrm{d}y, \quad \|\partial_x\lambda(t)\|_2^2 = \iint_D [(\partial_x u)^2 + (\partial_x v)^2]\,\mathrm{d}x\,\mathrm{d}y, \tag{4.11}$$

etc. We will also use

$$\|\partial_x\lambda(t)\|_4 = \left[\iint_D [(\partial_x u)^2 + (\partial_x v)^2]^2\,\mathrm{d}x\,\mathrm{d}y\right]^{1/4},$$

and

$$\|\lambda(t)\|_\infty = \sup_{(x, y)\in D}\{|u(x, y, t)|, |v(x, y, t)|\}.$$

If $\{u, v\}$ happen to be y-independent, then the y-integration in (4.11) is trivial.

It follows from (4.10) and the Cauchy–Schwarz inequality that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\lambda\|_2^2\right) \leqslant |\dot{\phi}|\|\lambda\|_2^2 + |\dot{\phi}|\iint_D\left[(u^2+v^2)\frac{|v|}{|A|}\right]\mathrm{d}x\,\mathrm{d}y. \tag{4.12}$$

The integral in (4.12) represents the contribution from the nonlinear terms in (4.9). Neglecting the integral leads to a slightly improved form of (4.7*b*). Retaining the integral in (4.12), we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\lambda\|_2^2\right) \leqslant |\dot{\phi}|\|\lambda\|_2^2\left[1 + \frac{\|\lambda\|_\infty}{|A|}\right]. \tag{4.13}$$

A similar calculation leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\partial_x\lambda\|_2^2\right) \leqslant |\dot{\phi}|\|\partial_x\lambda\|_2^2 + \frac{2}{|A|}|\dot{\phi}|\iint_D\left[|2u(\partial_x u\partial_x v) + ((\partial_x v)^2 - (\partial_x u)^2)v|\right]\mathrm{d}x\,\mathrm{d}y$$
$$+ \frac{2}{|A|^2}|\dot{\phi}|\iint_D\left[|(u^2 - v^2)(\partial_x u\partial_x v) + ((\partial_x v)^2 - (\partial_x u)^2)uv|\right]\mathrm{d}x\,\mathrm{d}y, \quad (4.14)$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\partial_x\lambda\|_2^2\right) \leqslant |\dot{\phi}|\|\partial_x\lambda\|_2^2\left\{1 + \frac{4\|\lambda\|_\infty}{|A|} + \frac{4\|\lambda\|_\infty^2}{|A|^2}\right\}. \quad (4.15)$$

In the one-dimensional problem (i.e. $\partial_y \equiv 0$), nonlinear stability follows from (4.13) and (4.15). To see this, define a Sobolev norm (Adams 1975):

$$\|\lambda(t)\|_{1,2}^2 = \|\lambda(t)\|_2^2 + \|\partial_x\lambda(t)\|_2^2 = \int_D [u^2 + v^2 + (\partial_x u)^2 + (\partial_x v)^2]\,\mathrm{d}x. \quad (4.16)$$

Then it follows from (4.13) and (4.15) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\lambda\|_{1,2}^2\right) \leqslant |\dot{\phi}|\|\lambda\|_{1,2}^2\left\{1 + \frac{4\|\lambda\|_\infty}{|A|} + \frac{4\|\lambda\|_\infty^2}{|A|^2}\right\}.$$

Hence, if there is a constant, N, such that $\|\lambda(t)\|_\infty \leqslant N < \infty$ for all $t \geqslant 0$, then by Gronwall's inequality (Coddington & Levinson 1955),

$$\|\lambda(t)\|_{1,2}^2 \leqslant \|\lambda(0)\|_{1,2}^2 \exp\left\{|\phi(t))|\left(1 + \frac{2N}{|A|}\right)^2\right\}, \quad (4.17a)$$

where

$$\phi(t) = \frac{\gamma|A|^2}{\delta}(1 - \mathrm{e}^{-2\delta t}). \quad (4.17b)$$

However, it is known (Adams 1975, p. 97) that in one spatial dimension, there is a constant C_1, independent of λ, such that

$$\|\lambda\|_\infty \leqslant C_1\|\lambda\|_{1,2}. \quad (4.18)$$

Then it follows from (4.17a) and (4.18) that for all $t \geqslant 0$,

$$\|\lambda(t)\|_{1,2}^2 < \varepsilon$$

provided

$$\|\lambda(0)\|_{1,2}^2 < \Delta \leqslant \varepsilon \exp\left\{-\frac{|\gamma||A|^2}{\delta}\left(1 + \frac{2C_1\sqrt{\varepsilon}}{|A|}\right)^2\right\}. \quad (4.19)$$

Therefore the Stokes solutions of (2.5) and of (2.1), both given in (2.8), are nonlinearly stable to small perturbations in one spatial dimension. This proves theorem 2 in one dimension. In two dimensions, the proof follows similar lines but is more complicated because (4.18) no longer holds. See Appendix B for that proof.

5. Predictions of experimental observations

In § 6, we describe experiments on nearly monochromatic waves of moderate amplitude in deep water. The experiments were conducted to test the dissipative theory presented here. A one-dimensional model is adequate to describe those experiments, so in this section we set $\partial_y \equiv 0$ in (2.1) and in (2.5), and derive predictions of the dissipative model, to be compared with experimental data in § 6. The results obtained

in this section are approximate, and they follow from (2.5). Alternatively, we can obtain predictions directly from (2.5) by integrating that equation numerically. We use both kinds of prediction in § 6.

In the experiments to be discussed, a spatially uniform one-dimensional wavetrain of moderate amplitude is perturbed by two wavetrains, each of small amplitude and each travelling in the same direction as the carrier wave. One of these perturbation wavetrains has waves with frequency slightly higher than that of the carrier wave, while the other has a frequency slightly lower. For our experiments, the two frequency differences are equal, so the initial data for (1.1) or (2.1) have approximately the form

$$\psi(x,0) = A + a_1 \,\mathrm{e}^{\mathrm{i}bx} + a_{-1}\,\mathrm{e}^{-\mathrm{i}bx}, \tag{5.1}$$

where $\{A, a_1, a_{-1}\}$ can be complex, with $|a_1| \ll |A|$, $|a_{-1}| \ll |A|$, and b represents the frequency difference between the carrier wave and either perturbation. The initial data are prescribed at one end of a test section, and the evolution occurs as the wave propagates along the test section. (Recall that in § 3 we defined x to represent clock-time, so b is a frequency in (5.1).)

5.1. *The spectrum of the one-dimensional wave pattern*

The data in (5.1) are periodic in x, so for either (1.1) or (2.1) with $\partial_y \equiv 0$, the solution that evolves from these data must be periodic in x as well. Necessarily it has the form

$$\psi(x,t) = \mathrm{e}^{-\delta t} \sum_{-\infty}^{\infty} a_n(t)\,\mathrm{e}^{\mathrm{i}nbx} \quad (\delta \geqslant 0), \tag{5.2}$$

so the only frequency differences generated by nonlinear interactions are integer multiples of b. This effect can be seen in a spectral representation of the data at any fixed 'time' (i.e. at any fixed location along the test section).

5.2. *A weakly nonlinear theory*

Solutions of (2.5) corresponding to (5.1) have the form $\mu(x,t) = \sum_{n=-\infty}^{\infty} a_n(t)\,\mathrm{e}^{\mathrm{i}nbx}$; $\mu(x,t)$ is complex, so $a_n(t)$ can be complex as well. In this representation, (2.6a, b) become

$$M_\mu = \sum_{n=-\infty}^{\infty} |a_n(t)|^2 = \text{const}, \tag{5.3a}$$

$$P_\mu = 2b \sum_{n=-\infty}^{\infty} n|a_n(t)|^2 = \text{const}. \tag{5.3b}$$

In addition, (2.5) reduces to an infinite set of coupled ODEs. For every integer n,

$$\mathrm{i}\dot{a}_n - \alpha(nb)^2 a_n + \gamma\,\mathrm{e}^{-2\delta t} \sum_{m,p=-\infty}^{\infty} \sum a_m a_p a^*_{m+p-n} = 0. \tag{5.4}$$

In our experiments, the initial data are rarely clean enough to satisfy (5.1) exactly. The theory we derive next applies under somewhat relaxed assumptions. Suppose

$$\left.\begin{aligned} &a_0(t) = O(\sqrt{M_\mu}), \quad a_1(t) \ll \sqrt{M_\mu}, \quad a_{-1}(t) = O(a_1), \\ \text{and for } |n| > 1, \quad &\frac{a_n(t)}{\sqrt{M_\mu}} = O\left(\left(\frac{a_1}{\sqrt{M_\mu}}\right)^{|n|}\right). \end{aligned}\right\} \tag{5.5}$$

Rewrite (5.4) to reflect this ordering. For $n=0$,

$$\begin{aligned}
&\mathrm{i}\dot{a}_0 + \gamma \mathrm{e}^{-2\delta t}\left[M_\mu a_0 + a_0 \sum_{p=1}^{\infty}\{|a_p|^2 + |a_{-p}|^2\} + 2a_0^* \sum_{p=1}^{\infty} a_p a_{-p}\right] \\
&\quad + \gamma \mathrm{e}^{-2\delta t} \sum \sum_{m,p=1}^{\infty} (a_m a_p a_{m+p}^* + a_{-m} a_{-p} a_{-m-p}^*) \\
&\quad + \gamma \mathrm{e}^{-2\delta t} \sum \sum_{m,p=1}^{\infty} 2(a_m^* a_{-p} a_{m+p} + a_{-m}^* a_p a_{-m-p}) = 0. \qquad (5.6a)
\end{aligned}$$

For $n=1$,

$$\begin{aligned}
&\mathrm{i}\dot{a}_1 - \alpha b^2 a_1 + \gamma \mathrm{e}^{-2\delta t}\left[2M_\mu a_1 - |a_1|^2 a_1 + a_0^2 a_{-1}^*\right] \\
&\quad + 2\gamma \mathrm{e}^{-2\delta t} \sum_{p=1}^{\infty} (a_0^* a_{-p} a_{p+1} + a_0 a_p^* a_{p+1} + a_0 a_{-p} a_{-p-1}^*) \\
&\quad + \gamma \mathrm{e}^{-2\delta t} \sum \sum_{m,p=1}^{\infty} (2a_{-m} a_p^* a_{m+p+1} + a_{-m} a_{-p} a_{-m-p-1}^*) \\
&\quad + \gamma \mathrm{e}^{-2\delta t}\left[2\sum_{p=1}^{\infty}\sum_{m=2}^{\infty} a_m a_{-p}^* a_{1-m-p} + \sum \sum_{m,p=2}^{\infty} a_m a_p a_{m+p-1}^*\right] = 0, \qquad (5.6b)
\end{aligned}$$

etc. No approximations to (2.5) have been made to this point.

For the remainder of this section, assume that (5.5) holds. Then it follows from (5.3*a*) that $M_\mu = |a_0|^2 + O((a_1)^2)$, and (5.6*a*) becomes

$$\mathrm{i}\dot{a}_0 + \gamma \mathrm{e}^{-2\delta t}[M_\mu a_0] = O((a_1)^2), \qquad (5.7)$$

with the solution

$$a_0(t) = A \exp\left\{\frac{\mathrm{i}\gamma M_\mu}{2\delta}(1 - \mathrm{e}^{-2\delta t})\right\} + O((a_1)^2), \qquad (5.8a)$$

where A is a complex-valued constant. If $a_n \equiv 0$ for $n \neq 0$, then $M_\mu = |a_0|^2$ and (5.8*a*) becomes (2.8*b*). In what follows, we make use of (4.17*b*) and (4.9*c*). Then (5.8*a*) becomes

$$a_0(t) = A\,\mathrm{e}^{\mathrm{i}\phi/2} + O((a_1)^2). \qquad (5.8b)$$

Thus, at leading order in this weakly nonlinear theory, $|a_0(t)|^2 = |A|^2 = \text{constant}$.

5.3. *Evolution of the excited sideband*

From (5.1), the excited sideband has the form $[a_1(t)\,\mathrm{e}^{\mathrm{i}bx} + a_{-1}(t)\,\mathrm{e}^{-\mathrm{i}bx}]$. For small $\{|a_1(t)|, |a_{-1}(t)|\}$, these amplitudes evolve according to

$$\begin{aligned}
\mathrm{i}\dot{a}_1 - \alpha b^2 a_1 + \dot{\phi}\left[a_1 + \tfrac{1}{2}\,\mathrm{e}^{\mathrm{i}\phi + 2\mathrm{i}\arg\{A\}} a_{-1}^*\right] &= O((a_1)^3), \\
\mathrm{i}\dot{a}_{-1} - \alpha b^2 a_{-1} + \dot{\phi}\left[a_{-1} + \tfrac{1}{2}\,\mathrm{e}^{\mathrm{i}\phi + 2\mathrm{i}\arg\{A\}} a_1^*\right] &= O((a_1)^3).
\end{aligned} \qquad (5.9)$$

Define

$$c_1(t) = a_1(t) + a_{-1}(t), \quad s_1(t) = a_1(t) - a_{-1}(t),$$

so that

$$[a_1\,\mathrm{e}^{\mathrm{i}bx} + a_{-1}\,\mathrm{e}^{-\mathrm{i}bx}] = c_1 \cos(bx) + \mathrm{i}s_1 \sin(bx).$$

Thus $c_1(t)$ represents the part of this sideband that is even in x, while $s_1(t)$ represents the part that is odd in x. It follows from (5.9) that to $O((a_1)^3)$,

$$\left.\begin{aligned} \mathrm{i}\dot{c}_1 - \alpha b^2 c_1 + \dot{\phi}\left[c_1 + \tfrac{1}{2}\,\mathrm{e}^{\mathrm{i}\phi + 2\mathrm{i}\arg\{A\}} c_1^*\right] = 0, \\ \mathrm{i}\dot{s}_1 - \alpha b^2 s_1 + \dot{\phi}\left[s_1 - \tfrac{1}{2}\,\mathrm{e}^{\mathrm{i}\phi + 2\mathrm{i}\arg\{A\}} s_1^*\right] = 0. \end{aligned}\right\} \tag{5.10}$$

Both $c_1(t)$ and $s_1(t)$ are complex-valued, so set $c_1(t) = [u_1(t) + \mathrm{i}v_1(t)]\,\mathrm{e}^{\mathrm{i}\phi(t)/2 + \mathrm{i}\arg\{A\}}$, and observe that $\{u_1(t),\ v_1(t)\}$ satisfy (4.3a, b) with $\{\beta l^2 = 0,\ m = b\}$ in those equations. Similarly, set $\mathrm{i}s_1(t) = [U_1(t) + \mathrm{i}V_1(t)]\,\mathrm{e}^{\mathrm{i}\phi(t)/2 + \mathrm{i}\arg\{A\}}$ and observe that $\{U_1(t),\ V_1(t)\}$ also satisfy (4.3a, b) with $\{\beta l^2 = 0,\ m = b\}$. It follows from these observations and (4.6) that

$$|a_1(t)|^2 \leqslant \tfrac{1}{2}[|c_1(0)|^2 + |s_1(0)|^2]\,\mathrm{e}^{2|\phi(t)|}, \quad |a_{-1}(t)|^2 \leqslant \tfrac{1}{2}[|c_1(0)|^2 + |s_1(0)|^2]\,\mathrm{e}^{2|\phi(t)|},$$

for as long as the linearized equations remain valid approximations. More precise information about $a_1(t)$ and $a_{-1}(t)$ can be obtained by integrating (4.3) directly, as we do in § 6.

Note that the symmetric (c_1) and antisymmetric (s_1) parts of this sideband can each grow for a while, but then must oscillate in time. If the initial seeded perturbation happens to be even in x at $t = 0$ (so $a_{-1}(0) = a_1(0)$), then it remains even as it evolves. If it has any initial asymmetry, then this asymmetry can also grow, but then must oscillate.

Specifically, each of (1.1), (2.1) and (2.5) is invariant under $\{x \to -x\}$. Even so, if the initial data for these equations are asymmetric in x, then that asymmetry will typically grow during the growth phase of the evolution. We show in § 6 that this kind of asymmetric growth commonly occurs in practice, and that (2.5) describes the asymmetric growth accurately, even though (2.5) is symmetric $\{x \to -x\}$.

5.4. *Evolution of other sidebands*

Nonlinear interactions of the first sideband with itself and with the carrier wave generate the second sideband, $[a_2(t)\,\mathrm{e}^{2\mathrm{i}bx} + a_{-2}(t)\,\mathrm{e}^{-2\mathrm{i}bx}]$. If (5.5) holds, then the amplitudes of this sideband satisfy

$$\left.\begin{aligned} \mathrm{i}\dot{a}_2 - 4\alpha b^2 a_2 + \dot{\phi}\left[a_2 + \tfrac{1}{2}\,\mathrm{e}^{\mathrm{i}\phi + 2\mathrm{i}\arg\{A\}} a_{-2}^*\right] + \gamma \mathrm{e}^{-2\delta t}\left[2a_0 a_1 a_{-1}^* + a_0^* a_1^2\right] = O((a_1)^4), \\ \mathrm{i}\dot{a}_{-2} - 4\alpha b^2 a_{-2} + \dot{\phi}\left[a_{-2} + \tfrac{1}{2}\,\mathrm{e}^{\mathrm{i}\phi + 2\mathrm{i}\arg\{A\}} a_2^*\right] + \gamma\,\mathrm{e}^{-2\delta t}\left[2a_0 a_{-1} a_1^* + a_0^* a_{-1}^2\right] = O((a_1)^4). \end{aligned}\right\} \tag{5.11}$$

As with (5.9), define

$$c_2(t) = a_2(t) + a_{-2}(t), \quad s_2(t) = a_2(t) - a_{-2}(t),$$

and set

$$c_2 = [u_2 + \mathrm{i}v_2]\,\mathrm{e}^{\mathrm{i}\phi/2 + \mathrm{i}\arg\{A\}}, \quad \mathrm{i}s_2 = [U_2 + \mathrm{i}V_2]\,\mathrm{e}^{\mathrm{i}\phi/2 + \mathrm{i}\arg\{A\}}.$$

Then (5.11) provides two pairs of real-valued equations:

$$\dot{u}_2 - 4\alpha b^2 v_2 + \frac{\dot{\phi}}{2|A|}(u_1 v_1 - U_1 V_1) = 0, \tag{5.12a}$$

$$-\dot{v}_2 - 4\alpha b^2 u_2 + \dot{\phi} u_2 + \frac{\dot{\phi}}{4|A|}\left(3u_1^2 + v_1^2 - 3U_1^2 - V_1^2\right) = 0, \tag{5.12b}$$

and

$$-\dot{V}_2 - 4\alpha b^2 U_2 + \dot{\phi} U_2 + \frac{\dot{\phi}}{2|A|}(3u_1 U_1 + v_1 V_1) = 0, \tag{5.12c}$$

$$\dot{U}_2 - 4\alpha b^2 V_2 + \frac{\dot{\phi}}{2|A|}(u_1 V_1 + v_1 U_1) = 0. \tag{5.12d}$$

With $\{\phi(t), u_1(t), v_1(t), U_1(t), V_1(t)\}$ known (from having integrated (4.3) twice), (5.12) contains forced linear equations that can be solved explicitly. It follows from (5.12) that even if $a_2(0)=0=a_{-2}(0)$ initially, these amplitudes grow until $|a_2(t)|=O(|a_1(t)|^2/\sqrt{M_\mu})$ and $|a_{-2}(t)|=O(|a_1(t)|^2/\sqrt{M_\mu})$, as asserted in (5.5). In fact, we can show that

$$|c_2(t)| \leqslant \frac{2}{|A|}[|c_1(0)|^2+|s_1(0)|^2]\,\mathrm{e}^{|\phi|}\big(1-\mathrm{e}^{-|\phi|/2}\big)+|c_2(0)|\,\mathrm{e}^{|\phi|/2},$$

$$|s_2(t)| \leqslant \frac{2}{|A|}[|c_1(0)|^2+|s_1(0)|^2]\,\mathrm{e}^{|\phi|}\big(1-\mathrm{e}^{-|\phi|/2}\big)+|s_2(0)|\,\mathrm{e}^{|\phi|/2}.$$

More precise information about $\{a_2(t), a_{-2}(t)\}$ can be obtained by integrating (5.12) directly.

We can continue in this way to obtain weakly nonlinear approximations for all of the amplitudes identified in (5.2). The next set of sidebands satisfies

$$\mathrm{i}\dot a_3 - 9\alpha b^2 a_3 + \dot\phi\big[a_3+\tfrac12\,\mathrm{e}^{\mathrm{i}\phi+2\mathrm{i}\arg(A)}a_{-3}^*\big] + 2\gamma\mathrm{e}^{-2\delta t}\big[a_0a_1a_{-2}^*+a_0^*a_1a_2+a_0a_{-1}^*a_2+\tfrac12 a_1^2a_{-1}^*\big] = O((a_1)^5), \quad (5.13a)$$

and

$$\mathrm{i}\dot a_{-3} - 9\alpha b^2 a_{-3} + \dot\phi\big[a_{-3}+\tfrac12\,\mathrm{e}^{\mathrm{i}\phi+2\mathrm{i}\arg(A)}a_3^*\big] + 2\gamma\mathrm{e}^{-2\delta t}\big[a_0a_{-1}a_2^*+a_0^*a_{-1}a_{-2}+a_0a_1^*a_{-2}+\tfrac12 a_{-1}^2a_1^*\big] = O((a_1)^5). \quad (5.13b)$$

We omit calculations for higher modes, because amplitudes smaller than $\mathrm{O}(a_3(t))$ were too small to measure accurately in our experiments.

5.5. *Higher-order evolution of the carrier wave*

According to (5.8), $|a_0(t)|$ is constant at leading order. However, from (5.3a), the other modes cannot grow significantly unless $a_0(t)$ loses energy to these modes. This transfer of energy from the carrier wave to the sidebands occurs at $O((a_1)^2)$, as we show next. A better approximation to (5.6a) than (5.7) is

$$\mathrm{i}\dot a_0 + \gamma\mathrm{e}^{-2\delta t}[M_\mu a_0 + a_0(|a_1|^2+|a_{-1}|^2)+2a_0^*a_1a_{-1})] = O((a_1)^4).$$

Represent $a_1(t)$ and $a_{-1}(t)$ as before, but now set $a_0(t)=\{|A|+u_0(t)+\mathrm{i}v_0(t)\}\times\mathrm{e}^{\mathrm{i}\phi(t)/2+\mathrm{i}\arg(\{A\}}$, where we expect $(u_0, v_0)=O((a_1)^2)$. The result is:

$$\dot u_0 + \frac{\dot\phi}{2\sqrt{M_\mu}}(u_1v_1+U_1V_1)=0, \quad -\dot v_0+\frac{\dot\phi}{2\sqrt{M_\mu}}\big(u_1^2+U_1^2\big)=0. \quad (5.14)$$

Comparing (5.14a) with (5.10), we can show that

$$u_0(t)=u_0(0)+\frac{1}{4\sqrt{M_\mu}}[|c_1(0)|^2+|s_1(0)|^2-|c_1(t)|^2-|s_1(t)|^2],$$

$$v_0(t)=v_0(0)+\frac{1}{2\sqrt{M_\mu}}\int_0^t\big(u_1^2+U_1^2\big)\dot\phi\,\mathrm{d}\tau. \quad (5.15)$$

We can choose A so that $u_0(0)=0=v_0(0)$. Then

$$|a_0(t)|^2=|A|^2+\tfrac12(|c_1(0)|^2-|c_1(t)|^2)+\tfrac12(|s_1(0)|^2-|s_1(t)|^2)+O((a_1)^4).$$

Alternatively, we can solve (5.3a) for $|a_0(t)|^2$ to obtain

$$|a_0(t)|^2=M_\mu-\left[|a_1(t)|^2+|a_{-1}(t)|^2+|a_2(t)|^2+|a_{-2}(t)|^2)+O\left(\frac{a_1^6}{M_\mu^2}\right)\right]. \quad (5.16)$$

5.6. *Evolution of the second harmonic of the carrier wave*

Equations (2.1) and (1.1) each describe approximately the evolution of a nearly monochromatic wavetrain of moderate amplitude, with and without dissipation, respectively. Each model is designed to be accurate for waves with frequency near that of the carrier wave. The second harmonic of the carrier wave has a frequency twice that of the carrier wave, so it lies outside the region of validity of either model. Even so, the derivation of either model from the original physical problem implies a necessary relation between the amplitude of the harmonic and that of the carrier wave. For waves in deep water, this relation is given in (3.2). In this way, the evolution of the second harmonic is controlled by the evolution of the carrier wave.

In the problem under consideration, at leading order

$$\psi(x,t) = \mathrm{e}^{-\delta t}\left[A\mathrm{e}^{\mathrm{i}\phi/2} + o(1)\right].$$

When combined with (3.2), this implies the dominant behaviour of the second harmonic:

$$\|\psi_2(x,t)\| \sim \mathrm{e}^{-2\delta t}\mathrm{const.} \tag{5.17}$$

Thus, (2.1) predicts that as the carrier wave decays, its second harmonic also decays, with a decay rate twice that of the carrier wave.

In fact, (3.2) provides more information about this harmonic. Once the amplitude of the carrier wave has been measured, then (3.2) predicts the amplitude of the second harmonic, with no adjustable parameters. Lake & Yuen (1977) emphasized this point in comparing their experiments with the B–F theory. To see this, represent both sides of (3.2) in Fourier series in x, and single out the mean term of each side. The result is

$$a_h(t) = \left(\frac{1+\tau k_0^2/g}{1-2\tau k_0^2/g}\right) k_0\left[a_0^2(t) + 2a_1(t)a_{-1}(t) + O\left(\frac{a_1^4}{M_\mu}\right)\right], \tag{5.18}$$

where $a_h(t)$ is the complex amplitude of the second harmonic. This specifies $a_h(t)$ completely, to this order of approximation.

5.7. *The most unstable wavenumber*

In the absence of dissipation, the growth rate of each unstable wavenumber is given by (3.5), so we can find a most unstable wavenumber and a maximal growth rate, as shown in (3.6). The situation changes if $\delta > 0$, because the most unstable wavenumber at one instant of time ceases to be the most unstable at a later instant. According to (4.3), all modes stop growing eventually, so we can ask for the wavenumber with the maximal total growth. Even this can be ambiguous for an experiment of limited duration, because the maximal total growth can depend on the duration of the experiment (i.e. the length of the test section). Generally, the concept of a most unstable wavenumber is unambiguous and easily testable in an experiment without dissipation, but less well-defined in an experiment with dissipation.

5.8. *Frequency downshifting*

In their experimental study of waves in deep water, Lake *et al.* (1977) discovered frequency downshifting, a process in which the frequency of the carrier wave slowly shifts to lower values as the waveform evolves. Afterwards, frequency downshifting was also observed and explained in optics (Gordon 1986). To our knowledge, downshifting cannot be explained by either (1.1) or (2.1), with or without (5.5). These models are too simple to explain this effect. We intend to report on downshifting in a subsequent paper.

6. Experimental tests of the dissipative theory

Our main point, that any amount of dissipation (of the correct kind) stabilizes the Benjamin–Feir instability for waves of moderate amplitude, contradicts well-established beliefs with regard to water waves, to optics and to plasmas. Therefore, we conducted laboratory experiments on waves in deep water, to see how well (2.1) describes the evolution of nearly monochromatic waves of moderate amplitude, as these waves propagate down a long wave tank. In this section, we describe both the experiments and their comparisons with the predictions of (2.1), given in § 5. We also compare the experimental results with predictions from two non-dissipative models, (1.1) and Dysthe's (1979) equation.

6.1. *Experimental apparatus and procedures*

The experimental apparatus comprised a wave channel, wavemaker, wave gauges, computer systems and water supply.

The wave channel is 43 ft long and 10 in wide with glass bottom and sidewalls. At one end is a sanded-glass beach to minimize reflections. Its toe is at 32.8 ft from the wavemaker. At the other end is a plunger-type wavemaker that oscillates vertically and has an exponential (E) cross-section that spans the width of the tank. The idea behind the E-plunger is to mimic more closely the (linearized) velocity field of a deep-water wave, which falls off exponentially in the vertical. The E-plunger used herein falls off with the wavenumber of a 3.33 Hz wave, the carrier wave used in these experiments. The wavemaker is driven by a motor under digital control using dual closed-loop feedbacks (position and velocity) to follow faithfully the command signals of position, velocity and time. This is the most precise linear motion technology available today.

For the experiments in this section η_w, the displacement of the wavemaker in the undisturbed water level, is prescribed to be a modulated sine wave

$$\eta_w(T) = 2\tilde{a}_0 \sin(\omega_0 T)[1 + 2r \sin(\omega_p T + \varphi)], \tag{6.1}$$

where the subscripts '0' and 'p' indicate carrier and perturbation modes, $r = \tilde{a}_p/\tilde{a}_0$ is the ratio of the perturbation amplitude to the carrier wave amplitude, and $\{\tilde{a}_n\}$ denotes a real-valued measured wave amplitude. We note that with the definition of Fourier amplitudes implied by (5.2), the crest-to-trough amplitude of an oscillatory wave is $2|\tilde{a}_n|$. The vertical velocity of the wavemaker is prescribed as the time-derivative of η_w. In this way, we seeded a particular perturbation mode in each experiment. If we did not seed modes for carrier waves of moderate amplitude, then the perturbation did not grow enough in our test section to observe measurable growth. If we did not seed modes for carrier waves of large amplitude, then perturbations grew measurably, but M_μ and/or $\boldsymbol{P}_\mu$ was not conserved.

To compare measurements with predictions from the theory, note that (6.1) can be written as

$$\begin{aligned}\eta_w(T) = {} & \mathrm{e}^{\mathrm{i}\omega_0 T}[-\mathrm{i}\tilde{a}_0 - \tilde{a}_p \mathrm{e}^{\mathrm{i}\omega_p T + \mathrm{i}\varphi} + \tilde{a}_p \mathrm{e}^{-\mathrm{i}\omega_p T - \mathrm{i}\varphi}] \\ & + \mathrm{e}^{-\mathrm{i}\omega_0 T}[\mathrm{i}\tilde{a}_0 - \tilde{a}_p \mathrm{e}^{-\mathrm{i}\omega_p T - \mathrm{i}\varphi} + \tilde{a}_p \mathrm{e}^{\mathrm{i}\omega_p T + \mathrm{i}\varphi}].\end{aligned}$$

This is consistent with the notation in (3.1) and (5.1) if we identify

$$\varepsilon b = \frac{\omega_p}{\omega_o}, \tag{6.2a}$$

$$\varepsilon\psi(x, 0) = [-\mathrm{i}\tilde{a}_0 - \tilde{a}_p \mathrm{e}^{\mathrm{i}bx + \mathrm{i}\varphi} + \tilde{a}_p \mathrm{e}^{-\mathrm{i}bx - \mathrm{i}\varphi}]. \tag{6.2b}$$

The wavegauge is a non-intrusive capacitance-type gauge that spans 12.7 cm of the width of the tank and 6 mm in the direction of wave propagation. It was calibrated dynamically by oscillating it at the carrier wave frequency and it has a remarkably linear calibration that remains constant over a period of days. Time series of water surface displacement were recorded at sampling rates of either 300 or 350 Hz for 25 s using LabView software and National Instruments hardware.

In all of the experiments discussed in this section, the wavemaker was oscillated with the motion (6.1). One set of experiments consisted of N experiments – typically, $N = 12$ or 14. In the nth experiment of the set ($n = 1, 2, \ldots, N$), the wavegauge was fixed at a distance of $X_n = 128 + 50(n - 1)$ cm from the wavemaker. A time series of water surface displacement at that point was measured. The $n = 1$ location was chosen to be out of the regime of evanescent waves near the wavemaker. We computed the Fourier transform of the nth time series to obtain the complex Fourier amplitudes with a frequency resolution of ± 0.27 rad s^{-1} or ± 0.04 Hz. From the Fourier transforms of each of the n time series, we evaluated: (i) M_μ from (5.3a); (ii) P_μ from (5.3b); (iii) a_0, the Fourier amplitude of the carrier wave at frequency ω_0; (iv) a_h, the Fourier amplitude of its harmonic at frequency $2\omega_0$; (v) $\{a_1, a_{-1}\}$, the Fourier amplitudes of the two seeded sidebands, at frequencies $\omega_0 \pm \omega_p$; (vi) $\{a_2, a_{-2}\}$, the Fourier amplitudes of the second sidebands, at frequencies $\omega_0 \pm 2\omega_p$; and (vi) $\{a_3, a_{-3}\}$, the Fourier amplitudes of the third sidebands, at frequencies $\omega_0 \pm 3\omega_p$.

Note that each of these Fourier amplitudes is complex-valued. Both the magnitude and phase of each Fourier amplitude, measured at the $n = 1$ location, are required as initial data for the predictive differential equations like (4.3) and (5.12). In the comparisons that follow, we plot only magnitudes of Fourier amplitudes, like $|a_0|$, but both magnitude and phase are required for a complete theory.

The validity of the exponential decay model is crucial for the validity of (2.1); thus, we also conducted experiments using unseeded wavetrains that varied wave amplitude and water surface age, to determine how the waves decay. These experiments are discussed in Appendix A. Our procedure for the experiments in this section was to clean the tank with alcohol before filling it to a depth of $h > 20$ cm. A brass rod that spans the width of the tank and is mounted on a carriage above the tank was skimmed along the surface, scraping the surface film to the end of the tank. There the film was vacuumed with a wet vac. The wet vac was used further to vacuum the water surface throughout the tank until the depth was $h = 20$ cm as measured by a point gauge mounted above the tank.

The experiments in Appendix A show that the waves decayed approximately exponentially and that the decay rate varied with surface age. However, the decay rate was essentially constant during the first two hours after cleaning the water surface. Thus, before each set of N experiments, the surface was cleaned as described above. The tank was allowed to settle for 10 min between the n and $n+1$ experiments, and the set of N experiments was conducted within about a 2 h period after cleaning the surface.

We obtained the value for $\tilde{\delta}$, the measured spatial decay rate, in three ways and compared the results: (i) we found the L_2 norm of the nth time series by numerically integrating the data, and then fit an exponential through those results; (ii) we computed values for M_μ of the nth time series from (5.3a) by summing the Fourier coefficients for all of the modes resolved in the nth time series, and then fit an exponential through those results; and (iii) we computed values for M_μ of the nth time series from (5.3a) by summing the Fourier coefficients for a band of modes around the carrier, and then fit an exponential through those results. In all cases,

$\omega_0/2\pi$ (Hz)	$\omega_p/2\pi$ (Hz)	$2a_0(0)$ (cm)	$(\lvert a_{-1}\rvert + \lvert a_1\rvert)/(2\lvert a_0\rvert)^{-1}$
3.33	0.17	−0.116 + 0.183 i	0.11
$2u_1$ (0) (cm)	$2v_1(0)$ (cm)	$2U_1$ (0) (cm)	$2V_1$ (0) (cm)
−0.032	0.051	−0.009	0.007
$2u_2(0)$ (cm)	$2v_2(0)$ (cm)	$2U_2(0)$ (cm)	$2V_2(0)$ (cm)
0.000	−0.006	0.002	−0.009
$\mathrm{Re}(2a_3(0))$	$\mathrm{Im}(2a_3(0))$	$\mathrm{Re}(2a_{-3}(0))$	$\mathrm{Im}(2a_{-3}(0))$
0.001	0.001	−0.001	0.000

TABLE 1. Experimental parameters measured at the first measurement site for the frequency of the carrier wave, the perturbation frequency, the complex amplitude of the carrier wave, the ratio of amplitudes of the first set of sidebands to that of the carrier wave, and the initial conditions used in computations of (4.3) for the first set of sidebands, of (5.12) for the second set, and of (5.13) for the third set. For this set of experiments, $\tilde{\delta} = 0.110\,\mathrm{m}^{-1}$.

the spatial decay rate was determined in terms of distance down the wavetank. For the first set of experiments we describe in this paper (figures 2–8), method (i) gave $\tilde{\delta} = 0.115\,\mathrm{m}^{-1}$, while methods (ii) and (iii) gave $\tilde{\delta} = 0.110\,\mathrm{m}^{-1}$. For comparison with theory we used the latter value.

The dimensionless δ used in the theory is related to $\tilde{\delta}$, the measured spatial decay rate, by

$$\tilde{\delta} = \varepsilon^2 k_0 \delta, \tag{6.3}$$

where k_0 is the wavenumber of the carrier wave, and ε is the small expansion parameter. We used $\varepsilon = 2k_0|a_0(0)|$.

6.2. *Our fundamental comparison with experiments*

In this subsection we compare the theory presented in § 5 with data from a typical set of experiments, with parameters given in table 1. These experiments had a measure of nonlinearity of $\varepsilon = 2k_0|a_0(0)| = 0.10$ ($k_0 = 0.441$ for the 3.33 Hz wave used herein). Additional experiments were conducted that varied carrier wave amplitude or initial perturbation wave amplitude with similar results, except where $P_\mu(t)$ was not conserved.

Figures 2 and 3 show in column 1 measured water surface displacement (in cm) of the slowly modulated wavetrain described by (6.1) and table 1 at the $n = 1, 2, \ldots,$ 12 different measurement sites. Comparing data from the different measurement sites shows the slow evolution of the modulations, along with the overall decay of all wave amplitudes. In column 2 are the modal amplitudes (in cm) obtained from the corresponding Fourier transforms. The prominent features at X_1 are:

(i) the carrier wave at 3.33 Hz;
(ii) two sidebands resulting from the seeded perturbation at 3.16 and 3.50 Hz;
(iii) the second harmonics near 6.7 Hz;
(iv) the third harmonics near 10 Hz.

As this modulated wavetrain propagated downstream, the amplitudes at these frequencies changed slowly. In addition, other sidebands with frequencies near that of the carrier wave also grew, as predicted by either the non-dissipative model in (1.1) or the dissipative model in (2.1). By X_3, there was observable energy in higher-frequency sidebands at about 3.67, 3.84 and 4.02 Hz, and in a lower-frequency sideband at 2.99 Hz. The time series show that the wavetrain evolved due both to an overall decay

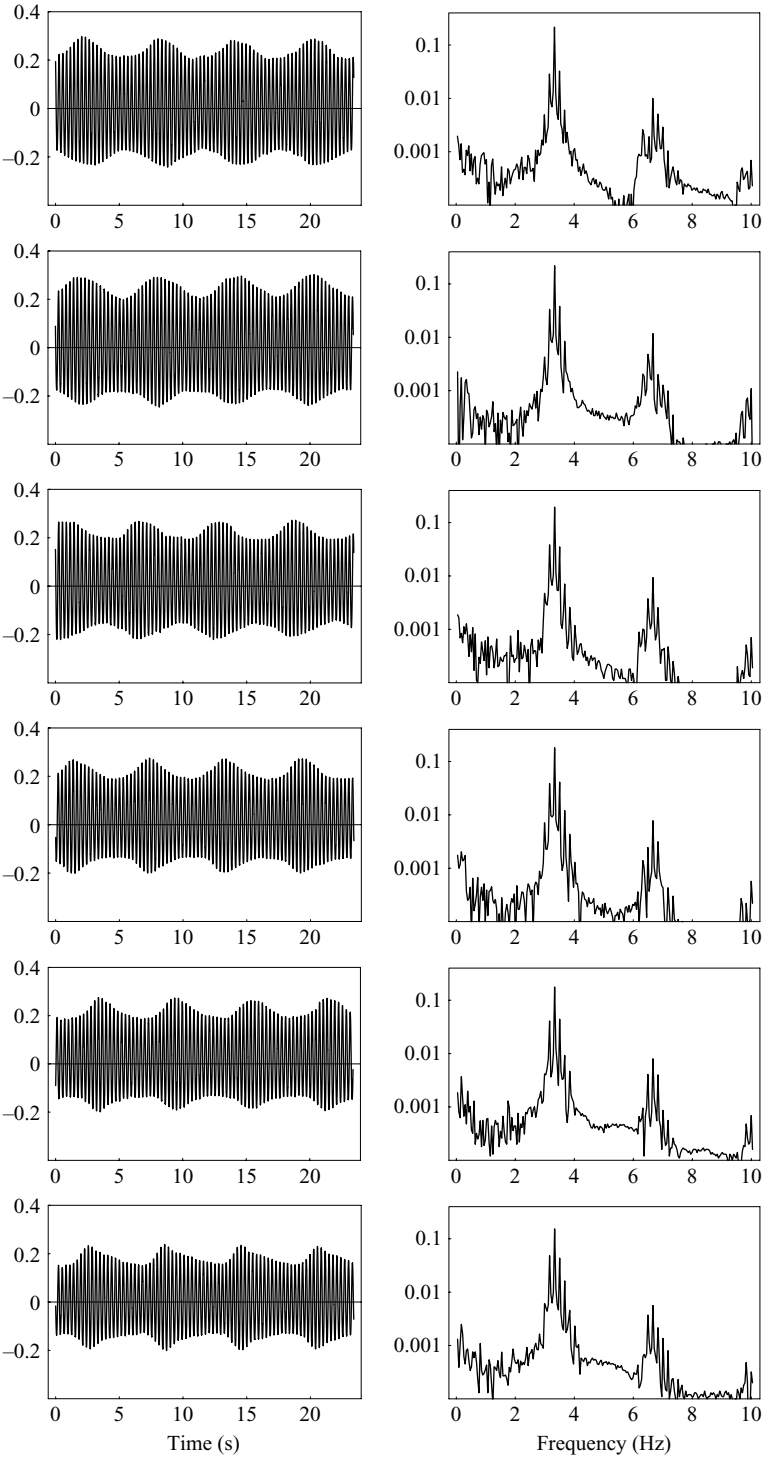

FIGURE 2. Water surface displacement (cm) as a function of time, and corresponding Fourier coefficients (cm) obtained from 6 experiments when the wave gauge was fixed at $128 + 50(n-1)$ cm from the wavemaker, for $(n = 1, 2, 3, 4, 5, 6)$.

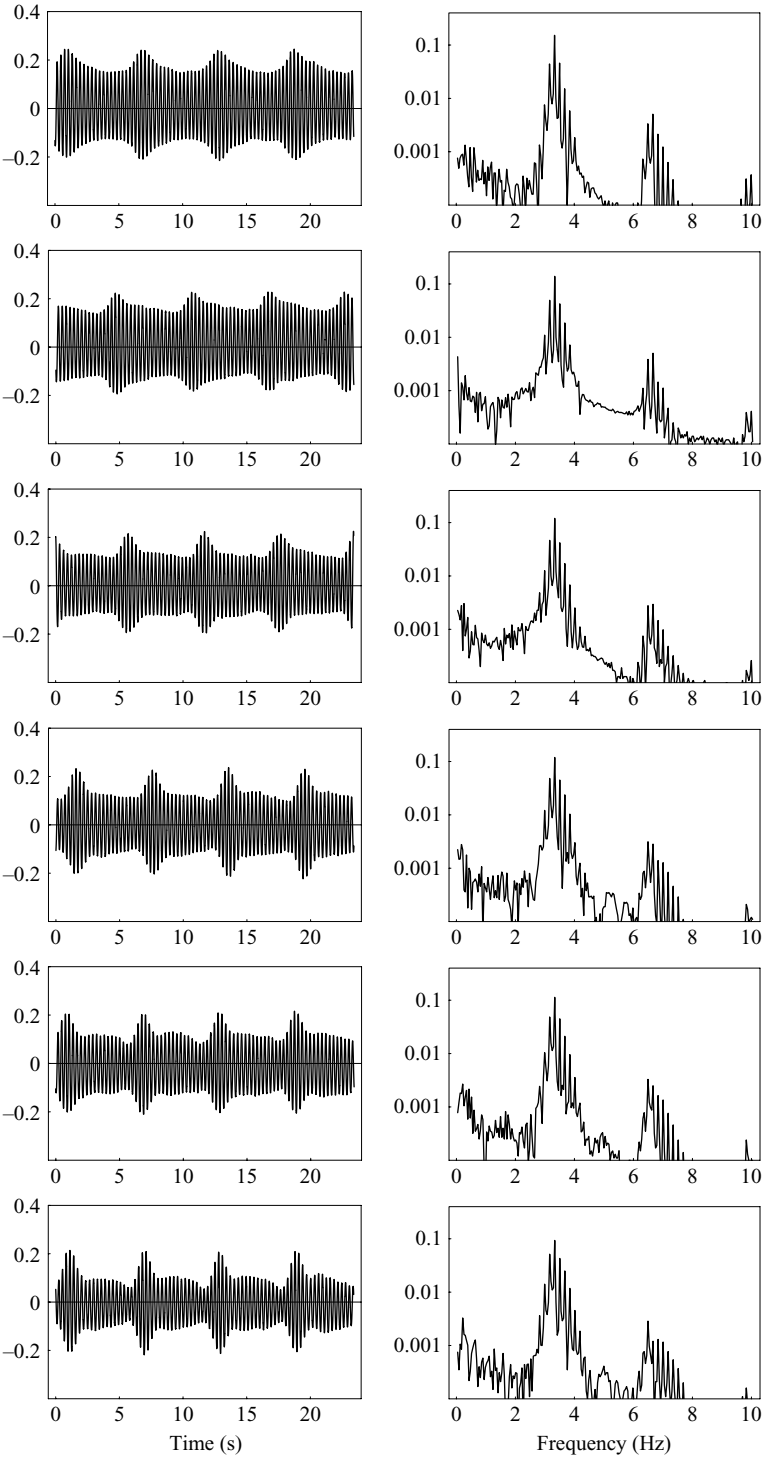

FIGURE 3. Water surface displacement (cm) as a function of time, and corresponding Fourier coefficients (cm) obtained from 6 experiments when the wave gauge was fixed at $128 + 50(n-1)$ cm from the wavemaker, for $(n = 7, 8, 9, 10, 11, 12)$.

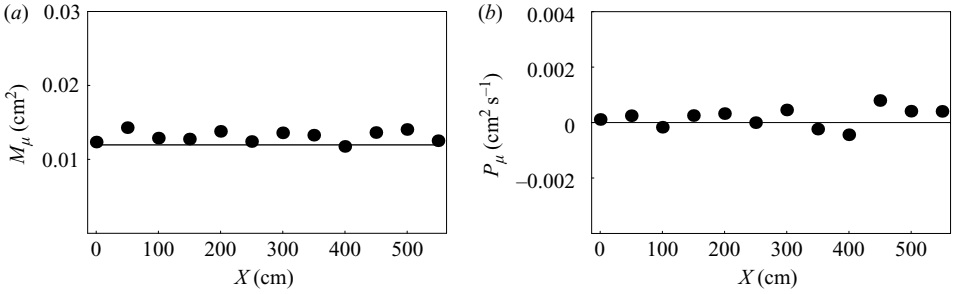

FIGURE 4. Measurements (dots) of (*a*) M_μ and (*b*) P_μ as functions of distance from the wavemaker. ($X = 0$ cm is 128 cm from the wavemaker.) The line in (*a*) is simply a horizontal line continued from the first data point.

and to the growth of these sidebands. Nevertheless, the growth of the sidebands was bounded, as predicted by (2.1). In these experiments, no sideband amplitude grew to even half the amplitude of the carrier wave.

Our experiments regularly show an asymmetric distribution of energy in these sidebands, with more sidebands excited at higher frequencies than at lower frequencies. Earlier experiments by Lake *et al.* (1977) also show this asymmetric growth. As we discuss below, in this set of experiments *the asymmetric growth can be attributed to growth from asymmetric initial conditions*, as predicted by (2.1). For larger amplitude waves, the asymmetric growth of the sideband spectrum becomes more pronounced (see § 6.5), and it cannot be predicted by (2.1). In such experiments, the asymmetric growth seems to be related to frequency downshifting. We plan to discuss the connection between asymmetric spectra and frequency downshifting in a separate paper. In the present paper, we compare the predicted and observed evolution of the carrier wave (3.33 Hz), the sidebands that resulted from the seeded perturbation (3.16, 3.50 Hz), the second sidebands (2.99, 3.67 Hz), the third sidebands (2.82, 3.84 Hz), and the second harmonic of the carrier wave (6.67 Hz).

For the evolution of wavetrains like those in figures 2 and 3, (2.1) requires that M_μ and P_μ in (5.3) be constant in the decaying reference frame (i.e. with each measured amplitude magnified by $e^{+\delta t} (= e^{+\tilde{\delta} x})$, to filter out the overall decay). Figure 4 shows the measured values of the conserved quantities for the set of experiments described by (6.1) and table 1. The data have (twice) the decay rate filtered out and are essentially constant. For M and P, the measurement errors for the experiments shown in figures 2 and 3 are ± 0.001 cm^2 and ± 0.0004 cm^2 s^{-1}, respectively. In this paper, we do not apply the predictions outlined in § 5 to experiments for which M_μ and P_μ show a non-constant trend in the decaying reference frame. Such experiments include one set using a larger-amplitude carrier wave (relative to the experiments shown in figures 2–8) with the same ratio of perturbation to carrier wave amplitudes, and one set using a perturbation with a larger amplitude (shown in § 6.5). These experiments with waves of large amplitude fall outside the scope of this paper; we will address them in a separate paper.

Figure 5 shows the measured and predicted amplitudes of the set of seeded sidebands, $\{a_{-1}(t), a_1(t)\}$. In figures 5, 6 and 7*a*, each measured amplitude is magnified by $e^{+\delta t}$ to filter out the overall decay, where δ was deduced from measured values of $M(t)$, as described above. The measured amplitudes in figure 5 have errors of ± 0.001 cm. The predicted values of the amplitudes (solid curves) were obtained by integrating (4.3) numerically, starting with initial values for $\{u_1(0), v_1(0), U_1(0), V_1(0)\}$ given in table 1. These starting values in turn were obtained from the (complex) values

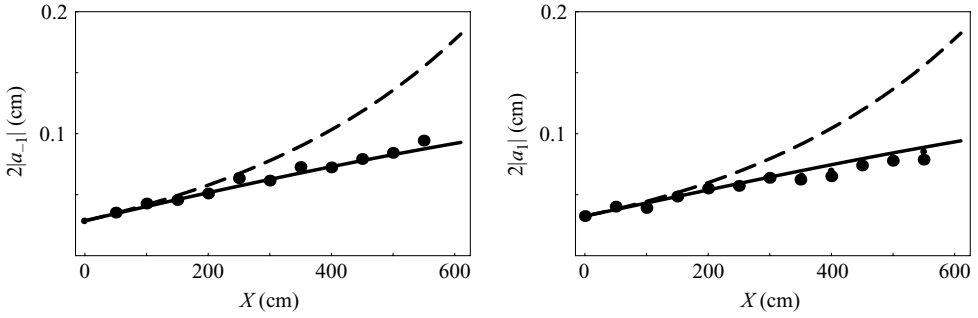

FIGURE 5. Predictions (solid curves) from (4.3) and measurements (dots) of the amplitudes of the two seeded sidebands, $|a_{-1}|$ and $|a_1|$, as functions of distance from the wavemaker. ($X = 0$ cm is 128 cm from the wavemaker.) Measured values were taken from figures 2 and 3, but amplified by $e^{+\delta X}$ to filter out the overall decay. The dashed curves show the classic Benjamin–Feir (1967) prediction of constant growth rate.

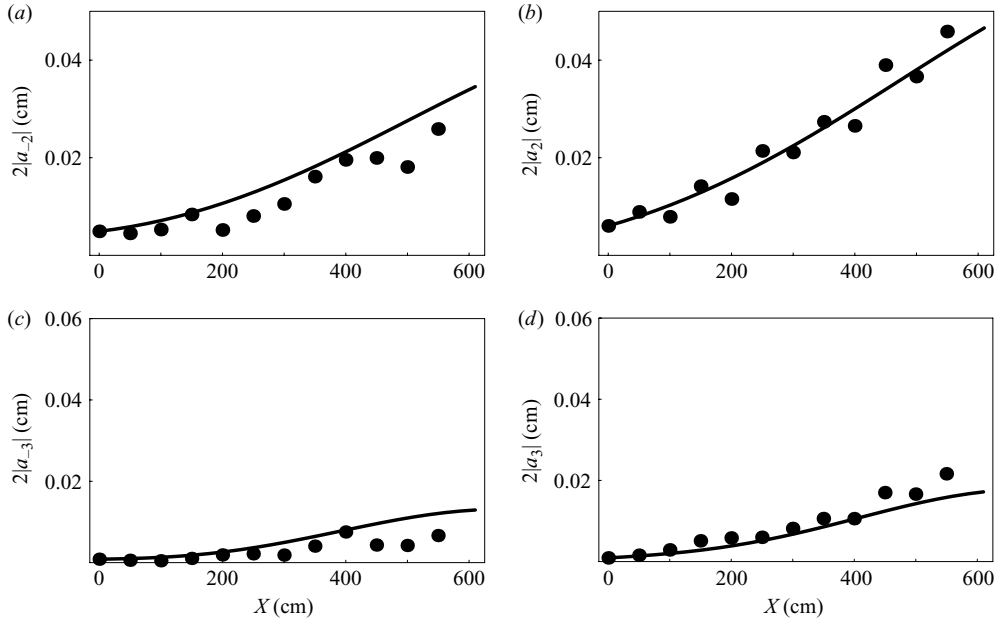

FIGURE 6. Predictions (solid curves) from (5.12) and (5.13) and measurements (dots) of the amplitudes of the second set of sidebands, (*a*) $|a_{-2}|$ and (*b*) $|a_2|$, and the amplitudes of the third set of sidebands, (*c*) $|a_{-3}|$ and (*d*) $|a_3|$, as functions of distance from the wavemaker. ($X = 0$ cm is 128 cm from the wavemaker.) Measured values were taken from the data of figures 2 and 3, but amplified by $e^{+\delta X}$ to filter out the overall decay.

of $\{a_{-1}(0), a_1(0)\}$ measured at the $n = 1$ location. The dashed curves show the growth predicted by Benjamin & Feir (1967). In the comparisons of theory and experiment shown in figure 5 and in all of the subsequent comparisons, no free parameters were available to help fit the data.

Figure 5 shows that for short 'times' (i.e. for approximately 1 m down the tank in these experiments), the damped theory (solid curve), Benjamin–Feir theory (dashed curve), and measured data all agree. For longer times, decay of the carrier wave slows the growth rate of the sidebands, as observed in the data and in the damped theory,

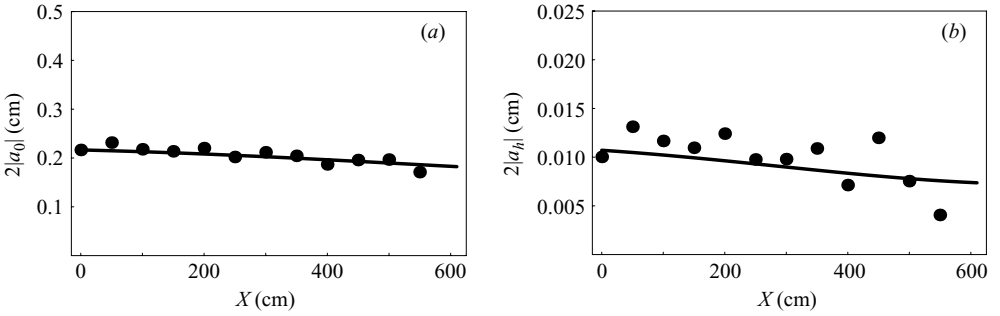

FIGURE 7. Results for the (a) carrier wave amplitude and (b) its second harmonic as functions of distance from the wavemaker. ($X = 0$ cm is 128 cm from the wavemaker.) The solid curve in (a) is the prediction from (5.16); the dots are the measured Fourier amplitudes at the carrier wave frequency, but amplified by $e^{+\delta X}$. The solid curve in (b) is the prediction from (5.18). The dots are the measured Fourier amplitudes at twice the carrier wave frequency, amplified by $e^{+2\delta X}$ as required by (5.17).

but not in the undamped theory. Recall that the undamped theory is being compared to data that has had the damping factored out. So, correcting the inviscid growth rate by subtracting the decay rate from it is inadequate for these waves.

Over the duration of these experiments, the damped theory predicts the measured growth of the sidebands from their starting values with reasonable accuracy. Equation (4.3) predicts that this growth must eventually stop completely. Continuing the computation of (4.3) beyond 6 m, we find that $|a_{-1}(t)|$ and $|a_1(t)|$ would have achieved maximum amplitudes of about 0.18 cm, or about 5.5 times their initial amplitudes, at about 10.7 m downstream from the wavepaddle. We ended our experiments before that distance to minimize the effects of reflections from the beach. As a result, we did not observe the bound on growth in this experiment, but we do see it in a second set of experiments, shown in figure 9.

Figure 6 shows the growth of the next two sets of sidebands, $\{a_{-2}, a_2, a_{-3}, a_3\}$. None of these sidebands was seeded, so they started with smaller amplitudes than $\{a_{-1}, a_1\}$, and they remained smaller. The predicted values for $\{a_{-2}, a_2\}$ were obtained by integrating (5.12) numerically, while the predictions for $\{a_{-3}, a_3\}$ were obtained by integrating (5.13). In all cases, the starting (complex) values were taken from data measured at the $n = 1$ location.

As in figure 5, the damped theory predicts the observed data over the duration of the experiment with good accuracy, and with no free parameters. The predictions for all six sidebands are in fairly good agreement with the data, which have errors of ± 0.001 cm. Note that $|a_2(t)|$ grows nearly twice as much as $|a_{-2}(t)|$ during the experiments, and that (2.5) accurately predicts this asymmetric growth. Equation (2.5) is symmetric under $\{x \to -x\}$, but figure 6 shows that (2.5) admits solutions with growing asymmetry. The initial data $\{a_{-2}(0), a_2(0)\}$ must be asymmetric, and those modes must be in the unstable region for a time. Then the asymmetric part of the solution grows approximately exponentially, for a while. Note also that $\{a_{-3}, a_3\}$ are asymmetric, but the asymmetry is less pronounced. According to (2.5), the $\{a_{-3}, a_3\}$ modes lay outside the unstable region for the entire experiment, so these two modes grew asymmetrically only because of asymmetric forcing by $\{a_{-1}, a_1\}$ and $\{a_{-2}, a_2\}$.

The theory in § 5 predicts how the carrier wave and its second harmonic evolve. Figure 7 shows the comparison of these predictions with the measured evolution.

Figure 7(*a*) shows the measured values of the carrier wave amplitude obtained from the Fourier transforms, with the overall decay rate filtered out. This amplitude varies slowly in this reference frame as it loses energy to the sidebands, as described by (5.16). The curve in figure 7(*a*) is from (5.16), using values for the sidebands given by the computed solutions of (4.3) and (5.12).

Figure 7(*b*) shows the measured and predicted evolution of the second harmonic of the carrier wave. Recall from (5.17) that the second harmonic decays with twice the decay rate of the carrier wave. Thus, twice the usual decay rate has been factored out in figure 7(*b*). The dots in figure 7(*b*) are the measured amplitudes of the harmonic, magnified by $e^{2\delta t}$. The curve shows the evolution of the harmonic predicted by (5.18), using the computed solutions of (4.3) and (5.15) with $u_0(0) = v_0(0) = 0$. Thus, the curve takes into account the evolution of the carrier wave and the first set of sidebands.

Note that the vertical scale in figure 7(*b*) is finer than that in figure 7(*a*); thus, figure 7(*b*) shows that (3.2) predicts the evolution of the harmonic quite accurately. Lake & Yuen (1977) found that (3.2) did not predict accurately the measurements in their experiments, but we found no such problems.

6.3. *Comparisons with other theories*

Figures 4–7 show that (2.1) predicts all the easily measured features of the data shown in figures 2 and 3 with good accuracy, using no adjustable parameters. Even so, this good agreement does not rule out the possibility that another theory might also predict these data accurately. For example, it is known that the initially exponential growth rate, predicted by Benjamin & Feir (1967) and shown in figure 5, can last only until nonlinear interactions among sidebands become important. For longer times, (1.1) predicts that the growth of the seeded sidebands, $\{a, a_1\}$, must diminish as these growing modes begin to lose energy to higher sidebands. Thus even with no damping ($\delta = 0$), a nonlinear theory like (1.1) also predicts that the initially vigorous growth of unstable sidebands must eventually slow down, consistent with the behaviour shown in figure 5. In terms of the behaviour of $\{a_{-1}, a_1\}$, the differences between the two theories are that:

(i) the mechanism for the slowing down is different (nonlinear interactions for (1.1) *vs.* damping of the carrier wave for (2.1)); and

(ii) the time scales on which this slowing down occurs are typically different.

Dysthe's (1979) model could also be used to predict the evolution shown in figures 2 and 3. He derived his higher-order correction to (1.1) in order to predict the behaviour of nonlinear events more accurately. In the form given by Lo & Mei (1985), using the notation given herein, Dysthe's model in one spatial dimension can be written as

$$\mathrm{i}\partial_t\psi + \alpha\partial_x^2\psi + \gamma|\psi|^2\psi + 8\mathrm{i}\varepsilon\gamma|\psi|^2\partial_x\psi - 4\varepsilon\gamma\,\partial_x\phi(|\psi|^2)\psi = 0, \tag{6.4}$$

where

$$\partial_x\phi(f) = -\frac{1}{4\pi}\int |k|\hat{f}(k,t)\,\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k, \quad f(x,t) = \frac{1}{2\pi}\int \hat{f}(k,t)\,\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k,$$

$\{\alpha, \gamma\}$ were given in § 3, and $\varepsilon = 2k_0|a_0(0)|$. Lo & Mei (1985) showed that (6.4) predicts the evolution of some narrow-banded wave packets in deep water more accurately than does (1.1).

Figure 8 repeats the data shown in figures 5, 6 and 7, showing the evolution of $\{a_0, a_{-1}, a_1, a_{-2}, a_2, a_{-3}, a_3\}$ as functions of time, with each measured amplitude magnified by $e^{+\delta t}$ to factor out the overall decay. We have also plotted the predicted evolution of each Fourier amplitude, according to (1.1), (2.1) in the form of (2.5), and

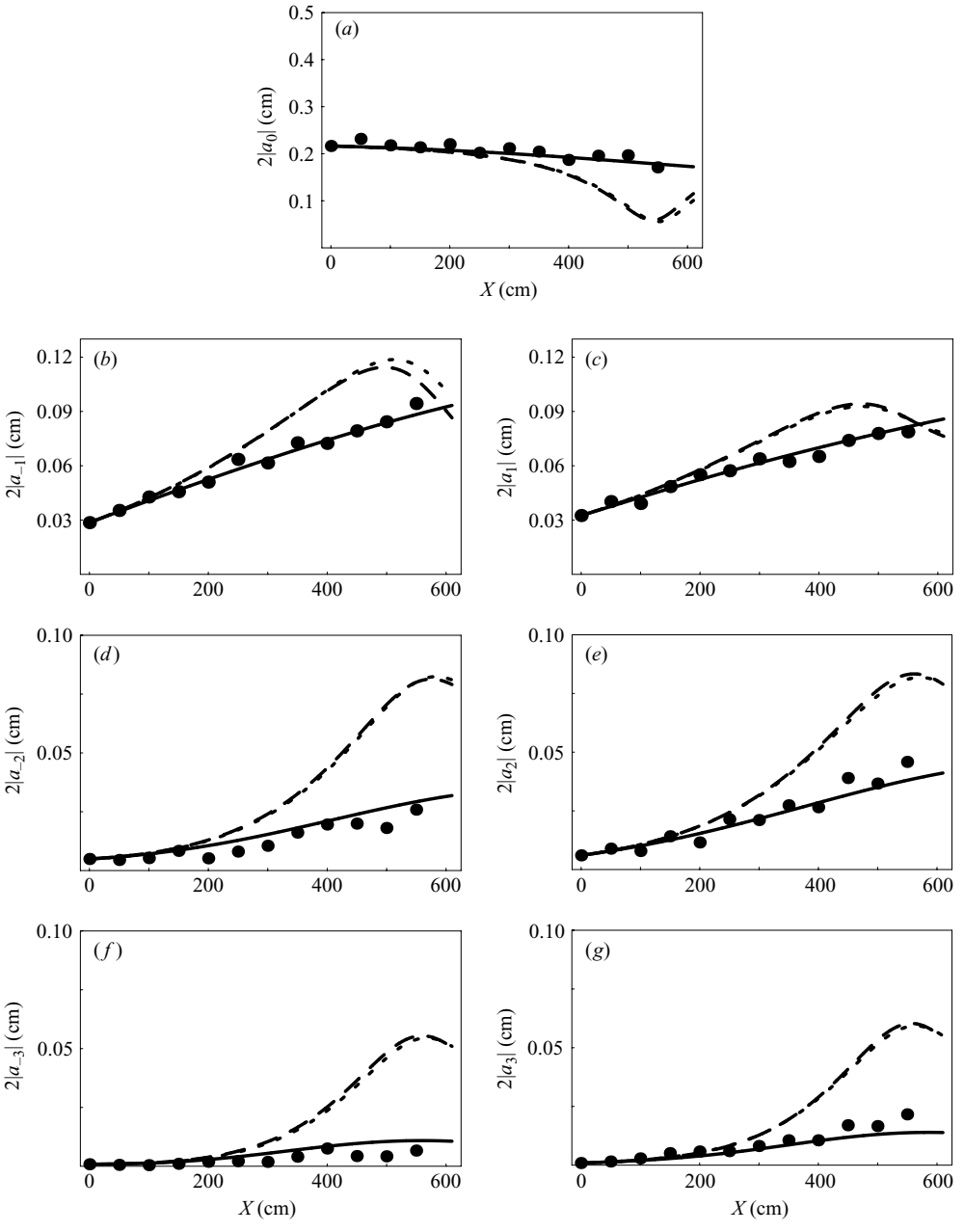

FIGURE 8. Predictions from (1.1) (dashed curves), (2.5) (solid curves) and (6.4) (dotted curves); and measurements (dots) of sidebands amplitudes: (*a*) $|a_0|$, (*b*) $|a_{-1}|$, (*c*) $|a_1|$, (*d*) $|a_{-2}|$, (*e*) $|a_2|$, (*f*) $|a_{-3}|$, (*g*) $|a_3|$ as functions of distance from the wavemaker. ($X = 0$ cm is 128 cm from the wavemaker.) Measured values were taken from the data of figures 2 and 3, but amplified by $e^{+\delta X}$ to filter out the overall decay.

(6.4). We note that for the experiment shown in figure 8, the results between NLS (1.1) and the NLS with higher-order terms (6.4) are almost indistinguishable for most of the Fourier amplitudes.

For this set of experiments, the damped NLS model, (2.1) in the form of (2.5), predicts the evolution of every measured amplitude much more accurately than either of the undamped models, (1.1) or (6.4). This striking discrepancy in accuracy among these three mathematical models illustrates one of our main points. Equations (1.1) and (6.4) both predict that unstable sidebands stop growing after their amplitudes become large enough that nonlinear interactions among sidebands become dynamically important. Equation (2.1) provides another option: damping of the carrier wave can slow and eventually stop the growth of the unstable sidebands altogether, before their amplitudes become large enough that nonlinear interactions play a role. When this happens, sideband amplitudes always remain small, and the difference between nonlinear terms in (1.1) and (6.4) has little effect on the evolution of the wavetrain. Figure 8 demonstrates that this option occurs in the experiments shown in figures 2 and 3: damping controls the growth of the sidebands, precluding serious nonlinear effects.

We are not suggesting that in the presence of damping, the processes described by (1.1) or (6.4) never occur. If the initial amplitudes of the seeded perturbations had been larger, if the amplitude of the carrier wave had been larger, or if the damping rate had been smaller, then nonlinear interactions among growing sidebands might have become important before damping effects took over. Instead, theorems 1 and 2 assert that for fixed carrier-wave amplitude and fixed damping rate (so $|\gamma||A|^2/\delta$ is fixed in (4.8) or (4.19)), then a sideband perturbation that is small enough initially must remain small forever. In this way, a uniform train of plane waves of moderate amplitude in deep water is stable, for any $\delta > 0$.

6.4. *A second set of experiments*

Figure 9 shows results from a second set of experiments, in which the seeded perturbation was chosen so that the B–F theory would predict instability even when the decay rate was subtracted from the growth rate, while the dissipative theory would predict a bound to growth that could be observed within the wave tank. For this experiment, $\omega_0/2\pi = 3.33\,\text{Hz}$, $\omega_p/2\pi = 0.39\,\text{Hz}$ and $\tilde{\delta} = 0.115\,\text{m}^{-1}$. The time series shown in figures 9(*a*) and 9(*c*) were obtained at X_1 and X_{14}, 650 cm apart. In this experiment there is little evolution of the wavetrain other than that due to damping, because the sidebands never grow much. The Fourier amplitudes of the carrier wave and the two seeded sidebands are shown in figures 9(*e*)–9(*g*). The dots are the measurements. The curves show computations from the linearized stability theory that includes damping, (4.3) (solid curves), and the NLS equation that includes higher-order terms, (6.4)–(dotted curves). For both sets of curves, the starting values of the complex Fourier amplitudes were:

$$2a_0 = -0.014 - 0.229\,\text{i},\ 2a_{-1} = 0.012 + 0.001\,\text{i},\ 2a_1 = -0.007 + 0.009\,\text{i},\ |a_n| = 0,\ |n| > 1.$$

As in figure 8, the data in figure 9 agree reasonably well with predictions from the stability theory for the uniform solution of (2.5), and agree less well with predictions from (6.4). According to (2.1), damping essentially stops the sideband growth, consistent with the observed behaviour. By contrast, this effect of damping is neglected by both (1.1) and (6.4), which predict continued growth of sidebands.

We reiterate: in the presence of damping, growth of an unstable sideband can stop either because of nonlinear interactions among sidebands, or because damping has quenched the growth of the sideband by decreasing the amplitude of the carrier wave (or both). For a fixed value of $\{|\gamma||A|^2/\delta\}$, we can always guarantee that

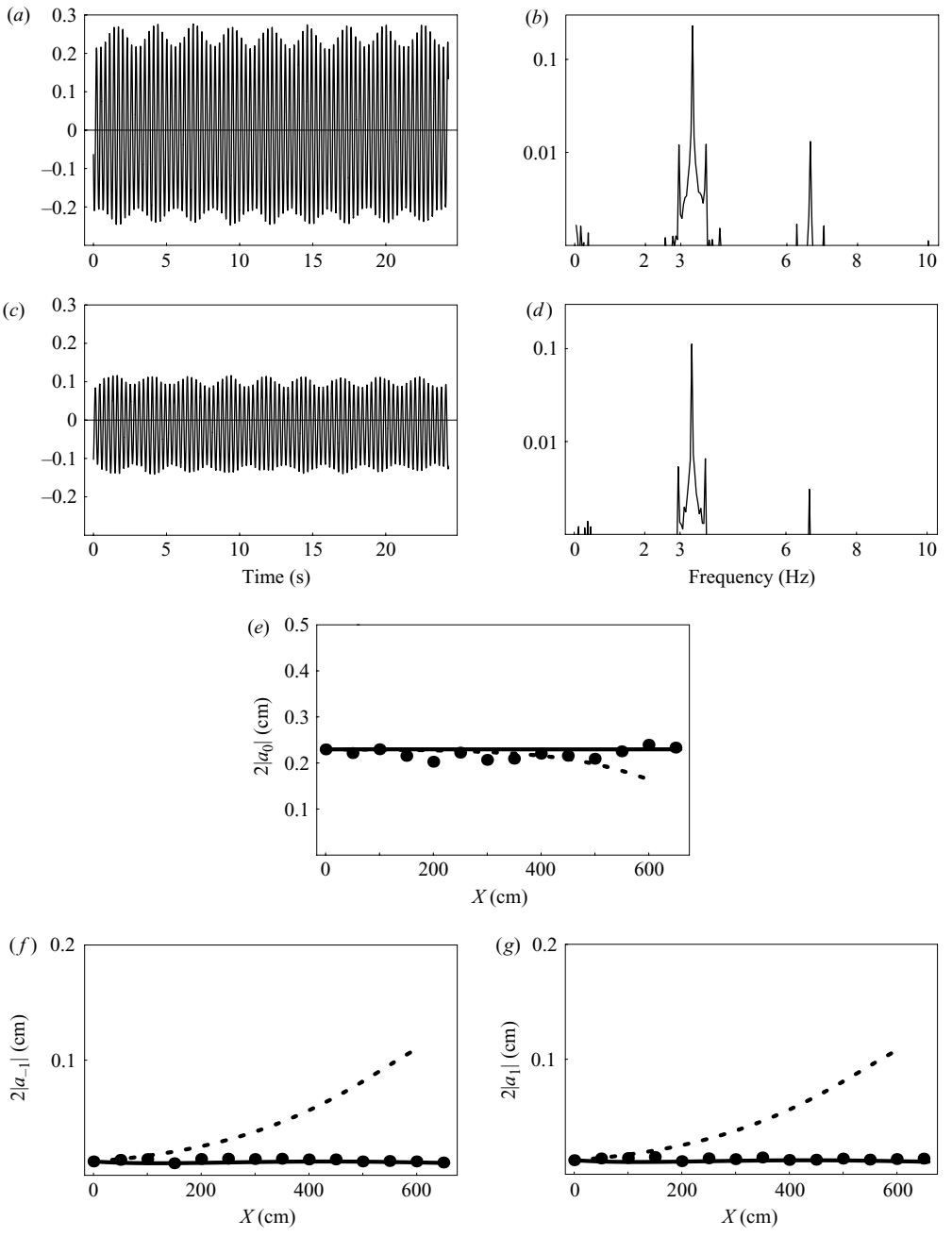

FIGURE 9. Water surface displacement (in cm) obtained at (*a*) X_1 and (*c*) X_{14} with corresponding Fourier coefficients (in cm) ((*b*) and (*d*)) from an experiment in which unstable sideband growth stopped within the test section. Comparisons for the carrier wave, (*e*) $|a_0|$, and the first set of sidebands, (*f*) $|a_{-1}|$, and (*g*) $|a_1|$, show measured amplitudes (dots) as functions of distance from the wavemaker ($X = 0$ cm is 128 cm from the wavemaker) amplified by $e^{+\tilde{\delta}X}$ to filter out the overall decay, along with predictions from (*e*) (5.15) (solid curve) and (6.4) (dotted curve) and (*f*), (*g*) (4.3) (solid curve) and (6.4) (dotted curve).

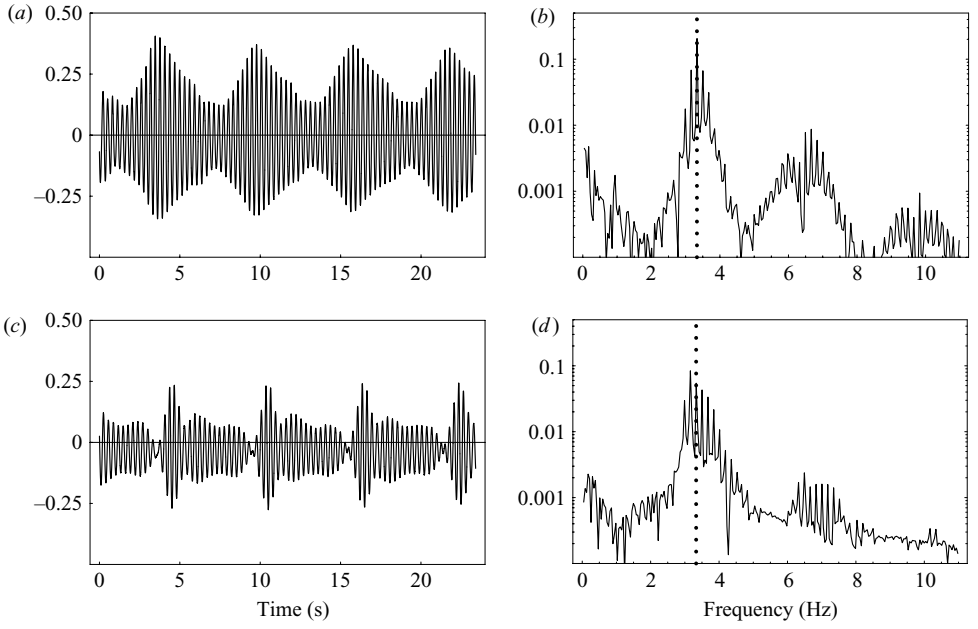

FIGURE 10. Water surface displacement (in cm) obtained at (*a*) X_1 and (*c*) X_{12} with corresponding Fourier coefficients (in cm) ((*b*) and (*d*)) for an experiment outside the range of validity of (2.1). The vertical dotted lines in (*b*) and (*d*) show the location of the carrier wave frequency.

damping quenches sideband growth before nonlinear interactions among sidebands become important, by requiring that the initial sideband amplitudes be small enough.

6.5. *Large-amplitude waves*

In the two sets of experiments shown, in figures 2–8 and in figure 9, (2.1) predicts the observed evolution of all the easily measurable features of the data with reasonable accuracy. We found this kind of agreement for all of our experiments in which the carrier wave amplitudes were small or moderate and the initial perturbation amplitudes were also small relative to that of the carrier wave. In the set of experiments discussed next, the carrier wave amplitude was about the same as that used in the experiments shown in figures 2–8, but the initial perturbation amplitude was larger. The result was that (2.1) failed to predict the wave evolution accurately. Thus, the next experiments demonstrate that (2.1) is valid only when perturbations are initially relatively small.

Figure 10 shows the evolution of a modulated wavetrain, as it propagates down the wavetank. Many of the parameters in this set of experiments ($\omega_0/2\pi = 3.33$ Hz, $\omega_p/2\pi = 0.17$ Hz, $\tilde{\delta} = 0.120\,\mathrm{m}^{-1}$, $2k_0|a_0(0)| = 0.093$), are comparable to those in the experiments shown earlier. One difference is that the initial amplitude of the perturbative waves was larger so that the ratio of perturbation to carrier wave amplitudes, $(|a_{-1}| + |a_1|)/(2|a_0|)$, was 0.33, as opposed to 0.14 for the experiments in figures 2–8 and 0.05 for those shown in figure 9. The data in figure 10 look qualitatively similar to those in figures 2, 3 and 9, with the modulation being more

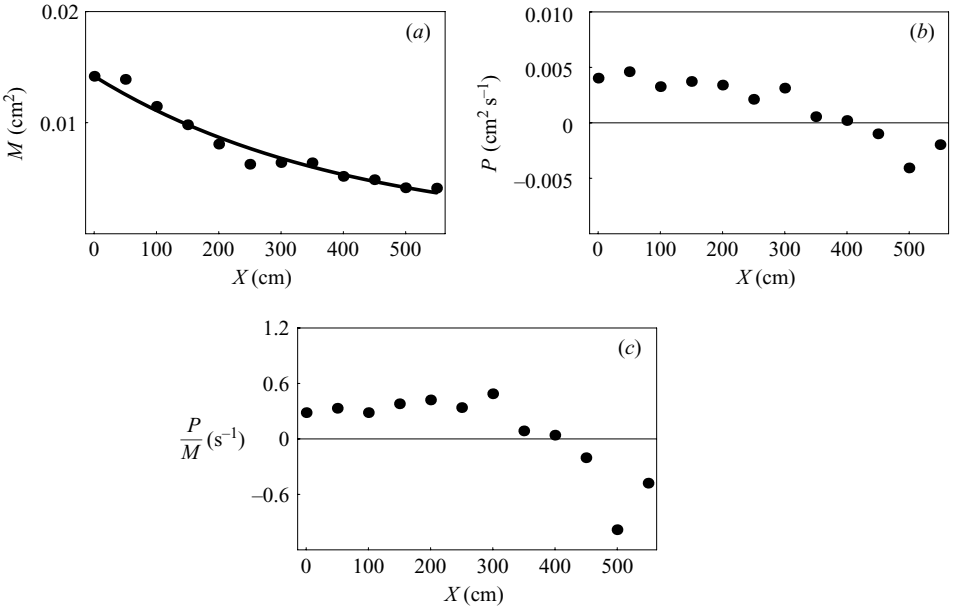

FIGURE 11. Integral quantities, (*a*) $M(t)$, (*b*) $P(t)$, and (*c*) their ratio, in an experiment with an initially large perturbation amplitude. The curve in (*a*) is an exponential fit with $\tilde{\delta} = 0.12\,\mathrm{m}^{-1}$. Here, the entire measured spectra are used to compute the quantities.

pronounced. A striking difference between this experiment and the previous two is seen in the Fourier transform of the time series obtained at the last measurement station, X_{12}. There, the energy of the first lower sideband had grown larger than the energy of the carrier wave. This downshift of energy to the lower sideband is what we refer to as 'frequency downshifting'. We saw no evidence of wave breaking in these experiments. Although the time series in figure 10 are qualitatively similar to those in figures 2, 3 and 9, the wave evolution shown in figure 10 differs significantly from that seen in figures 2, 3 and 9. Figure 11 shows the evolution of the integral quantities, $M(t)$ and $\boldsymbol{P}(t)$, defined in (2.2), as well as $\{\boldsymbol{P}(t)/M(t)\}$ for the data in figure 10. We can fit an exponential to the decay of $M(t)$, leading to $\tilde{\delta} = 0.120\,\mathrm{m}^{-1}$ for this experiment. However, $\boldsymbol{P}(t)$ changes sign, so clearly $\boldsymbol{P}(t)$ is not decaying exponentially. Recall from (2.3) that if (2.1) is a valid model, then both $M(t)$ and $\boldsymbol{P}(t)$ must decay to zero, with the same decay rate. Equivalently, $\{\boldsymbol{P}(t)/M(t)\}$ must be constant in time, as must $\mathrm{e}^{2\delta t}P(t) = \boldsymbol{P}_{\mu}(t)$. Figure 11(*c*) clearly shows that $\{\boldsymbol{P}(t)/M(t)\}$ varies in time. Thus, figure 11 shows that the experimental data shown in figure 10 lie outside the range of validity of (2.1).

This failure is not restricted to (2.1). We can show by direct calculation that both $M(t)$ and $\boldsymbol{P}(t)$ are constants of the motion for (1.1), and also for (6.4). Thus, $\{\boldsymbol{P}(t)/M(t)\}$ is a constant of the motion for any of these models: (1.1), (2.1), (6.4). Figure 11 shows that the experimental data shown in figure 10 lie outside the range of validity of all three of these mathematical models.

The simultaneous failure of all three of these mathematical models is important physically, because it suggests that the effects of 'large amplitude' might appear in wave data at much smaller amplitudes than had previously been recognized. Such an important point ought not to depend critically on the specific definitions of

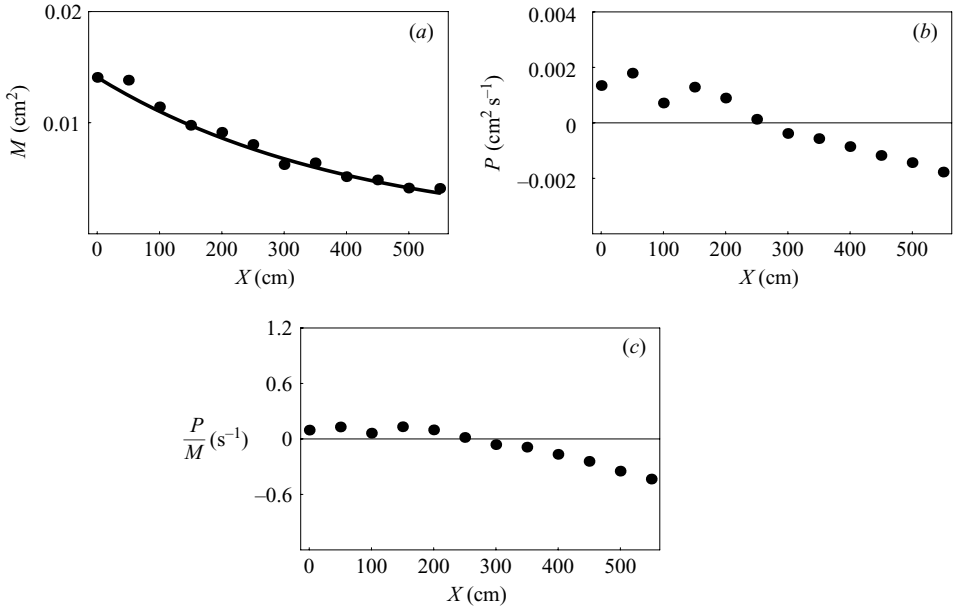

FIGURE 12. Integral quantities, (*a*) $M(t)$, (*b*) $P(t)$, and (*c*) their ratio, in an experiment with an initially large perturbation amplitude. The curve in (*a*) is an exponential fit with $\tilde{\delta} = 0.120\,\mathrm{m}^{-1}$. Here the quantities are computed using a portion of the measured spectra around the carrier wave.

the integrals involved. In figure 11, $M(t)$ and $\boldsymbol{P}(t)$ were calculated from the entire signal measured at each location X_n, $n = 1, 2, \ldots, N$. However, the three mathematical models in question, (1.1), (2.5) and (6.4), all assume 'narrow bandwidths', i.e. they assume that the energy is restricted to a relatively narrow band of frequencies near that of the carrier wave. This assumption suggests that we should filter our data, and calculate $M(t)$ and $\boldsymbol{P}(t)$ based only on frequencies within a narrow range of the carrier wave. Figure 12 shows the measured evolution of $M(t)$ and $\boldsymbol{P}(t)$, after the measured signal has been filtered by retaining the signal only within $1.67 < \omega/(2\pi) < 5.00$. Comparing the results in figures 11 and 12 suggests that this filtering changes $M(t)$ very little; it smoothes $\boldsymbol{P}(t)$ somewhat, but it does not change our main qualitative point: $\{\boldsymbol{P}(t)/M(t)\}$ varies in time for these experiments, so these data lay outside the range of validity of any of the models, (1.1), (2.1) or (6.4). For this experiment the resolutions for $M(t)$ and $\boldsymbol{P}(t)$ are $0.001\,\mathrm{cm}^2$ and $0.001\,\mathrm{cm}^2\mathrm{s}^{-1}$. Though the variation in $\boldsymbol{P}(t)$ is a small effect, it is resolvable.

Results similar to those shown in figures 10–12 also occurred in experiments in which we increased the carrier wave amplitude and kept the initial perturbation amplitudes small. We classify all such experiments as 'large amplitude'. They lie outside the range of the validity of any of (1.1), (2.1) or (6.4).

For comparison, recall that the original derivation by Benjamin & Feir (1967) was limited to waves of small or moderate amplitude, as was the derivation of (1.1) by Zakharov (1968), and the derivation of (6.4) by Dysthe (1979). If we think of the 'Benjamin–Feir instability' as limited by the assumptions made in their original derivation, then our results show that any amount of dissipation stabilizes the Benjamin–Feir instability. Alternatively, if we use 'Benjamin–Feir instability' to

describe events outside of the range stated in the original work of Benjamin & Feir, then the previous sentence is too strong.

6.6. *What does 'small or moderate amplitude' mean?*

The data shown in figures 2–9 demonstrate that the theory presented in this paper describes accurately the evolution of a nearly uniform train of nearly monochromatic waves of small or moderate amplitude in deep water, provided the perturbations are small enough. The data in figures 10–12 show what can happen if either the amplitude of the carrier wave is large enough, or if the perturbations are large enough. For waves with larger amplitudes, we have found experimentally that our theory loses accuracy, and also that $\{\boldsymbol{P}(t)/M(t)\}$ decreases monotonically, as shown in figures 11(*c*) or 12(*c*). This ratio should be constant according to any of (1.1), (2.1) or (6.4), so its monotonic decrease indicates that all of these theories fail together. As far as we can tell, the process that forces $\{\boldsymbol{P}(t)/M(t)\}$ to decrease is dissipative. If so, then any generalization of (1.1) or (6.4) that includes more terms in an asymptotic expansion within a non-dissipative model like Euler's (conservative) equations will fail to capture the physics of large-amplitude events.

What happens to waves of larger amplitude, which fall outside the range of validity of the theory presented here? This topic is worthy of a separate study, and we do not address it here. For this paper, we restrict our comments simply to defining what we mean by 'moderate amplitude' and 'large amplitude'. Based on our experimental results, the critical issue seems to be whether $\{\boldsymbol{P}(t)/M(t)\}$ is conserved during the evolution. If $\{\boldsymbol{P}(t)/M(t)\}$ is conserved as the wavetrain evolves, then we classify the wavetrain as having moderate amplitude and use (2.1). In our experiments, $\{\boldsymbol{P}(t)/M(t)\}$ was not conserved when using wavetrains with larger carrier wave amplitudes or larger perturbation amplitudes; in these experiments, (2.1) often failed to predict the evolution of more than one measured quantity.

We can state this criterion in various ways. Instead of monitoring $\{\boldsymbol{P}(t)/M(t)\}$, we can use the evolution of $M(t)$ to define δ for a particular set of experiments; then the wavetrain in those experiments has moderate amplitude if $\mathrm{e}^{2\delta t}\boldsymbol{P}(t)$ is constant (to within experimental error) during the experiments. This is mathematically equivalent to monitoring $\boldsymbol{P}_{\mu}(t)$, defined in (2.6).

A disadvantage of this criterion is that it is not predictive: we must run the experiment and observe $\{\boldsymbol{P}(t)/M(t)\}$ in order to use the criterion. Two referees of this paper asked that we provide a simple a *priori* estimate to predict whether a given wavetrain lies in the range of validity of the theory presented here.

This is an important question, but we are currently unable to give a complete answer to it. Here is what we know.

(i) A common measure of nonlinearity for a nearly monochromatic train of waves in deep water is the dimensionless slope of the wave, $\varepsilon = 2k_0|a_0(0)|$. (Recall that the extra factor of 2 arises because our crest-to-trough amplitude is $2|a_0(0)|$.)

(ii) The theory presented here fails for experiments that show wave-breaking, frequency downshifting or both; neither of these phenomena is predicted by (1.1), (2.1) or (6.4). Melville (1982) found experimentally that nearly monochromatic spatially uniform wavetrains showed signs of both wave-breaking and frequency downshifting for $\varepsilon > 0.16$.

(iii) Our experiments showed measurable changes in $\mathrm{e}^{2\delta t}\boldsymbol{P}(t)$ for $\varepsilon > 0.14$. This is consistent with Melville's observations.

(iv) The experiments in figure 10 show that frequency downshifting can occur even for wavetrains with $\varepsilon < 0.14$, if the perturbations are large enough.

(v) Finding an accurate, unambiguous criterion to determine which wavetrains are 'large' enough to produce frequency downshifting or wave breaking is an important open question, in our minds.

When the model (2.1) is valid, the stability theory outlined in § 4 says that dissipation bounds the growth of sidebands before nonlinear interactions become important, provided the sidebands have a small enough initial amplitude. Here are two thoughts concerning the application of this theory.

(i) A bound on growth due to dissipation is given by (4.5) to be $e^{|\gamma|M/\delta}$. An approximate WKB analysis of (4.3), which we do not outline here, suggests that the total growth is more likely to be about $e^{|\gamma|M/3\delta}$ when M/δ is large. In the experiments presented herein, dissipation bounded the growth of sidebands before nonlinear interactions became important. We do not have a bound above which nonlinear interactions become important, but no downshifting occurs.

(ii) For the experiments discussed herein, the initial data had a narrow bandwidth and dissipation kept the bandwidth narrow, so that the higher-order model (6.4) that includes broader bandwidth effects, but not dissipation, did not apply. If the data initially had a broader bandwidth, then a higher-order model might be more appropriate, but our results indicate that modifying (6.4) to include dissipation is likely to be necessary.

7. Comparison with earlier experimental results

Our main result is that even a small amount of dissipation controls the Benjamin–Feir instability for nearly monochromatic waves of moderate amplitude in deep water. However, this instability is well established in water waves, in optics and elsewhere. In this section, we re-examine some of the earlier experimental work that is cited in support of the B–F instability.

In the literature on waves in deep water, two kinds of experiment have been done that relate to the Benjamin–Feir instability. Experiments by Feir (1967), Yuen & Lake (1975), Su (1982), and Trulsen & Stansberg (2001) focused on the evolution of spatially localized wave packets, and these do not concern us here. Instead, we concentrate on experiments that involve modulations of a continuous train of nearly monochromatic waves, because they are consistent with the periodic boundary conditions used in this paper. Specifically, we now reconsider the experiments described by Benjamin (1967), Lake *et al.* (1977), Lake & Yuen (1977) and Melville (1982) on waves in deep water, and by Tai *et al.* (1986) on electromagnetic waves in an optical fibre. In all of these experiments, the observed waves were essentially one-dimensional, corresponding to $\beta = 0$ or $\partial_y \equiv 0$ in (1.1) or (2.1).

In our reading of these papers, all of the experimental studies designed to test the Benjamin–Feir mechanism found discrepancies between the non-dissipative theory and physical experiments. Benjamin (1967) described preliminary experimental results, which provided qualitative support for the new theory of Benjamin & Feir (1967). Those experiments showed:

(i) energy grows in sideband frequencies, at the expense of the carrier wave;

(ii) the range of unstable sideband frequencies is set by the amplitude of the carrier wave (as shown in (4.4), with $\beta = 0$ for one-dimensional waves and $\delta = 0$ for no dissipation);

(iii) no actual, physical growth occurs unless the inviscid growth rate exceeds the viscous damping rate.

All of these results are consistent with (2.1). It is not necessary to conclude from them that there is an instability, as defined by (2.10) and (2.11).

In his figure 4, Benjamin (1967) compared measured growth rates with values predicted by the inviscid theory, with a correction for viscous decay. (His correction was: {observed growth rate = predicted growth rate – viscous decay rate}. Our procedure of amplifying the measured data by $e^{\delta t}$ is nearly equivalent to his.) His presentation of the data inherently assumes that any sideband growth is exponential, so it precludes the kind of varying growth shown in our figure 5. He found that observed growth rates are consistently smaller than those predicted by the Benjamin–Feir theory, even after correcting for damping, with discrepancies ranging from 40 % to 100 %. As our figures 5, 6, 8, 9 show, actual growth rates decrease with 'time' (i.e. as the wave propagates down the tank). Benjamin's correction gives a growth rate that is correct for short times, but too large for longer times.

Thus, the ground-breaking experiments reported by Benjamin (1967) show clearly that energy in sidebands can grow at the expense of the carrier wave, but they do not actually establish an instability. Further, they show that the Benjamin–Feir theory over-predicts the observed growth rate, even with a correction for damping. All of this is consistent with (2.1).

Benjamin (1967) reports a measured damping rate of 0.01/wavelength in his experiments. In Benjamin's notation, our measured damping rate is 0.016/wavelength $[=\tilde{\delta}(2\pi/k_0)]$ for the experiments shown in figures 2–8. Thus, our damping rate was comparable to Benjamin's. None of the other experimental papers on waves in deep water (Feir 1967; Lake & Yuen 1977; Lake *et al.* 1977; Melville 1982; Su 1982; Tulin & Waseda 1999; Trulsen & Stansberg 2001) reported measured damping rates.

The discrepancies shown by Benjamin (1967) inspired Lake, Yuen, and their colleagues to conduct a second set of experiments to test the Benjamin–Feir theory. Lake & Yuen (1977) reported that in their experiments, the carrier wave and its second harmonic did not satisfy our (3.2). They attributed this difficulty to problems with their paddle, which could be consistent with the problems with paddle motion cited by Flick & Guza (1980). However, in a later paper Lake & Yuen state, 'this effect is far less significant than was believed and should be disregarded' (See Crawford *et al.* 1981, p. 184). Regardless of their reasoning, our figure 7(*b*) shows that our data do satisfy (3.2), so their problems with (3.2) are not inherent in deep-water waves.

Based on our own experiments, we wonder if some of their problems with (3.2) can be attributed to reflected waves in their tank. (See Lake *et al.* 1977, p. 52: 'Each experimental condition or run was recorded for at least 12 min of real time.' Their tank was 40 ft long, so a deep-water wave with a frequency of 3 Hz travels at its own group velocity down the tank and back in about 94 s, much shorter than 12 min.) For comparison, our experiments ran for much shorter time intervals, and our figure 7(*b*) shows good agreement with (3.2). In an experiment run for their long times, the tank probably contains waves running both upstream and downstream. A probe at a fixed location cannot determine the direction of propagation of a wave passing the probe, but only waves moving downstream should be included in (3.2).

A major conclusion of Lake & Yuen (1977) was that the one-dimensional wave mode with the maximal growth rate did not satisfy our (3.7*b*) with $l = 0$, as predicted by the Benjamin–Feir theory. Translating their result into our notation, they found that the most unstable mode satisfied

$$m = 0.78\sqrt{\frac{\gamma}{\alpha}}|A|, \quad \text{instead of} \quad m = \sqrt{\frac{\gamma}{\alpha}}|A|$$

as predicted by (3.7*b*). This finding is consistent with (2.1), because the mode that is most unstable at $t = 0$ ceases to be most unstable for $t > 0$. As a result, modes with wavenumbers smaller than $m = \sqrt{\gamma/\alpha}|A|$ can show more total growth over the length of the test section, and this is what Lake & Yuen (1977) observed.

A second major conclusion of Lake & Yuen (1977, p. 56) is that '... the data show that the exponential rate of sideband growth remains essentially constant, indicating that the second-order effect of dissipation is also unimportant during the initial stage of evolution...' This claim is consistent with our figure 5 for about the first 2 m of propagation in our experiments, but it contradicts our results for the remaining 4 m. In order to create a longer 'effective' test section, Lake & Yuen spliced together results from several experiments in creating their figure 3. As far as we can tell, their splicing procedure did not allow the carrier wave to decay as it would have in a single experiment. We wonder about the validity of their splicing procedure. Certainly our results contradict theirs, except for short distances.

Melville (1982) studied the relation between the B–F instability and wave breaking in deep water. As mentioned above, the original analysis of Benjamin & Feir (1967) was restricted to waves of moderate amplitude, so Melville's definition of 'B–F instability' is generalized from that originally discussed by Benjamin & Feir to include waves of large amplitude. Melville's carrier wave amplitudes were much larger than those considered here, which led him to conclude that his experiments were outside the range of validity of (1.1). There seems to be little overlap between his experimental regime and ours: our results apply to waves of small and moderate amplitude, while he considered primarily waves of large amplitude. Tulin & Waseda (1999) were also interested in breaking waves, and considered primarily waves of large amplitude.

Finally, Tai *et al.* (1986) described their experiments on modulational (i.e. B–F) instabilities of electromagnetic waves in an optical fibre. Like the earlier experiments of Benjamin (1967), their experiments show that sidebands can gain energy at the expense of the carrier wave, but it is impossible to tell from their data whether this growth is exponential, or variable (as predicted by (2.1)). No actual growth rates are presented. Like Lake & Yuen (1977), they found that the frequency-difference between the carrier wave and the most unstable sideband mode increased as the carrier wave amplitude increased, but the increase was slower than that predicted by (3.6*b*). All of this is consistent with (2.1) instead of (1.1), so it is consistent with a conclusion of sideband growth, but no actual instability.

In summary, all of the studies known to us that compare the predictions of Benjamin & Feir (1967) with experimental data show that energy in sidebands can grow at the expense of the carrier wave, and our experiments show this as well. We found no experimental studies showing that the growth is actually exponential, instead of the variable growth predicted by (2.1). Similarly, we found no experimental studies showing that the growth is part of an actual instability in the sense of (2.10) and (2.11). Our own experiments, discussed in § 6, show that the growth is not exponential, and that there is no actual instability for waves of moderate amplitude.

We are grateful to Bernard Deconinck for useful conversations, and to Guillemette Caulliez for suggesting the experiment shown in figure 9. We dedicate this paper to the memory Joe Hammack, who passed away during its publication, and to the memory of Bill Pritchard, who built the wave tank used herein with the idea of revisiting the Benjamin–Feir instability. This work was funded in part by the National Science Foundation, under grants DMS-9972210, DMS-0139771, DMS-0139742, DMS-0139847, and by Packard and Sloan Fellowships.

Appendix A. Experiments on damping of monochromatic waves in deep water

Here we present experiments whose purpose was to determine if narrow-banded, deep-water waves decay exponentially, and if so, how the decay rate depends on wave amplitude and the age of the water surface. The experiments presented in this Appendix differ from those presented in § 6 in two fundamental ways:

(i) The wavemaker was programmed with the position and velocity of a uniform wavetrain, $\eta_w(T) = 2\tilde{a}_w \sin(\omega_0 T)$, rather than a modulated wavetrain. For all of the experiments presented in this Appendix, $\omega_0 = (2\pi)3.33$ rad s^{-1}.

(ii) The gauge was mounted on a carriage that was programmed to travel down the tank at 12.21 cm s^{-1}, half the linearized group velocity of a 3.33 Hz wavetrain. Thus, for the results presented in this Appendix, only one experiment was conducted to obtain each decay rate, rather than N experiments.

Two plunger wavepaddles were used for these experiments: the one discussed in § 6 that had an exponential cross-section (E), and one that had a triangular cross-section (T). The procedure for these experiments was as follows.

(i) The tank was cleaned with alcohol and overfilled.

(ii) The surface was cleaned as described in § 6. The age of the water surface was $T_s = 0$ upon cleaning.

(iii) The T-plunger or E-plunger was oscillated as in (i) above at 3.33 Hz to generate monochromatic waves.

(iv) The gauge travelled down the wavetank at 12.21 cm s^{-1} after the wavemaker ran for 20 s, to avoid transients at the wave front. It measured 12 000 samples at 200 Hz. Figure 13(*a*) shows time series from a set of experiments using the T-plunger at an oscillation amplitude of 0.5 cm. We can see a Doppler shift at about 5 s when the carriage began travelling down the tank. (This experiment corresponds to that of the second column in table 2.)

(v) We cut off the part of the series obtained before the carriage began moving and broke up the remaining time series into intervals that were 1024 pts each and that overlapped at the midpoints.

(vi) We integrated numerically the time series to find the L_2 norm of each interval.

(vii) We found $\tilde{\delta}$, the measured spatial decay rate, by fitting an exponential to the values of the L_2 norm at positions downstream of the wavepaddle, according to $L_2(X) = L_2(0)\mathrm{Exp}(-2\tilde{\delta}X)$. The dots in figure 13(*b*) show the measurements of the L_2 norm (from step (vi)) and the line shows the exponential fit.

(viii) Without further cleaning of the water surface, we repeated the experiment at later times up to 6 days to determine the measured spatial decay rate as a function of water surface age, T_s.

Table 2 shows the results of these experiments. The first column has the surface age, the amount of time that passed after the surface was cleaned, either in minutes or in days. The remaining columns have the measured decay rates when either the triangular or exponential plunger type wave-paddle was oscillated at the given amplitude. The result was that $\tilde{\delta}$, the measured spatial decay rate, was about constant for 2 h after cleaning the surface and then increased over a period of days. We note that the decay rate did not appear to depend on the carrier wave amplitude.

Miles (1967) provided formulae for theoretical values of damping rates that take into account boundary layers at the sidewalls and the bottom, and surface effects. There are three standard ways to model the effects of the surface on damping. (*a*) Assume the surface is clean. Then for the 3.33 Hz carrier wave, the theoretical spatial damping rate is 0.037 m^{-1}, which is about 1/3 the value we measured for $T_s < 2$ h. (*b*) Assume the surface is immobile; it can oscillate vertically, but cannot stretch horizontally. Then

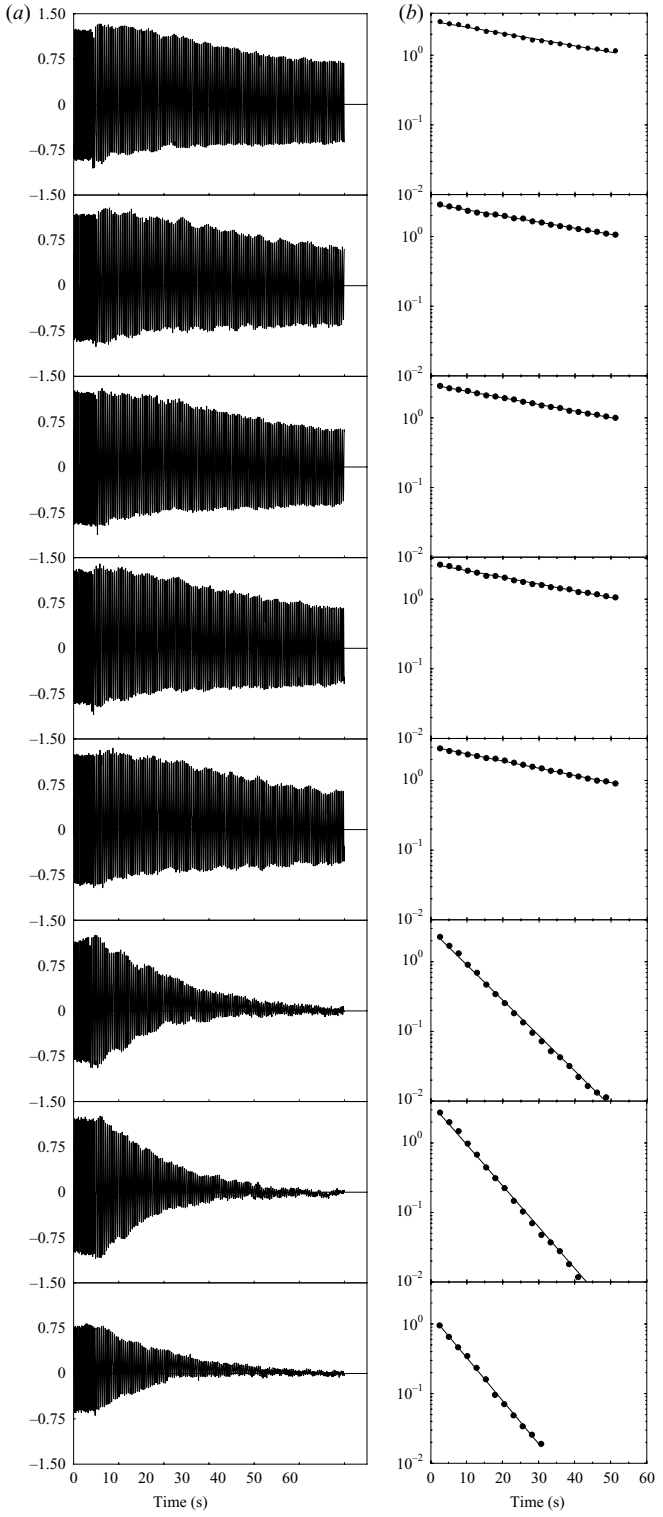

FIGURE 13. Results of experiments listed in column 2 of table 2 at times: 15, 45, 60, 80, 120 min and 1, 2, and 6 days. (*a*) Time series (in volts) of water surface displacement. (*b*) L_2 norm of pieces of the corresponding time series as described in the text.

T_s	T: $2\tilde{a}_w =$ 0.5 cm	E: $2\tilde{a}_w =$ 0.5 cm	E: $2\tilde{a}_w =$ 0.6 cm	E: $2\tilde{a}_w =$ 0.7 cm	E: $2\tilde{a}_w =$ 0.8 cm	E: $2\tilde{a}_w =$ 0.9 cm	E: $2\tilde{a}_w =$ 1.0 cm
15 min	0.086	0.111		0.098	0.107		0.098
20 min			0.082			0.094	
30 min		0.127		0.107	0.107		0.107
45 min	0.082	0.107	0.094	0.107	0.098	0.110	0.098
60 min	0.090		0.098	0.098	0.098	0.110	0.094
80 min	0.090	0.107	0.098			0.107	
90 min				0.094	0.098		0.115
100 min		0.111	0.098				
120 min	0.98	0.143	0.094	0.094	0.098	0.115	0.135
1 day	0.479		0.246	0.176	0.115	0.414	0.418
2 day	0.561		0.209	0.270	0.299	0.275	0.279
5 day			0.598	0.500	0.549	0.549	0.598
6 day	0.586						

TABLE 2. Measured spatial damping rate, $\tilde{\delta}$ (m^{-1}), of the amplitude of a monochromatic wave train at 3.33 Hz for various wave-paddle amplitudes and for various surface ages. The wave-paddle had either a triangular (T) or exponential (E) cross-section.

the theoretical spatial damping rate is $0.316\,\mathrm{m}^{-1}$, almost 3 times greater than the value we measured for $T_s < 2$ h. (*c*) Assume that the surface has elasticity due to surfactants. This model requires knowledge of the surface contamination, so we cannot compute a number here; however, it is known (cf. Henderson 1998) that surface contaminants can cause the damping rate to vary non-monotonically between the clean surface value and about 10 times that value. Perhaps the effects of elasticity are the cause for the damping rate to increase non-monotonically for some experiments during the first few days after surface cleaning. It is also true that in experiments on two-dimensional surface patterns (Hammack *et al.* 2005), the decay rate did not increase monotonically with surface age.

In these experiments and those discussed in § 6, we made an effort to minimize reflections. In particular, we chose test sections and measurement time intervals so that energy travelling at the linearized group velocity of the carrier wave did not have time to reflect and return to the measurement sites. Nevertheless, long wave energy due to wavemaker start-up may have been present in the tank, and the phase velocity of the waves is large enough that some reflected wave motion may have been present in the measurements. The presence of the beach and viscous decay helps these unwanted motions to be small.

Appendix B. Proof of nonlinear stability in two spatial dimensions

For the two-dimensional problem, proof of nonlinear stability of the Stokes solution of (2.5) begins with (4.9), as in the one-dimensional case. Again, we need (4.13) and (4.15), but in two dimensions we also control $\|\partial_y\lambda(t)\|_2^2$. As with (4.15), we find that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\partial_y\lambda\|_2^2\right) \leqslant |\dot{\phi}|\,\|\partial_y\lambda\|_2^2\left\{1+\frac{4\|\lambda\|_\infty^2}{|A|}+\frac{4\|\lambda\|_\infty^2}{|A|^2}\right\}. \tag{B 1}$$

In addition, we must control $\|\partial_x^2\lambda(t)\|_2^2$, $\|\partial_x\partial_y\lambda(t)\|_2^2$ and $\|\partial_y^2\lambda(t)\|_2^2$. Calculations for all these terms are similar, so we derive only one. Just as we obtained (4.14), we find

that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\partial_x\partial_y\lambda\|_2^2\right) \leqslant |\dot{\phi}|\|\partial_x\partial_y\lambda\|_2^2$$
$$+\frac{2}{|A|}|\dot{\phi}|\iint_D [|2u(\partial_x\partial_y u)(\partial_x\partial_y v)+v((\partial_x\partial_y v)^2-(\partial_x\partial_y u)^2)|]\,\mathrm{d}x\,\mathrm{d}y$$
$$+\frac{2}{|A|}|\dot{\phi}|\iint_D [|(3\partial_x u\partial_y u+\partial_x v\partial_y v)\cdot\partial_x\partial_y v-(\partial_x u\partial_y v+\partial_x v\partial_y u)\cdot\partial_x\partial_y u|]\,\mathrm{d}x\,\mathrm{d}y$$
$$+\frac{2}{|A|^2}|\dot{\phi}|\iint_D [|(u^2-v^2)(\partial_x\partial_y u)(\partial_x\partial_y v)+uv\cdot((\partial_x\partial_y v)^2-(\partial_x\partial_y u)^2)|]\,\mathrm{d}x\,\mathrm{d}y$$
$$+\frac{2}{|A|^2}|\dot{\phi}|\iint_D [|(\partial_x u\partial_y v+\partial_x v\partial_y u)\cdot(v\partial_x\partial_y v-u\partial_x\partial_y u)|]\,\mathrm{d}x\,\mathrm{d}y$$
$$+\frac{2}{|A|^2}|\dot{\phi}|\iint_D [|(3\partial_x u\partial_y u+\partial_x v\partial_y v)(u\partial_x\partial_y v)-(3\partial_x v\partial_y v+\partial_x u\partial_y u)(v\partial_x\partial_y u)|]\,\mathrm{d}x\,\mathrm{d}y. \tag{B 2}$$

The following bounds are helpful for the cubic terms:

$$\iint_D |2u(\partial_x\partial_y u)(\partial_x\partial_y v)|\,\mathrm{d}x\mathrm{d}y \leqslant \|\lambda\|_\infty\iint_D |2(\partial_x\partial_y u)(\partial_x\partial_y v)|\,\mathrm{d}x\mathrm{d}y \leqslant \|\lambda\|_\infty\|\partial_x\partial_y\lambda\|_2^2;$$
$$\iint_D |v((\partial_x\partial_y v)^2-(\partial_x\partial_y u)^2)|\,\mathrm{d}x\mathrm{d}y \leqslant \|\lambda\|_\infty\|\partial_x\partial_y\lambda\|_2^2;$$
$$|\partial_x u\partial_y v+\partial_x v\partial_y u| = |(\partial_x u,\,\partial_x v)\cdot(\partial_y v,\,\partial_y u)| \leqslant \sqrt{(\partial_x u)^2+(\partial_x v)^2}\sqrt{(\partial_y u)^2+(\partial_y v)^2},$$

so

$$\iint_D |(\partial_x u\partial_y v+\partial_x v\partial_y u)\partial_x\partial_y u|\,\mathrm{d}x\,\mathrm{d}y \leqslant \iint_D \sqrt{(\partial_x u)^2+(\partial_x v)^2}\sqrt{(\partial_y u)^2+(\partial_y v)^2}|\partial_x\partial_y u|\,\mathrm{d}x\,\mathrm{d}y$$
$$\leqslant \left[\iint_D [(\partial_x u)^2+(\partial_x v)^2][(\partial_y u)^2+(\partial_y v)^2]\,\mathrm{d}x\,\mathrm{d}y\right]^{1/2}\left[\iint_D (\partial_x\partial_y u)^2\,\mathrm{d}x\,\mathrm{d}y\right]^{1/2}$$
$$\leqslant \left[\iint_D [(\partial_x u)^2+(\partial_x v)^2]^2\,\mathrm{d}x\,\mathrm{d}y\right]^{1/4}\left[\iint_D [(\partial_y u)^2+(\partial_y v)^2]^2\,\mathrm{d}x\,\mathrm{d}y\right]^{1/4}\|\partial_x\partial_y\lambda\|_2$$
$$= \|\partial_x\lambda\|_4\|\partial_y\lambda\|_4\|\partial_x\partial_y\lambda\|_2.$$

Similarly,

$$\iint_D |(3\partial_x u\partial_y u+\partial_x v\partial_y v)\partial_x\partial_y v|\,\mathrm{d}x\,\mathrm{d}y \leqslant 3\|\partial_x\lambda\|_4\|\partial_y\lambda\|_4\|\partial_x\partial_y\lambda\|_2.$$

For the quartic terms:

$$\iint_D |(u^2-v^2)(\partial_x\partial_y u)(\partial_x\partial_y v)|\,\mathrm{d}x\,\mathrm{d}y \leqslant \|\lambda\|_\infty^2\|\partial_x\partial_y\lambda\|_2^2,$$
$$\iint_D |uv((\partial_x\partial_y v)^2-(\partial_x\partial_y v)^2)|\,\mathrm{d}x\,\mathrm{d}y \leqslant \|\lambda\|_\infty^2\|\partial_x\partial_y\lambda\|_2^2,$$
$$\iint_D |v(\partial_x\partial_y v)(\partial_x v\partial_y u+\partial_x u\partial_y v)|\,\mathrm{d}x\,\mathrm{d}y \leqslant \|\lambda\|_\infty\|\partial_x\lambda\|_4\|\partial_y\lambda\|_4\|\partial_x\partial_y\lambda\|_2,$$

$$\iint_D |u(\partial_x\partial_y u)(\partial_x v\partial_y u + \partial_x u\partial_y v)|\,\mathrm{d}x\,\mathrm{d}y \leqslant \|\lambda\|_\infty \|\partial_x\lambda\|_4 \|\partial_y\lambda\|_4 \|\partial_x\partial_y\lambda\|_2,$$
$$\iint_D |u(\partial_x\partial_y v)(3\partial_x u\partial_y v + \partial_x v\partial_y u)|\,\mathrm{d}x\,\mathrm{d}y \leqslant 3\|\lambda\|_\infty \|\partial_x\lambda\|_4 \|\partial_y\lambda\|_4 \|\partial_x\partial_y\lambda\|_2,$$
$$\iint_D |v(\partial_x\partial_y u)(3\partial_x v\partial_y v + \partial_x u\partial_y u)|\,\mathrm{d}x\,\mathrm{d}y \leqslant 3\|\lambda\|_\infty \|\partial_x\lambda\|_4 \|\partial_y\lambda\|_4 \|\partial_x\partial_y\lambda\|_2.$$

Using these bounds in (B 2) and regrouping leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\partial_x\partial_y\lambda\|_2^2\right) \leqslant |\dot\phi|\|\partial_x\partial_y\lambda\|_2^2\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]^2 + |\dot\phi|\|\partial_x\partial_y\lambda\|_2\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]\left(\frac{8\|\partial_x\lambda\|_4\|\partial_y\lambda\|_4}{|A|}\right). \quad \text{(B 3)}$$

In this same way, we also obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\left\|\partial_x^2\lambda\right\|_2^2\right) \leqslant |\dot\phi|\left\|\partial_x^2\lambda\right\|_2^2\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]^2 + |\dot\phi|\left\|\partial_x^2\lambda\right\|_2\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]\left(\frac{8\|\partial_x\lambda\|_4^2}{|A|}\right),$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\left\|\partial_y^2\lambda\right\|_2^2\right) \leqslant |\dot\phi|\left\|\partial_y^2\lambda\right\|_2^2\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]^2 + |\dot\phi|\left\|\partial_y^2\lambda\right\|_2\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]\left(\frac{8\|\partial_y\lambda\|_4^2}{|A|}\right).$$

Adding these yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\left\|\partial_x^2\lambda\right\|_2^2 + \|\partial_x\partial_y\lambda\|_2^2 + \left\|\partial_y^2\lambda\right\|_2^2\right) \leqslant |\dot\phi|\left(\left\|\partial_x^2\lambda\right\|_2^2 + \|\partial_x\partial_y\lambda\|_2^2 + \left\|\partial_y^2\lambda\right\|_2^2\right)\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]^2 + \frac{8}{|A|}|\dot\phi|\left[1+\frac{2\|\lambda\|_\infty}{|A|}\right]\left(\left\|\partial_x^2\lambda\right\|_2\|\partial_x\lambda\|_4^2 + \|\partial_x\partial_y\lambda\|_2\|\partial_x\lambda\|_4\|\partial_y\lambda\|_4 + \left\|\partial_y^2\lambda\right\|_2\|\partial_y\lambda\|_4^2\right). \quad \text{(B 4)}$$

Now we define a Sobolev norm different from that in (4.16):

$$\|\lambda\|_{2,2}^2 = \|\lambda\|_2^2 + \|\partial_x\lambda\|_2^2 + \|\partial_y\lambda\|_2^2 + \|\partial_x^2\lambda\|_2^2 + \|\partial_x\partial_y\lambda\|_2^2 + \|\partial_y^2\lambda\|_2^2. \quad \text{(B 5)}$$

Then we invoke more Sobolev inequalities (Adams 1975, p. 97). In two spatial dimensions:

(i) there is a positive constant C_2, independent of λ, such that

$$\|\lambda\|_\infty \leqslant C_2\|\lambda\|_{2,2}; \quad \text{(B 6)}$$

(ii) there is a positive constant C_3, independent of λ, such that

$$\|\partial_x\lambda\|_4 \leqslant C_3\left[\|\partial_x\lambda\|_2^2 + \left\|\partial_x^2\lambda\right\|_2^2 + \|\partial_x\partial_y\lambda\|_2^2\right]^{1/2}, \quad \text{(B 7)}$$

and

$$\|\partial_y\lambda\|_4 \leqslant C_3\left[\|\partial_y\lambda\|_2^2 + \|\partial_x\partial_y\lambda\|_2^2 + \left\|\partial_y^2\lambda\right\|_2^2\right]^{1/2}.$$

Now combine (4.13), (4.15), (B 1) and (B 5), make use of (B 6) and (B 7), to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\lambda\|_{2,2}^2\right) \leqslant |\dot\phi|\|\lambda\|_{2,2}^2\left[1+\frac{2C_2\|\lambda\|_{2,2}}{|A|}\right]^2 + |\dot\phi|\|\lambda\|_{2,2}^2\left[1+\frac{2C_2\|\lambda\|_{2,2}}{|A|}\right]\left(24C_3^2\frac{\|\lambda\|_{2,2}}{|A|}\right). \quad \text{(B 8)}$$

Hence, if there is a constant K, such that $\|\lambda(t)\|_{2,2} \leqslant K < \infty$ for all $t \geqslant 0$, then it follows from (B 8) and Gronwall's inequality that

$$\|\lambda(t)\|_{2,2}^2 \leqslant \|\lambda(0)\|_{2,2}^2 \exp\left\{|\phi(t)\left[1 + \frac{2C_2K}{|A|}\right]\left[1 + \frac{\left(2C_2 + 24C_3^2\right)K}{|A|}\right]\right\}. \tag{B 9}$$

Thus, for all $t \geqslant 0$,

$$\|\lambda(t)\|_{2,2}^2 < \varepsilon,$$

provided at $t = 0$,

$$\|\lambda(0)\|_{1,2}^2 < \Delta \leqslant \varepsilon \exp\left\{-\frac{|\gamma||A|^2}{\delta}\left(1 + \frac{2C_2\sqrt{\varepsilon}}{|A|}\right)\left[1 + \frac{\left(2C_2 + 24C_3^2\right)\sqrt{\varepsilon}}{|A|}\right]\right\}, \tag{B 10}$$

where we have chosen $K \leqslant \sqrt{\varepsilon}$. This completes the proof of theorem 2 in two dimensions: the Stokes solutions of (2.5) and of (2.1) are nonlinearly stable to small perturbations in either one or two dimensions.

REFERENCES

Ablowitz, M. J. & Segur, H. 1981 *Solitons and the Inverse Scattering Transform.* SIAM, Philadelphia, 391 pp.

Adams, R. A. 1975 *Sobolev Spaces.* Academic, 259 pp.

Anderson, D. & Lisak, M. 1984 Modulational instability of coherent optical-fibre transmission signals. *Opt. Lett.* **9**, 468–470.

Benjamin, T. B. 1967 Instability of periodic wavetrains in nonlinear dispersive systems. *Proc. R. Soc.* A **299**, 59–75.

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

Benney, D. J. & Newell, A. C. 1967 The propagation of nonlinear wave envelopes. *Stud. Appl. Maths* **46**, 133–139.

Cherry, T. M. 1926 On periodic solutions of Hamiltonian systems of differential equations. *Phil. Trans. R. Soc. Lond.* A **227**, 137–221.

Coddington, E. A. & Levinson, N. 1955 *Theory of Ordinary Differential Equations.* McGraw-Hill, 429 pp.

Crawford, D. R., Lake, B. M., Saffman, P. G. & Yuen, H. C. 1981 Stability of deep-water waves in two and three dimensions. *J. Fluid Mech.* **105**, 177–191.

Dysthe, K. 1979 Note on a modification to the nonlinear Schrodinger equation for application to deep water waves. *Proc. R. Soc.* A **369**, 105–114.

Feir, J. E. 1967 Discussion: some results from wave pulse experiments. *Proc. R. Soc.* A **299**, 54–58.

Flick, R. E. & Guza, R. T. 1980 Paddle generated waves in laboratory channels. *J. Waterway Port Cosatal Ocean Div. WW1 ASCE* **106**, 79–97.

Gordon, J. P. 1986 Theory of the soliton self-frequency shift. *Optics Lett.* **11**, 662–664.

Guenther, R. B. & Lee, J. W. 1988 *Partial Differential Equations of Mathematical Physics and Integral Equations.* Dover, 557 pp.

Hammack, J. L. & Henderson, D. M. 2003 Experiments on deep-water waves with two-dimensional surface patterns. *J. Offshore Mech. Arctic Engng* **125**, 48–53.

Hammack, J. L. Henderson, D. M. & Segur, H. 2005 Progressive waves with persistent two-dimensional surface patterns in deep water. *J. Fluid Mech.* **532**, 1–51.

Hasegawa, A. 1971 Theory and computer experiment on self-trapping instability of plasma cyclotron waves. *Phys. Fluids* **15**, 870–881.

Hasegawa, A. & Kodama, Y. 1995 *Solitons in Optical Communications.* Clarendon, 320 pp.

Henderson, D. M. 1998 Effects of surfactants on Faraday wave dynamics. *J. Fluid Mech.* **365**, 89–107.

Ince, E. L. 1956 *Ordinary Differential Equations.* Dover, 558 pp.

Karlsson, M. 1995 Modulational instability in lossy optical fibers. *J. Opt. Soc. Am.* B **12**, 2071–2077.

Lake, B. M. & Yuen, H. C. 1977 A note on some water-wave experiments and the comparison of data with theory. *J. Fluid Mech.* **83**, 75–81.

Lake, B. M., Yuen, H. C., Rungaldier, H. & Ferguson, W. E. 1977 Nonlinear deep-water waves: theory and experiment. Part 2. Evolution of a continuous wave train. *J. Fluid Mech.* **83**, 49–74.

Lighthill, M. J. 1965 Contribution to the theory of waves in nonlinear dispersive systems. *J. Inst. Math. Applics.* **1**, 269–306.

Lo, E. & Mei, C. C. 1985 A numerical study of water-wave modulations based on a higher-order nonlinear Schrödinger equation. *J. Fluid Mech.* **150**, 395–416.

Luther, G. G. & McKinstrie, C. J. 1990 Transverse modulational instability of collinear waves. *J Opt. Soc. Am.* B **7**, 1125–1141.

McKinstrie, C. J. & Bingham, R. 1989 The modulational instability of coupled waves. *Phys. Fluids* B **1**, 230–237.

Mei, C. C. & Hancock, M. J. 2003 Weakly nonlinear surface waves over a random seabed. *J. Fluid Mech.* **475**, 247–268.

Melville, W. K. 1982 The instability and breaking of deep-water waves. *J. Fluid Mech.* **115**, 65–185.

Miles, J. W. 1967 Surface-wave damping in closed basins. *Proc. R. Soc.* A **297**, 459–475.

Nemytskii, V. V. & Stepanov, V. V. 1960 *Qualitative Theory of Differential Equations.* Princeton University Press, 519 pp.

Ostrovsky, L. A. 1967 Propagation of wave packets and space-time self-focussing in a nonlinear medium. *Sov. Phys. J. Exp. Theor. Phys.* **24**, 797–800.

Stokes, G. G. 1847 On the theory of oscillatory waves. *Trans. Camb. Phil. Soc.* **8**, 441. Also in *Mathematical and Physical Papers*, vol. 1, pp. 197–229.

Su, Y.-M. 1982 Evolution of groups of gravity waves with moderate to high steepness. *Phys. Fluids* **25**, 2167–2174.

Sulem, C. & Sulem, P. L. 1999 *The Nonlinear Schrodinger Equation*, Springer, 345 pp.

Tai, K., Tomita, A. & Hasegawa, A. 1986 Observation of modulational instability in optical fibres. *Phys. Rev. Lett.* **56**, 135–138.

Tulin, M. P. & Waseda, T. 1999 Laboratory observations of wave group evolution, including breaking effects. *J. Fluid Mech.* **378**, 197–232.

Trulsen, K., Kliakhandler, I., Dysthe, K. B. & Velarde, M. G. 2000 On weakly nonlinear modulations of waves in deep water. *Phys. Fluids*, **12**, 2432–2437.

Trulsen, K. & Stansberg, C. T. 2001 Spatial evolution of water surface waves. *Proc. Eleventh Intl Offshore & Polar Engng Conf.* pp. 71–77.

Whitham, G. B. 1967 Nonlinear dispersion of water waves. *J. Fluid Mech.* **27**, 399–412.

Yuen, H. C. & Lake, B. M. 1975 Nonlinear deep water waves: theory and experiment *Phys. Fluids* **18**, 956–960.

Zakharov, V. E. 1967 Instability of self-focusing of light. *Sov. Phys. J. Exp. Theor. Phys.* **24**, 455–459.

Zakharov, V. E. 1968 Stability of periodic waves of finite amplitude on the surface of a deep fluid. *J. Appl. Mech. Tech. Phys.* **2**, 190–194.