Phil. Trans. R. Soc. Lond. A **302**, 457–510 (1981) [457]
Printed in Great Britain

AN EVALUATION OF A MODEL EQUATION FOR WATER WAVES

BY J. L. BONA,† W. G. PRITCHARD‡ AND L. R. SCOTT§
† *Department of Mathematics, University of Chicago, Chicago, Illinois* 60637, *U.S.A.*
‡ *Department of Mathematics, University of Essex, Colchester, Essex CO4 3SQ, U.K.*
§ *Department of Mathematics, University of Michigan, Ann Arbor, Michigan* 48109, *U.S.A.*

(*Communicated by T. B. Benjamin, F.R.S. – Received* 18 *November* 1980)

CONTENTS

 [Published 24 September 1981

The aim of this paper is to assess how well the partial differential equation

$$\eta_t + \eta_x + \tfrac{3}{2}\eta\eta_x - \mu\eta_{xx} - \tfrac{1}{6}\eta_{xxt} = 0 \qquad (*)$$

describes the propagation of surface water waves in a channel. In (*) the variables are all scaled, with x proportional to the horizontal coordinate along the channel, t proportional to the elapsed time, and η proportional to the vertical displacement of the surface of the water from its equilibrium position. The parameter μ is a non-negative constant.

A numerical scheme has been developed to solve (*) in the domain $\{(x,t): x, t > 0\}$, subject to the initial condition $\eta(x,0) \equiv 0$ and to the boundary condition $\eta(0,t) = h(t)$. The specified function h corresponds to a given displacement of the free surface at one end of the channel. The numerical scheme, which introduces some novel ideas for the approximation of solutions of certain partial differential equations, is explicit, unconditionally stable, and has fourth-order accuracy in both the spatial and temporal variables. The errors inherent in the integration procedure are rigorously analysed, and convergence tests of the computer code are presented.

A comparison is made between the predictions of the theoretical model and the results of laboratory experiments. The experiments, which were performed at a fixed wavelength λ and at a number of wave amplitudes a, covered a range of the parameter $S(= a\lambda^2/d^3)$ of more than two orders of magnitude. Here d is the depth of the undisturbed water. The model was found to give quite a good description of the spatial and temporal development of periodically generated waves over a wide range of the parameter S. It was noteworthy that, at least on laboratory scales, an allowance for dissipative effects was crucial in obtaining good agreement between experimental observations and the predictions of a theoretical model.

At the larger wave amplitudes used in the experiments there were important differences between the forecast of the model and the empirical results. Possible reasons for these discrepancies are discussed.

1. Introduction

This study assesses a particular model for the unidirectional propagation of water waves, comparing its predictions with the results of a set of laboratory experiments. The equation to be tested is a one-dimensional representation of weakly nonlinear, dispersive waves in shallow water. A model for such flows was proposed by Korteweg & de Vries (1895) and this has provided the theoretical basis for a number of laboratory experiments. Some recent studies that have been made in the area are those of Zabusky & Galvin (1971), Hammack (1973) and Hammack &

Segur (1974). In each case the theoretical model gave a good qualitative account of the experiments, but the quantitative comparisons were not very extensive. One of the purposes of this paper is to provide a more detailed quantitative assessment of a particular model than has been given to date.

An important aspect of the formulation of the theoretical model is the specification of the initial conditions and the boundary conditions for the equation. Zabusky & Galvin (1971) considered an initial-value problem having spatial periodicity, whereas Hammack (1973) and Hammack & Segur (1974) considered an initial-value problem posed on the real line. In contrast, we shall consider an initial-value problem posed on the half line with boundary data specified at the origin. This problem was chosen to correspond with an experiment in which waves were generated at one end of a long channel, and obviates certain difficulties inherent in the other formulations.

Approximate solutions to the mathematical problem thus posed have been found numerically. The method we have developed (see § 4) is an unconditionally stable, explicit scheme having fourth-order accuracy in both the spatial and the temporal variables. It is based on the discretization of an integral representation of the solution of the equation, a method which, although commonly used for ordinary differential equations, is unusual for a partial differential equation. An analysis of the scheme is presented in § 5, where we give both *a priori* error bounds and explicit *a posteriori* error estimates based on the observed maximum value of the discrete solution.

In making the experiments we have covered a range of conditions for which it was expected that the nonlinear effects would be relatively unimportant, that the nonlinear and dispersive effects would be approximately balanced and that the dispersive effects would be less significant than the nonlinear effects. One feature to emerge from the study was that, in all cases, it was important to make an allowance for dissipative effects to obtain reasonable agreement between the empirical results and the theoretical model. But, with this proviso, the model appeared to give a good description of the experimental results at the smaller wave amplitudes. At the larger amplitudes the agreement was not so good, and possible reasons for the discrepancies are discussed (see §§ 7.4, 7.5 and 7.6).

The structure of the paper is as follows. In § 2 we first discuss model equations apposite to the present study, together with the concomitant assumptions underlying their derivation. Then, in view of these constraints, we examine some of the criteria involved in the design of the laboratory experiment. Some theoretical properties of the solutions of the differential equation to be tested are given in § 3. These are used in the derivation of the error estimates for the numerical scheme, the scheme being described and analysed in §§ 4 and 5. Convergence tests made with the program are also given. In § 6 the experimental procedure is described, and in § 7 the main results are presented. A *résumé* of the results is given in § 8.

We shall use the abbreviation KdV to refer to the Korteweg–de Vries equation, as defined by (2.1). An alternative model for the same physical situation, the model that will form the focus of this study, will be referred to as (M) and is defined in equation (2.3). Two variants of this basic model will also be considered. One is (M*) (see equation (2.6)) and the other is referred to as (M†) (defined by (7.1)). The dispersion relations for the various models are referred to throughout the paper. In this context the term 'exact' refers to the dispersion relation for the Euler equation with the linearized form of the boundary condition at the free surface (see (2.4))

2. Experimental design

2.1. *Model equations*

Consider two-dimensional surface waves propagating along a uniform horizontal channel. Suppose that the waves propagate only in the positive x-direction and that the undisturbed depth of the liquid in the channel is d. All the variables used here are dimensionless, with the length scale taken to be the equilibrium depth d and the time scale $(d/g)^{\frac{1}{2}}$, where g is the acceleration due to gravity. Let t be time, let x be the horizontal coordinate and let $\eta = \eta(x,t)$ represent the vertical displacement of the surface of the liquid from its equilibrium position. If the horizontal scale δ^{-1} of the motions is large and the maximum amplitude ϵ of the waves is sufficiently small, then a model for the propagation of irrotational waves is afforded by the KdV equation (see Whitham 1974)

$$\eta_t + \eta_x + \tfrac{3}{2}\eta\eta_x + \tfrac{1}{6}\eta_{xxx} = 0. \qquad (2.1)\ (\text{KdV})$$

The primary terms η_t and η_x represent a uniform translation of a wave and it is proposed that the secondary terms $\frac{3}{2}\eta\eta_x$ and $\frac{1}{6}\eta_{xxx}$ account respectively for the modification of the wave through the separate influences of nonlinear and dispersive effects. The relative importance of the nonlinear and the dispersive effects is given by the parameter $S = \epsilon\delta^{-2}$, and an important assumption in the derivation of KdV is that this parameter is $O(1)$ (cf. Meyer 1972;† Whitham 1974). (Here, and in what follows, the symbol $O(\cdot)$ will be used informally in the way that is common in the construction and formal analysis of model equations for physical phenomena. We shall always be concerned in principle with the limits $\epsilon \downarrow 0$ and $\delta \downarrow 0$ though, in fact, finite but small values of these parameters will be in question. Thus, $S = O(1)$ means that, as $\epsilon \downarrow 0$ and $\delta \downarrow 0$, $S = \epsilon\delta^{-2}$ takes values that are neither very large nor very small.) Note that, for waves whose length scale is λ and whose amplitude is a, $S = a\lambda^2/d^3$.

The above considerations suggest the introduction of a new dependent variable N and new independent variables ξ, τ such that

$$\eta = \epsilon N, \quad x = \epsilon^{-\frac{1}{2}}\xi, \quad t = \epsilon^{-\frac{1}{2}}\tau.$$

Thus, by assumption, N and its derivatives with respect to the new independent variables are all $O(1)$, and it follows that KdV can be written as

$$N_\tau + N_\xi + \tfrac{3}{2}\epsilon N N_\xi + \tfrac{1}{6}\epsilon N_{\xi\xi\xi} = O(\epsilon^2), \qquad (2.2)$$

showing explicitly the relative sizes of the various terms. (For convenience in the present discussion, S has been taken to be 1.) On the right-hand side of (2.2) we have indicated the relative size of the terms neglected in the formal derivation of the KdV model. A physical interpretation of (2.2) is that the small nonlinear and dispersive corrections can accumulate and, on time scales τ of $O(\epsilon^{-1})$ (or $t = O(\epsilon^{-\frac{3}{2}})$), make important modifications to the initial waveform. Moreover, since the terms neglected in (2.2) are $O(\epsilon^2)$, it follows that, on time scales $\tau = O(\epsilon^{-2})$ (or $t = O(\epsilon^{-\frac{5}{2}})$), the model can no longer be formally justified.

Because of the orders of magnitude of the terms in (2.2) an alternative model for the same physical situation, valid to the same accuracy as the KdV equation, is the equation (see Peregrine 1966, Benjamin *et al.* 1972)

$$N_\tau + N_\xi + \tfrac{3}{2}\epsilon N N_\xi - \tfrac{1}{6}\epsilon N_{\xi\xi\tau} = O(\epsilon^2).$$

In terms of the physical variables $\eta(x,t)$ this model takes the form

$$\eta_t + \eta_x + \tfrac{3}{2}\eta\eta_x - \tfrac{1}{6}\eta_{xxt} = 0. \qquad (2.3)\ (\text{M})$$

† *Note added in proof*: See also Meyer, R. E. 1979 *Bull. Cal. Math. Soc.* **71**, 121.

Thus, in summary, (KdV) and (M) have been proposed as models for the propagation of water waves under the following conditions.

(i) The waves effectively propagate in one direction. This precludes the possibility of interactions with reflected waves; in practical terms it means that any variations in the depth of the channel should be small or should occur on length scales very much larger than the horizontal scale of the waves.

(ii) The wave amplitudes are small (i.e. $\epsilon \ll 1$) and the horizontal length scale of the waves is large (i.e. $\delta \ll 1$).

(iii) The nonlinear and dispersive effects are comparable: $\epsilon\delta^{-2} = O(1)$.

(iv) The waves arise on an irrotational flow.

(v) There is no mechanical degradation of energy.

(vi) The influence of surface tension is negligible (though this restriction can be relaxed, cf. Korteweg & de Vries 1895).

We can expect significant modifications to a waveform on a time scale $O(\epsilon^{-\frac{3}{2}})$ and, from a formal viewpoint, the model cannot be justified on times which are $O(\epsilon^{-\frac{5}{2}})$.

2.2. *Previous studies*

In 1971 Zabusky & Galvin reported some experiments in which a train of initially sinusoidal waves propagated into still water. At stations further along the channel they found that, after the first few wave crests had passed, the wave profiles were very nearly periodic in time. This property suggested a numerical experiment in which a periodic version of KdV was integrated, with a sinusoidal waveform as the initial datum. Then, to compare the numerical computations with the experiments, the long-wave speed for linear disturbances was used as the basis for a transformation from time in the periodic problem to position in the experimental configuration. The experiments were made at values of S of 22, 95 and 153.† Fairly good qualitative agreement was obtained between the predicted wave shapes and those observed experimentally, but quantitative comparisons were not made, principally because viscous effects had a significant influence on the experimental results.

A study similar in concept to the programme to be described here was made by Hammack (1973). Water was displaced at one end of a channel generating an isolated waveform, the passage of which was observed at various positions along the channel. Comparisons made between the observed profiles and numerical solutions of (M) showed good qualitative agreement but, since the computations were not very accurate and since viscous effects were again important, detailed quantitative comparisons were not made. For these experiments the value of S lay between about 1 and 10.

In a subsequent experiment Hammack & Segur (1974) also followed the evolution of an isolated waveform propagating along a channel. Using the inverse-scattering methods developed for the KdV equation, they predicted both the number of solitons to emerge from the initial waveform and the amplitude of the largest soliton. The predicted number of emergent solitons was in agreement with the experimental observations, but the predicted amplitude of the leading soliton (after making a correction for viscous damping along the lines suggested by Keulegan 1948) differed by about 15–20 % from the observed values. These experiments were made at values of S ranging between 50 and 600.

† *Note added in proof*: These numbers correct the values given in colums 2, 3 and 5 of table 1 of Zabusky & Galvin (1971).

In each of the above studies, the theoretical predictions were obtained by solving a pure initial-value problem. However, the experimental data were not obtained in the form required for the theoretical model, which necessitated a transformation of the data set. Because the transformation used was inexact this may have led to significant errors in the solution. This issue is discussed in Appendix A.

2.3. *Allowance for dissipation*

One of the main conclusions to be drawn from the previous experimental studies is that useful quantitative predictions can be made only by taking account of dissipative effects. On the scale of the present experiment the main sources of wave damping appear to derive from viscous dissipation in the boundary layers on the sides and bottom of the channel, from the influence of the meniscus at the side walls of the channel and perhaps from damping at the free surface (see Barnard *et al.* 1977; Mahony & Pritchard 1980). It is possible to incorporate the effects of the boundary layers on the walls and bottom of the channel into the theories described above (see Kakutani & Matsuuchi 1975), but there are empirical difficulties in determining the properties of the liquid surface and theoretical uncertainties about the representation of the effects at the free surface and at the meniscus (see, for example, Miles 1967, Mei & Liu 1973). Thus, any attempts to quantify the dissipative effects must, to a certain extent, be guesswork.

A rationale behind the construction of models such as those described in §2.1 is that the various corrections to the primary terms in the equation can be calculated independently, with a composite model formed by including the modifications additively on the assumption that the coupling between them is negligible. Because of this it is sufficient, for the time being, to consider the effects of damping only on waves of extremely small amplitude, so that a linear model is applicable. Then the dispersion relation between the frequency ω and the wavenumber k is given by (cf. Introduction)

$$\omega = k(1-\tfrac{1}{6}k^2)\ (\mathrm{KdV}), \quad \omega = k(1+\tfrac{1}{6}k^2)^{-1}(\mathrm{M}), \quad \omega = (k\tanh k)^{\frac{1}{2}}\ (\text{exact}). \tag{2.4}$$

By construction the phase speeds ω/k for each of these relations are different only at the fourth order in k.

The theory of Kakutani & Matsuuchi (1975) indicates that the effect of dissipation in the boundary layers on the rigid surfaces of the channel is comparable with the nonlinear and the dispersive corrections from the inviscid theories when the wavenumber k is $O(R^{-\frac{1}{5}})$. Here the Reynolds number R is $(gd^3)^{\frac{1}{2}}/\nu$, where ν is the kinematic viscosity of the fluid. Under these conditions Kakutani & Matsuuchi showed that the dispersion relation for (KdV) should be modified to

$$\omega = k(1-\tfrac{1}{6}k^2) + \rho|k|^{\frac{1}{2}}, \tag{2.5}$$

where ρ is a complex number depending on R. Thus, not only do the boundary layers induce a damping of the waves but they also affect the phase speed slightly. Moreover, the analysis indicates that the boundary-layer damping can be neglected only when $(|k|/8R)^{\frac{1}{2}} \ll \frac{1}{6}|k|^3$, which we could, for example, interpret in the sense that the relative importance of these terms be $O(\delta^2)$. For $\delta = \frac{1}{12}$ this would suggest, as a rough guide, that a depth of a metre or more would be needed to make dissipative effects negligible. (Note that Zabusky & Galvin 1971, using $d \approx 15$ cm and $\delta \approx \frac{1}{16}$, concluded that dissipation was important in their experiment.) The damping introduced in (2.5) can, of course, also be incorporated into model (M). A term of this kind introduces a pseudo-differential operator into each of the model equations.

However, as indicated above, the boundary-layer theory considerably underestimates the damping rate (by about 40 % on the scale of the present experiment, according to Mahony &

Pritchard 1980). Because of the inadequacy of the theory in this respect we decided to use an *ad hoc* representation of the wave damping to preserve the simple structure of the model equation, rather than attempting a more complicated representation that could not be totally justified. Thus, we shall suppose that a wave of wavenumber k is damped at a rate proportional to k^2, which has the effect of introducing a term $-\mu\eta_{xx}$, $\mu \geqslant 0$, into the model equation. This can easily be incorporated into the numerical scheme of §4.

For the experiments to be described the waves were generated by a forced motion at a frequency ω_0, with the result that most of the energy should have resided in a single wavenumber k_0, say. Then, by choosing μ such that the damping of waves of wavenumber k_0 agreed with the experimental decay rate at very small amplitudes, we should at least have modelled correctly the dissipation of the fundamental wave, even if other wavenumbers are likely to have been dissipated at an incorrect rate.

These considerations indicate that we need to be circumspect about the representation of the dissipative effects, a point we shall consider in more detail in §7.5. However, for now let us take the model equation in the form

$$\eta_t + \eta_x + \tfrac{3}{2}\eta\eta_x - \mu\eta_{xx} - \tfrac{1}{6}\eta_{xxt} = 0. \qquad (2.6)\ (\mathrm{M}^*)$$

2.4. *Mathematical considerations*

Three kinds of mathematical problems have been studied in connection with (KdV) or (M):

(i) *Pure initial-value problems.* For this class of problem it is supposed that the surface profile is known at some instant, say $t = 0$. Mathematically this amounts to the specification

$$\eta(x, 0) = g(x), \quad \text{for} \quad x \in \mathbb{R}, \qquad (2.7)$$

where $\mathbb{R}$ denotes the set of all real numbers. Interest is focused on the solution of (KdV) or (M), defined for $t \geqslant 0$, that agrees with g at $t = 0$. If g is an element of a function class comprised of smooth functions that decay to zero sufficiently rapidly at $\pm\infty$, then it is known that the specification (2.7) constitutes a well posed problem in conjunction with (M) (see, for example, Benjamin *et al.* 1972) or in conjunction with (KdV) (see, for example, Bona & Smith 1975).

A physical realization of this formulation of the problem can be achieved in a long channel by establishing a wavetrain of restricted spatial extent that propagates from one end of the channel to the other. A photograph of the water surface at some instant could be used to determine the initial datum g and the wave profile at later times could be compared with, say, numerical solutions to the model problem. (This, in essence, is the kind of programme carried out by Hammack & Segur 1974. However their determination of $g(x)$ was made from a temporal wave record $g(x_0, t)$, x_0 fixed, together with the leading-order approximation $\eta_t + \eta_x = 0$ for the wavefield. It is shown in Appendix A that such a procedure can lead to significant errors.)

(ii) *Periodic initial-value problems.* These problems are the same as described in (i) except that the initial datum, g, is a given periodic function. Again, the mathematical problems for (KdV) and for (M) are well posed. However, the physical realization of such a model is very difficult to achieve. (Zabusky & Galvin 1971 used numerical solutions to a problem of this kind to explain qualitatively the behaviour of waves generated by the periodic motion of a wavemaker at one end of the channel, cf. §2.2.)

(iii) *Initial- and boundary-value problems.* For this class of problem we are interested in solutions $\eta(x, t)$ for $x, t > 0$ to the model equations, subject to the conditions

$$\eta(x, 0) = g(x), \quad x \geqslant 0, \quad \text{and} \quad \eta(0, t) = h(t), \quad t \geqslant 0. \qquad (2.8)$$

For compatibility we suppose that $g(0) = h(0)$. It has been shown by Bona & Bryant (1973) that, under these conditions, (M) constitutes a well posed problem if g and h are suitably smooth functions. A similar result has also recently been proved for (KdV) by Bona & Winther (1981).

Physically g represents the initial configuration of the water surface; usually we would expect at the outset the water to be undisturbed, in which case we would have $g = 0$. The function $h(t)$ represents an imposed displacement of the water surface at the left-hand end of the channel. Thus, we might think of this problem as a model for waves with known amplitude initiated at one end of a long channel.

2.5. *Practical considerations*

The issues raised in the preceding discussion impose considerable restrictions on the experimental design. If the models are to be of any real practical value they should be applicable to the kind of situation that usually obtains in the laboratory, namely the propagation of waves arising from the forcing effects of a wavemaker at one end of a channel. Since wavemakers are usually driven in a periodic motion, it would be nice to allow this feature in the model. Indeed, such forcing would be desirable here because the imposed frequency effectively establishes a length scale for the motions, allowing a fairly precise specification of the parameter S. To meet these requirements and to simplify the experimental procedure, it would appear that the most suitable kind of mathematical problem to model is the initial- and boundary-value formulation. (One of the main empirical difficulties in modelling the pure initial-value formulation is that of obtaining an instantaneous spatial measurement of the wavefield. Also the wave tank available to us would not have been long enough for such an experiment.) A convenient experimental procedure would be to start with the channel free of motion and then to set the wavemaker working at a fixed frequency and amplitude. This would initiate a train of waves that would propagate along the channel, retaining their unidirectional quality until they reached the end of the channel, when the experiment would have to cease. The boundary condition $h(t)$ in (2.8) could be specified by a temporal record of the wave amplitude (taken at a position far enough away from the wavemaker to avoid confusing the free waves with the parasitic field localized near the paddle and to ensure smooth changes in amplitude at the front of the wavetrain).

The wave tank available in our laboratory was only $5\frac{1}{2}$ m long. So, to allow enough time for the waves to show significant modifications before reaching the end of the tank, the basic wavelength had not to be too large. On the other hand, it had to be larger than the channel width (30 cm) to avoid spontaneous generation of transverse modes (cf. §7.2). A reasonable compromise for the wavelength appeared to be 36 cm. We decided to use a wavelength:depth ratio of 12:1.

In principle we should like the experiment to cover a range of wave amplitudes for which the parameter S, measuring the relative importance of nonlinear and dispersive effects, spans a fairly representative range of parameter space. Under the above conditions, S would take a value of 0.1 at a wave amplitude of 0.002 cm and would be 10 at a wave amplitude of 0.2 cm. So, to be sure of achieving linear motions at one end of the parameter range, it would be necessary to use *very* small wave amplitudes. It is fortunate that, in our experiments, this did not pose any major difficulties.

3. Properties of the exact solution of the model equation

In this section we study properties of the solution of the initial- and boundary-value problem

$$\eta_t+\alpha\eta_x+\beta\eta\eta_x-\mu\eta_{xx}-\gamma\eta_{xxt}=0, \quad \text{for} \quad x,t\geqslant 0, \tag{3.1a}$$

$$\eta(x,0)=0 \quad \text{for} \quad x\geqslant 0, \quad \text{and} \quad \eta(0,t)=h(t) \quad \text{for} \quad t\geqslant 0, \tag{3.1b}$$

where α, β, μ are non-negative constants and γ is a positive constant. We shall first discuss the questions of existence, uniqueness and *a priori* boundedness of η. Then, in preparation for *a posteriori* error estimates to be derived in § 5, bounds for derivatives of η are given in terms of assumed bounds on η. Finally, it is shown that η decays exponentially in space, which justifies the truncation of the spatial domain in numerical calculations.

3.1. *Existence, uniqueness and a priori bounds for* η

Suppose that the boundary data are 'smooth' in the sense that, for a given $T>0$ and an integer $l\geqslant 1$,

$$h\in\mathscr{C}^l([0,T]) \quad \text{and} \quad h(0)=0. \tag{3.2}$$

Then, by using the techniques of Bona & Bryant (1973), it follows that (3.1) has a unique solution $\eta\in\mathscr{C}_T^{l,k}$; that is, $(\partial/\partial x)^j(\partial/\partial t)^i\eta(x,t)$ exists and is bounded and continuous on $[0,\infty[\times[0,T]$ for $i=0,1,\ldots,l$ and $j=0,1,\ldots,k$. (Here k may be any positive integer.) Furthermore, these derivatives of η all tend to zero as $x\to\infty$, and η, η_x are square integrable in x on $[0,\infty[$. If $|h(t)|$ and $|h'(t)|$ are bounded by some constant, say M, for $t\in[0,T]$, then by using the methods of Bona & Bryant (1973) it can be shown, for $t\in[0,T]$, that

$$\max\{|\eta(x,s)|: x\geqslant 0,\ s\in[0,t]\}\leqslant b_1 e^{b_2 t}, \tag{3.3}$$

where b_1, b_2 are constants depending only on α, β, μ, γ and M. In addition it follows that the solution to (3.1) satisfies the equation

$$\begin{aligned}\eta_t(x,t)&=h'(t)e^{-x/\sqrt{\gamma}}+\int_0^\infty \tilde{K}(x,y)\,(\alpha\eta+\tfrac{1}{2}\beta\eta^2)(y,t)\,dy\\&\quad+\frac{\mu}{\gamma}[h(t)e^{-x/\sqrt{\gamma}}-\eta(x,t)]-\mu\int_0^\infty \tilde{H}(x,y)\,\eta(y,t)\,dy,\end{aligned} \tag{3.4}$$

where

$$\tilde{K}(x,y)=\frac{1}{2\gamma}[e^{-(x+y)/\sqrt{\gamma}}+\operatorname{sgn}(x-y)\,e^{-|x-y|/\sqrt{\gamma}}] \tag{3.5}$$

and

$$\tilde{H}(x,y)=\frac{1}{2\gamma^{\frac{3}{2}}}[e^{-(x+y)/\sqrt{\gamma}}-e^{-|x-y|/\sqrt{\gamma}}].$$

The numerical scheme to be described in § 4 is based on this formulation.

From the definitions it follows, for any non-negative integer k, that

$$\begin{aligned}&\max\left\{\int_0^x\left|\left(\frac{\partial}{\partial y}\right)^k\tilde{K}(x,y)\right|dy,\int_x^\infty\left|\left(\frac{\partial}{\partial y}\right)^k\tilde{K}(x,y)\right|dy: x>0\right\}\leqslant\gamma^{-\frac{1}{2}(k+1)},\\&\max\left\{\int_0^x\left|\left(\frac{\partial}{\partial y}\right)^k\tilde{H}(x,y)\right|dy,\int_x^\infty\left|\left(\frac{\partial}{\partial y}\right)^k\tilde{H}(x,y)\right|dy: x>0\right\}\leqslant\gamma^{-\frac{1}{2}(k+2)}.\end{aligned} \tag{3.6}$$

Remarks. (i) The *a priori* bound (3.3) can be improved considerably. For example, we have been able to show that $\max\{|\eta|\}$ grows no faster than t. However, (3.3) is sufficient to obtain *a posteriori* estimates (see § 5), which show that there is essentially no growth in the maximum of $|\eta|$, provided the same holds true for a discrete approximation to η.

(ii) The above theory holds when (3.1) is posed with non-zero initial data $\eta(x,0) = g(x)$ provided that $g \in \mathscr{C}^k([0,\infty[)$ for $k \geqslant 2$, that g and its derivatives tend to zero as $x \to \infty$, that g, g' are square integrable and that $g(0) = h(0)$.

3.2. *Bounds for the derivatives of* η

Bounds on the temporal and spatial derivatives of η are to be derived in terms of assumed bounds on the maximum of $|\eta|$ itself. Thus for $T > 0$ define

$$\sigma(T) = \max\{|\eta(x,t)| : x \geqslant 0,\ t \in [0,T]\}. \tag{3.7}$$

We shall use the notation

$$\|\phi\| \equiv \max\{|\phi(x)| : x \geqslant 0\}, \tag{3.8}$$

where ϕ represents η or its derivatives.

Bounds for η_t can be obtained directly from (3.4) through an application of Hölder's inequality (together with (3.6)). These imply that

$$\|\eta_t(\cdot,t)\| \leqslant h_m^{(1)}(t) + \gamma^{-\frac{1}{2}}[\alpha\sigma(t) + \tfrac{1}{2}\beta\sigma^2(t)] + 3(\mu/\gamma)\,\sigma(t), \tag{3.9}$$

where

$$h_m^{(k)}(t) \equiv \max\{|h^{(k)}(s)| : 0 \leqslant s \leqslant t\}, \tag{3.10}$$

with $t > 0$ and k a non-negative integer. (Note that $|h(t)| \leqslant \sigma(t)$.)

Bounds for the spatial derivatives may also be deduced from (3.4). By dividing the range of integration in (3.4) at $y = x$ and then differentiating with respect to x we have that

$$\begin{aligned}\eta_{xt}(x,t) = {} & -\gamma^{-\frac{1}{2}} h'(t)\, e^{-x/\sqrt{\gamma}} + \int_0^\infty \tilde{K}_x(x,y)\,(\alpha\eta + \tfrac{1}{2}\beta\eta^2)(y,t)\,dy \\ & -\gamma^{-1}(\alpha\eta + \tfrac{1}{2}\beta\eta^2)(x,t) - \mu\gamma^{-\frac{3}{2}} h(t)\, e^{-x/\sqrt{\gamma}} - \frac{\mu}{\gamma}\eta_x(x,t) - \mu\int_0^\infty \tilde{H}_x(x,y)\,\eta(y,t)\,dy. \end{aligned} \tag{3.11}$$

Multiplying this equation by $\eta_x(x,t)$ and using Hölder's inequality together with (3.6), it follows that

$$\tfrac{1}{2}(\partial/\partial t)[\eta_x^2(x,t)] + (\mu/\gamma)\,\eta_x^2(x,t) \leqslant M|\eta_x(x,t)|, \tag{3.12}$$

where

$$M \equiv \gamma^{-\frac{1}{2}} h_m^{(1)}(t) + 2\gamma^{-1}(\alpha\sigma(t) + \tfrac{1}{2}\beta\sigma^2(t)) + 2\mu\gamma^{-\frac{3}{2}}\sigma(t).$$

Gronwall's lemma implies that

$$|\eta_x(x,t)| \leqslant (M\gamma/\mu)\,(1 - e^{-\mu t/\gamma}) \equiv P_1(h_m^{(1)}(t), \sigma(t), t).$$

Thus, since $x \geqslant 0$ was arbitrary,

$$\|\eta_x(\cdot,t)\| \leqslant P_1(h_m^{(1)}(t),\ \sigma(t),\ t). \tag{3.13}$$

Note that $P_1(h_m^{(1)}(t), \sigma(t), t) \leqslant M \min\{t, \gamma/\mu\}$, so that P_1 is bounded by a polynomial linear in t and $h_m^{(1)}(t)$, and quadratic in $\sigma(t)$, having coefficients that are polynomials in α, β, μ and $\gamma^{-\frac{1}{2}}$. Moreover, the (explicit) dependence on t can be ignored for $t \geqslant \gamma/\mu$.

Bounds for higher-order spatial derivatives can be obtained inductively by similar arguments, leading to the following lemma.

LEMMA 3.1. *Let $T > 0$. Suppose that $h \in \mathscr{C}^1([0,T])$ and that $h(0) = 0$. Let η be the solution to* (3.1), *and let $h_m^{(1)}$ and σ be defined by* (3.10) *and* (3.7) *respectively. Then, for any positive integer k and for $t \in [0,T]$,*

$$\left\| \left(\frac{\partial}{\partial x}\right)^k \eta(\cdot,t) \right\| \leqslant P_k(h_m^{(1)}(t),\ \sigma(t),\ t),$$

where P_k can be bounded by a polynomial of degree k in $h_m^{(1)}(t)$ and $\min\{t, \gamma/\mu\}$, and of degree $k+1$ in $\sigma(t)$, having coefficients that are polynomials in α, β, μ and $\gamma^{-\frac{1}{2}}$.

Comment. A polynomial $P(x_1, \ldots, x_n)$ is said to have degree l_j in the variable x_j if P is a polynomial of degree at most l_j in the variable x_j, when all the other variables are held fixed.

3.3. *Decay rates for the exact solution*

LEMMA 3.2. *Let $T > 0$ and suppose that $h \in \mathscr{C}^1([0, T])$, with $h(0) = 0$. Let η be the solution to (3.1) and let $h_m^{(j)}, j = 0, 1$, and σ be defined by (3.10) and (3.7) respectively. Then, for any real number $r \in]0, \gamma^{-\frac{1}{2}}[$ there is a function $C = C(\sigma(t)) < \infty$ such that*

$$|\eta(x,t)| \leqslant [(\mu/\gamma)\, h_m^{(0)}(t) + h_m^{(1)}(t)]\, C^{-1}\, e^{Ct - rx}$$

for $t \in [0, T]$. Here $C(\xi) = (a/\gamma)\,[(\alpha + \frac{1}{2}\beta\xi) + \mu/\gamma^{\frac{1}{2}}]$, where $a = 2\gamma^{\frac{1}{2}}/(1 - \gamma r^2)$.

Remarks. (i) This estimate says, in effect, that signals propagating in accordance with (3.1) have a speed not exceeding C/r. When $\mu = 0$, the speed C/r is minimized when $r = (3\gamma)^{-\frac{1}{2}}$, with $C/r = 3\sqrt{3}(\alpha + \frac{1}{2}\beta\sigma)$.

(ii) Similar results also apply when (3.1) is posed with non-zero initial data $\eta(x, 0) = g(x)$, provided $g(0) = h(0)$, g' is square integrable and $|g(x)| \leqslant M e^{-r}$ for $x \geqslant 0$. The estimate is then modified by the addition of the term $M e^{Ct - rx}$.

Proof. Let $X > 0$ and define a weighting function w such that

$$w(x) \equiv \begin{cases} e^{rx}, & 0 \leqslant x \leqslant X, \\ e^{rX}, & x \geqslant X. \end{cases}$$

Set $v(x,t) = w(x)\,\eta(x,t)$ and multiply (3.4) by w. It follows that

$$\begin{aligned} v_t(x,t) &= h'(t)\, e^{-x/\sqrt{\gamma}}\, w(x) + \int_0^\infty \tilde{K}(x,y)\, \frac{w(x)}{w(y)}\, (\alpha + \tfrac{1}{2}\beta\eta(y,t))\, v(y,t)\, dy \\ &\quad + \frac{\mu}{\gamma}[h(t)\, e^{-x/\sqrt{\gamma}}\, w(x) - v(x,t)] - \mu \int_0^\infty \tilde{H}(x,y)\, \frac{w(x)}{w(y)}\, v(y,t)\, dy. \end{aligned} \tag{3.14}$$

When r is in the interval $]0, \gamma^{-\frac{1}{2}}[$, $e^{-x/\sqrt{\gamma}}\, w(x) \leqslant 1$ for any $x \geqslant 0$. Therefore, after multiplying (3.14) by $v(x,t)$ and applying the Hölder inequality, we have that

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial t} v^2(x,t) + \frac{\mu}{\gamma} v^2(x,t) &\leqslant \left\{ |h'(t)| + \frac{\mu}{\gamma}|h(t)| + (\alpha + \tfrac{1}{2}\beta\sigma(t))\, \|v(\cdot,t)\| \int_0^\infty |\tilde{K}(x,y)|\, \frac{w(x)}{w(y)}\, dy \right. \\ &\quad \left. + \mu \|v(\cdot,t)\| \int_0^\infty |\tilde{H}(x,y)|\, \frac{w(x)}{w(y)}\, dy \right\} |v(x,t)|. \end{aligned} \tag{3.15}$$

(Note that $\|v(\cdot,t)\| \leqslant \sigma(t)\, e^{rX} < \infty$ for $t \in [0, T]$.)

From the definitions (3.5) it follows that

$$|\tilde{K}(x,y)| \leqslant \gamma^{-1} e^{-|x-y|/\sqrt{\gamma}} \quad \text{and} \quad |\tilde{H}(x,y)| \leqslant \gamma^{-\frac{3}{2}} e^{-|x-y|/\sqrt{\gamma}}.$$

Moreover, $w(x)/w(y) \leqslant e^{(x-y)r}$ when $y \leqslant x$, and $w(x)/w(y) \leqslant 1$ when $y \geqslant x$, so that

$$\int_0^\infty e^{-|x-y|/\sqrt{\gamma}}\, \frac{w(x)}{w(y)}\, dy \leqslant (\gamma^{-\frac{1}{2}} - r)^{-1} + \gamma^{\frac{1}{2}} \equiv a_0. \tag{3.16}$$

Using these inequalities in (3.15), we see that

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial t} v^2(x,t) + \frac{\mu}{\gamma} v^2(x,t) &\leqslant \left\{ |h'(t)| + \frac{\mu}{\gamma}|h(t)| + a_0 \left[(\alpha + \tfrac{1}{2}\beta\sigma(t))/\gamma + \frac{\mu}{\gamma^{\frac{3}{2}}} \right] \|v(\cdot,t)\| \right\} |v(x,t)|, \\ &\leqslant \left\{ \frac{\mu}{\gamma}\, h_m^{(0)}(t) + h_m^{(1)}(t) + \tilde{c}(t)\, \|v(\cdot,t)\| \right\} |v(x,t)|, \end{aligned}$$

where $\tilde{c}(t) = a_0[(\alpha + \frac{1}{2}\beta\sigma(t))/\gamma + \mu/\gamma^{\frac{3}{2}}]$.

Write $S(t) \equiv \max\{\|v(\cdot, s)\| : 0 \leqslant s \leqslant t\}$. Then Gronwall's lemma implies, for any $t_0 < t \leqslant T$, that

$$|v(x,t)| \leqslant \left(\frac{\mu}{\gamma} h_m^{(0)}(t) + h_m^{(1)}(t) + \tilde{c}(t)\, S(t)\right) \frac{\gamma}{\mu} (1 - \mathrm{e}^{-\mu(t-t_0)/\gamma}) + |v(x,t_0)|\, \mathrm{e}^{-\mu(t-t_0)/\gamma}.$$

But, since x is an arbitrary point, it follows that

$$S(t) \leqslant \left(\frac{\mu}{\gamma} h_m^{(0)}(t) + h_m^{(1)}(t) + \tilde{c}(t)\, S(t)\right) \frac{\gamma}{\mu} [1 - \mathrm{e}^{-\mu(t-t_0)/\gamma}] + S(t_0).$$

Thus,
$$0 \leqslant \frac{S(t) - S(t_0)}{t - t_0} \leqslant \left(\frac{\mu}{\gamma} h_m^{(0)}(t) + h_m^{(1)}(t) + \tilde{c}(t)\, S(t)\right) \left\{\frac{1 - \mathrm{e}^{-\mu(t-t_0)/\gamma}}{\mu(t-t_0)/\gamma}\right\}.$$

Therefore S is a Lipschitz function. On letting $t_0 \to t$ it follows that

$$S'(t) \leqslant (\mu/\gamma)\, h_m^{(0)}(t) + h_m^{(1)}(t) + \tilde{c}(t)\, S(t),$$

except on a set of zero measure. A further application of Gronwall's inequality gives

$$S(t) \leqslant \left[\frac{\mu}{\gamma} h_m^{(0)}(t) + h_m^{(1)}(t)\right] \frac{\mathrm{e}^{\tilde{c}(t)t} - 1}{\tilde{c}(t)},$$

so that
$$|\eta(x,t)| \leqslant \left[\frac{\mu}{\gamma} h_m^{(0)}(t) + h_m^{(1)}(t)\right] \frac{\mathrm{e}^{\tilde{c}(t)t} - 1}{\tilde{c}(t)} \frac{1}{w(x)}.$$

So far we have held X fixed, but if we now let $X \to \infty$ the conclusion of the lemma follows except that a is replaced by a_0 in the function $\tilde{c}(t)$. But, having deduced that $\eta(x,t)$ decreases exponentially in x we can repeat the above argument with $w(x) = \mathrm{e}^{rx}$ for all $x \geqslant 0$. We now know that $\|v(\cdot,t)\| = \|w(\cdot,t)\| < \infty$, and the argument using Gronwall's inequality is therefore valid. The improvement in the constant a comes about because

$$\int_0^\infty \mathrm{e}^{-|x-y|/\sqrt{\gamma}}\, \mathrm{e}^{r(x-y)}\, \mathrm{d}y \leqslant (\gamma^{-\frac{1}{2}} - r)^{-1} + (\gamma^{-\frac{1}{2}} + r)^{-1} \equiv a.$$

Using this estimate instead of (3.16) leads to the stated result.

4. The numerical scheme

The numerical scheme is based on the integral equation (3.4). The equation is first discretized in space, its right-hand side being approximated by numerical quadrature; the resultant system of ordinary differential equations is integrated forward in time by a finite difference method.

4.1. *Spatial discretization*

The spatial discretization is effected by approximating the integrals of (3.4) by the trapezoidal rule with derivative correction at the end points of the domain (see Davis & Rabinowitz 1967). Thus, truncating the half line $[0, \infty[$ and introducing a uniform partition of $N+1$ points, $\{0, \Delta x, 2\Delta x, \ldots, N\Delta x\}$, we have, for any sufficiently smooth function $V(x)$, the approximation

$$\int_{j\Delta x}^{k\Delta x} V(y)\, \mathrm{d}y \approx I_{j,k}(V)$$
$$\equiv \Delta x \left[\tfrac{1}{2}(V(j\Delta x+) + V(k\Delta x-)) + \sum_{i=j+1}^{k-1} V(i\Delta x)\right] + \tfrac{1}{12}\Delta x^2 (V'(j\Delta x+) - V'(k\Delta x-)), \tag{4.1}$$

where j, k are natural numbers with $0 \leqslant j < k \leqslant N$. This approximation has fourth-order accuracy, provided V has four bounded continuous derivatives on the open interval $]j\Delta x, k\Delta x[$ (see §5.1 below).

In using (4.1) to approximate (3.4) we note that the function $V(y)$ takes the form $J(x,y)\,v(y)$, where $v(y)$ is assumed to have four bounded, continuous derivatives on $]0, N\Delta x[$, and $J(x,y)$ (which is used to symbolize either $\tilde{H}$ or $\tilde{K}$) has four bounded, continuous derivatives, as a function of y, on each of $]0,x[$ and $]x, N\Delta x[$. Thus, the approximation (4.1) is to be applied separately on each of these intervals and the sum is to be taken. When N is large enough it can be shown (see §5.1) that the contribution from the right-hand end point, $N\Delta x$, is negligibly small, so that terms arising there can be omitted from the numerical scheme. Therefore, if we denote $J(i\Delta x, y)$ by $J_i(y)$, $0 \leqslant i \leqslant N$, it follows that

$$\begin{aligned}\int_0^{N\Delta x} J_i(y)\,v(y)\,\mathrm{d}y \approx {} & \tfrac{1}{2}\Delta x[J_i(0)\,v(0) + (J_i(i\Delta x-) + J_i(i\Delta x+))\,v(i\Delta x)] + \Delta x \sum_{j=1,\,j\neq i}^{N} J_i(j\Delta x)\,v(j\Delta x) \\ & + \tfrac{1}{12}\Delta x^2[(J_i(y)\,v(y))'|_{0+} - (J_i(y)\,v(y))'|_{i\Delta x-} + (J_i(y)\,v(y))'|_{i\Delta x+}] \\ & (= (I_{0,i} + I_{i\,N})\,(J_i(y)\,v(y)) - \tfrac{1}{2}\Delta x\,J_i(N\Delta x)\,v(N\Delta x) + \tfrac{1}{12}\Delta x^2 (J_i(y)\,v(y))'|_{N\Delta x-}). \quad (4.2)\end{aligned}$$

If we further define $v_j \equiv v(j\Delta x)$ and $J_{ij} \equiv J(i\Delta x, j\Delta y)$ and introduce the particular forms for $\tilde{H}$ and $\tilde{K}$, we can (after some simplification) rewrite (4.2) in the form

$$\begin{aligned}\int_0^{N\Delta x} \tilde{H}_i(y)\,v(y)\,\mathrm{d}y \approx {} & \mathrm{e}^{-i\Delta x/\sqrt{\gamma}}\left[\tfrac{1}{2}(\Delta x/\gamma^{\frac{3}{2}}) \sum_{j=1}^{N} \mathrm{e}^{-j\Delta x/\sqrt{\gamma}} v_j - \tfrac{1}{12}(\Delta x/\gamma)^2 v_0\right] \\ & + \Delta x \sum_{j=1}^{N} H_{ij} v_j + \tfrac{1}{12}(\Delta x/\gamma)^2 v_i, \quad (4.3)\end{aligned}$$

and

$$\begin{aligned}\int_0^{N\Delta x} \tilde{K}_i(y)\,v(y)\,\mathrm{d}y \approx {} & \tfrac{1}{2}(\Delta x/\gamma)\,\mathrm{e}^{-i\Delta x/\sqrt{\gamma}}\left[v_0 + \sum_{j=1}^{N} \mathrm{e}^{-j\Delta x/\sqrt{\gamma}} v_j + \tfrac{1}{6}\Delta x\,v'(0+)\right] \\ & + \Delta x \sum_{j=1}^{N} K_{ij} v_j - \tfrac{1}{12}(\Delta x^2/\gamma)\,v'(i\Delta x), \quad (4.4)\end{aligned}$$

where

$$H_{ij} \equiv -(1/2\gamma^{\frac{3}{2}})\,\mathrm{e}^{-|i-j|\,\Delta x/\sqrt{\gamma}}, \quad \text{and} \quad K_{ij} \equiv (1/2\gamma)\,\mathrm{sgn}\,(i-j)\,\mathrm{e}^{-|i-j|\,\Delta x/\sqrt{\gamma}}, \quad i \neq j, \quad K_{ii} \equiv 0.$$

The continuous quantities $v'(0+)$ and $v'(i\Delta x)$ in (4.4) are still to be discretized. Since both these terms derive from the second-order correction terms to the trapezoidal rule it is sufficient to approximate them only to second order to retain the overall fourth-order approximation of the integral. Thus, writing

$$v'(0+) \approx (-v_2 + 4v_1 - 3v_0)/2\Delta x, \quad v'(i\Delta x) \approx (v_{i+1} - v_{i-1})/2\Delta x,$$

and incorporating these approximations in (4.4), we have

$$\begin{aligned}\int_0^{N\Delta x} \tilde{K}_i(y)\,v(y)\,\mathrm{d}y \approx {} & \tfrac{1}{2}(\Delta x/\gamma)\,\mathrm{e}^{-i\Delta x/\sqrt{\gamma}}\left\{\sum_{j=0}^{N} \mathrm{e}^{-j\Delta x/\sqrt{\gamma}} v_j + \tfrac{1}{12}(-v_2 + 4v_1 - 3v_0)\right\} \\ & + \Delta x \sum_{j=1}^{N} K_{ij} v_j - \tfrac{1}{24}(\Delta x/\gamma)\,(v_{i+1} - v_{i-1}). \quad (4.5)\end{aligned}$$

Using (4.3) and (4.5), we construct an approximation to the right-hand side of (3.4), at grid points $i\Delta x$, $1 \leqslant i \leqslant N$, of the form

$$F_i(t, ((\alpha\eta + \tfrac{1}{2}\beta\eta^2)(0,t), \ldots, (\alpha\eta + \tfrac{1}{2}\beta\eta^2)(N\Delta x, t))) - \mu G_i(\eta(0,t), \ldots, \eta(N\Delta x, t)), \quad (4.6)$$

where $\boldsymbol{F}$, $\boldsymbol{G}$ are vector functions with $\boldsymbol{F} \equiv (F_1, \ldots, F_N)$ and $\boldsymbol{G} \equiv (G_1, \ldots, G_N)$. It is convenient for computation to write each of $\boldsymbol{F}$ and $\boldsymbol{G}$ as the sum of two vectors, i.e.

$$\boldsymbol{F}(t, \boldsymbol{v}) \equiv \boldsymbol{F}^1(t, \boldsymbol{v}) + \boldsymbol{F}^2(\boldsymbol{v}), \quad \boldsymbol{G}(\boldsymbol{v}) \equiv \boldsymbol{G}^1(\boldsymbol{v}) + \boldsymbol{G}^2(\boldsymbol{v}),$$

where

$$\left.\begin{aligned} F_i^1 &= \mathrm{e}^{-i\Delta x/\sqrt{\gamma}}\left\{h'(t) + \tfrac{1}{2}(\Delta x/\gamma)\left[\sum_{j=0}^{N} \mathrm{e}^{-j\Delta x/\sqrt{\gamma}} v_j + \tfrac{1}{12}(-v_2 + 4v_1 - 3v_0)\right]\right\} \\ &\quad + \Delta x \sum_{j=1}^{N} K_{ij} v_j, \quad \text{for} \quad i = 0, 1, 2, \ldots, \\ F_i^2 &= \begin{cases} -\tfrac{1}{24}(\Delta x/\gamma)\,(v_{i+1} - v_{i-1}), & \text{for} \quad i = 1, 2, \ldots, N-1, \\ \tfrac{1}{24}(\Delta x/\gamma)\, v_{N-1}, & \text{for} \quad i = N, \end{cases} \end{aligned}\right\} \tag{4.7}$$

$$\left.\begin{aligned} G_i^1 &= \mathrm{e}^{-i\Delta x/\sqrt{\gamma}}\left\{-[\gamma^{-1} + \tfrac{1}{12}(\Delta x/\gamma)^2]\, v_0 + \tfrac{1}{2}(\Delta x/\gamma^{\frac{3}{2}}) \sum_{j=1}^{N} \mathrm{e}^{-j\Delta x/\sqrt{\gamma}} v_j\right\} \\ &\quad + \Delta x \sum_{j=1}^{N} H_{ij} v_j, \quad \text{for} \quad i = 0, 1, 2, \ldots, \\ G_i^2 &= [\gamma^{-1} + \tfrac{1}{12}(\Delta x/\gamma)^2] v_i, \quad \text{for} \quad i = 1, 2, \ldots, N. \end{aligned}\right\} \tag{4.8}$$

Here $\boldsymbol{v} = (v_0, \ldots, v_N)$. Note that F_i^1, G_i^1 are defined for all $i \geqslant 0$, even though they involve only $v_0, \ldots, v_N$.

4.2. *An efficient computational procedure*

Before discussing the discretization in time it is worth while to consider efficient ways of computing $\boldsymbol{F}^1$ and $\boldsymbol{G}^1$. Evaluating $\boldsymbol{F}^1$ and $\boldsymbol{G}^1$ directly would require $O(N^2)$ operations, although this can easily be reduced to $O(N \ln N)$ operations through the use of a fast convolution method. However, it is possible to view (4.6) as a difference approximation to (3.1 *a*) and to reduce the computation of $\boldsymbol{F}^1$ and $\boldsymbol{G}^1$ to $O(N)$ operations. To do this, we introduce a difference operator D^2 defined by

$$\begin{aligned} (\mathrm{D}^2 \boldsymbol{w})_i &= w_i - (w_{i+1} - 2w_i + w_{i-1})/(\mathrm{e}^{\Delta x/\sqrt{\gamma}} - 2 + \mathrm{e}^{-\Delta x/\sqrt{\gamma}}), \\ &\equiv A w_i + B(w_{i+1} + w_{i-1}), \end{aligned}$$

so that $A = 1 - 2B$. The operator D^2 is effectively an infinite-order approximation to $1 - \gamma \partial_x^2$, in the sense that, when D^2 is applied to the integral kernel for the inverse of $1 - \gamma \partial_x^2$, the Kronecker δ-function results, *exactly*. (To see this, note first that if $w_i = \mathrm{e}^{i\Delta x/\sqrt{\gamma}}$, $i \in \mathbb{Z}$, then $\mathrm{D}^2 \boldsymbol{w} \equiv 0$.) Thus, by applying D^2 to $\boldsymbol{F}^1$ and $\boldsymbol{G}^1$ (and after some simplification) it follows, for $2 \leqslant i \leqslant N-1$, that

$$\left.\begin{aligned} [\mathrm{D}^2 \boldsymbol{F}^1(t, \boldsymbol{v})]_i &= \tfrac{1}{2}(B\Delta x/\gamma)\,(v_{i+1} - v_{i-1}), \\ \text{and} \qquad [\mathrm{D}^2 \boldsymbol{G}^1(\boldsymbol{v})]_i &= (B\Delta x/\gamma^{\frac{3}{2}}) \sinh(\Delta x/\sqrt{\gamma})\, v_i. \end{aligned}\right\} \tag{4.9}$$

To complete the system of equations the values of $(\mathrm{D}^2 \boldsymbol{F}^1)_i$, $(\mathrm{D}^2 \boldsymbol{G}^1)_i$ for $i = 1$ and $i = N$ must be calculated, where, here only, we let $\boldsymbol{F}^1$ denote $(F_0^1, F_1^1, \ldots, F_{N+1}^1)$, and similarly for $\boldsymbol{G}^1$. This calculation provides, for $i = 1$,

$$\left.\begin{aligned} AF_1^1 + BF_2^1 &= -Bh'(t) + \tfrac{1}{24}(B\Delta x/\gamma)\,(13v_2 - 4v_1 - 9v_0), \\ \text{and} \qquad AG_1^1 + BG_2^1 &= (\gamma^{-1} + \tfrac{1}{12}(\Delta x/\gamma)^2)\, Bv_0 + (B\Delta x/\gamma^{\frac{3}{2}}) \sinh(\Delta x/\sqrt{\gamma})\, v_1, \end{aligned}\right\} \tag{4.10}$$

and, for $i = N$,

$$\left.\begin{aligned} BF_{N-1}^1 + AF_N^1 &= -\tfrac{1}{2}(B\Delta x/\gamma)\, v_{N-1} + (-BF_{N+1}^1), \\ \text{and} \qquad BG_{N-1}^1 + AG_N^1 &= (B\Delta x/\gamma^{\frac{3}{2}}) \sinh(\Delta x/\sqrt{\gamma})\, v_N + (-BG_{N+1}^1). \end{aligned}\right\} \tag{4.11}$$

We evaluate F^1, G^1 by solving the tridiagonal system of equations (4.9)–(4.11), which requires only $O(N)$ operations. To solve these equations we must first evaluate the terms F^1_{N+1} and G^1_{N+1} that appear in (4.11), the calculations for which can be made explicitly by using the formulae (4.7), (4.8). (Note that such a computation involves only $O(N)$ operations.) However, it can be shown (see § 5.1) that the retention of the quantities F^1_{N+1} and G^1_{N+1} is of only exponentially small consequence and it is more convenient simply to discard them from the system (4.9)–(4.11). We shall represent the solution of the resulting set of equations (i.e. the ones without F^1_{N+1}, G^1_{N+1}) by $\tilde{F}^1$, $\tilde{G}^1$.

Let us denote the *semi-discrete approximation* to η by the vector function

$$\boldsymbol{u}(t) = (u_0(t), u_1(t), \ldots, u_N(t)),$$

where $\boldsymbol{u}$ is defined by

$$\left.\begin{aligned} & u_0(t) = h(t), \quad t \geqslant 0, \\ \text{and} \quad & \dot{u}_i(t) = \tilde{F}_i[t, (\alpha \boldsymbol{u} + \tfrac{1}{2}\beta \boldsymbol{u} \circ \boldsymbol{u})(t)] - \mu \tilde{G}_i[\boldsymbol{u}(t)], \quad i = 1, \ldots, N, \end{aligned}\right\} \tag{4.12}$$

and $\tilde{\boldsymbol{F}} = \tilde{\boldsymbol{F}}^1 + \boldsymbol{F}^2$, $\tilde{\boldsymbol{G}} = \tilde{\boldsymbol{G}}^1 + \boldsymbol{G}^2$. Here $\boldsymbol{u} \circ \boldsymbol{u}$ denotes the vector $(u_0^2, u_1^2, \ldots, u_N^2)$. Then if h is identified with u_0 wherever it appears in the definition of $\tilde{F}$ and $\tilde{G}$, the set of equations (4.12) may be written as a system of ordinary differential equations,

$$\dot{\boldsymbol{u}} = \mathscr{F}(t, \boldsymbol{u}), \tag{4.13}$$

for the vector $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)$. This set of equations can be shown (see § 5.3) to have a solution on an interval $[0, T_0]$, where T_0 tends to infinity as both $\Delta x \to 0$ and $N\Delta x \to \infty$.

4.3. *Temporal discretization*

The temporal discretization of (4.13) has been effected through a prediction–correction method which, here, is efficient because the initial datum is zero and no start-up procedure is needed. The continuous quantity $h'(t)$ appearing in the definition of $\tilde{F}$ (cf. (4.7)) is calculated by the fourth-order central-difference formula

$$h'(n\Delta t) \approx \mathrm{d}h^n = (h^{n-2} - 8h^{n-1} + 8h^{n+1} - h^{n+2})/12\Delta t, \tag{4.14}$$

where $h^n = h(n\Delta t)$, $n \in \mathbb{N}$. Let $\mathscr{F}^n(\boldsymbol{v})$ denote the function obtained by substituting $\mathrm{d}h^n$ for $h'(n\Delta t)$ in $\mathscr{F}(n\Delta t, \boldsymbol{v})$ at that point. Then we take the fully discrete approximation to η to be the vector function given by Moulton's method (cf. Isaacson & Keller 1966), namely

$$\left.\begin{aligned} & \bar{\boldsymbol{u}}^{n+1} = \boldsymbol{u}^n + \tfrac{1}{24}\Delta t[55\mathscr{F}^n(\boldsymbol{u}^n) - 59\mathscr{F}^{n-1}(\boldsymbol{u}^{n-1}) + 37\mathscr{F}^{n-2}(\boldsymbol{u}^{n-2}) - 9\mathscr{F}^{n-3}(\boldsymbol{u}^{n-3})], \\ \text{and} \quad & \boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \tfrac{1}{24}\Delta t[9\mathscr{F}^{n+1}(\bar{\boldsymbol{u}}^{n+1}) + 19\mathscr{F}^n(\boldsymbol{u}^n) - 5\mathscr{F}^{n-1}(\boldsymbol{u}^{n-1}) + \mathscr{F}^{n-2}(\boldsymbol{u}^{n-2})]. \end{aligned}\right\} \tag{4.15}$$

Since, for $t \leqslant 0$, $\eta(x, t)$ is presumed to be zero, we shall take $\boldsymbol{u}^0, \boldsymbol{u}^{-1}, \ldots$ (and $h^0, h^{-1}, \ldots$) to be zero as the starting values for (4.15).

The error induced by using the above scheme to approximate the solution of (3.1) can be shown to be $O(\Delta x^4 + \Delta t^4)$ (see § 5) and, since $\mathscr{F}$ remains bounded as $\Delta x \to 0$, there are no stability limitations on Δt or Δx. The same methods can be used to develop schemes of arbitrary order of accuracy by using higher-order derivative corrections for the trapezoidal rule (i.e. the Euler–Maclaurin formula) and higher-order prediction–correction methods. But before studying the accuracy of the approximation theoretically we shall first describe some numerical tests made with the scheme.

4.4. *Convergence tests*

The theoretical convergence rate of the scheme was checked by comparing numerical solutions for the propagation of a solitary wave with the exact solution for the continuous equation. With $\mu = 0$ and for $x \in \mathbb{R}$, there is a family of exact solutions to (3.1 a) of the form

$$\eta = \eta_0 \operatorname{sech}^2\{[\beta\eta_0/12\gamma(\alpha+\tfrac{1}{3}\beta\eta_0)]^{\frac{1}{2}}[x+x_0-(\alpha+\tfrac{1}{3}\beta\eta_0)t]\}, \tag{4.16}$$

where $\eta_0 > 0$ is the (maximum) amplitude of the wave and x_0 is a real constant. The wave propagates without change of form at a steady speed $\alpha+\frac{1}{3}\beta\eta_0$. The constant x_0 is a parameter used to 'offset' the solitary wave so that, at $t = 0$, the wave crest is located at $x = -x_0$; alternatively, the wave crest passes an observer stationed at $x = 0$ at time $t = x_0/(\alpha+\frac{1}{3}\beta\eta_0)$. Suppose therefore, at $x = 0$ and for $t > 0$, that (4.16) were used as the boundary data $h(t)$, then there would result an exact solution to (3.1), subject of course to some non-zero initial data $g(x)$, say. By choosing x_0 sufficiently large, the exponential decay of (4.16) implies that the maximum value of $|g|$ can be made arbitrarily small, in which case the specification $g \equiv 0$ provides a close approximation to an exact solution of (3.1). It should, however, be noted that such a specification introduces an incompatibility at the origin and this may slightly pollute the numerical solutions.

TABLE 1. THE ERRORS $\bar{E}_m(T)$ INDUCED IN INTEGRATING A SOLITARY WAVE (4.16) WITH $\eta_0 = 0.25$, FOR A TIME T

($N\Delta x = 180.0$; $\Delta = \Delta t = \Delta x$; $\alpha = 1, \beta = \frac{3}{2}, \mu = 0, \gamma = \frac{1}{6}$. An entry in a row labelled 'ratio' is the ratio of the numbers above and below that entry.)

(a) $\eta(0,0)/\eta_0 = 0.1\times10^{-8}$, $x_0 \approx 27.029$

Δ	T 19.2	38.4	67.2	96.0	172.8
0.6	0.933(−4)	0.209(−1)	0.564(−1)	0.923(−1)	0.169
ratio	12.4	14.3	13.1	12.5	10.0
0.3	0.753(−5)	0.146(−2)	0.429(−2)	0.740(−2)	0.169(−1)
ratio	13.6	15.6	15.5	15.7	16.3
0.15	0.554(−6)	0.938(−4)	0.276(−3)	0.470(−3)	0.104(−2)
ratio	14.2	15.8	16.1	16.3	16.9
0.075	0.389(−7)	0.595(−5)	0.171(−4)	0.288(−4)	0.616(−4)
max $\|\eta\|$	0.110(−1)	0.250	0.250	0.250	0.250

(b) $\eta(0,0)/\eta_0 = 0.1\times10^{-9}$, $x_0 \approx 29.899$

Δ	T 19.2	38.4	67.2
0.15	0.556(−7)	0.790(−4)	0.259(−3)
ratio	14.2	15.7	16.1
0.075	0.391(−8)	0.502(−5)	0.161(−4)
ratio	15.0	15.8	16.1
0.0375	0.260(−9)	0.317(−6)	0.100(−5)
max $\|\eta\|$	0.110(−2)	0.250	0.250

Nevertheless, for the initial data $g(x) = 0$, we have taken (4.16) as an 'exact' solution and have made a convergence test for the scheme, the results of which are given in table 1. Let $u_i(T)$ be the computed solution at time T and at $x = i\Delta x$, $0 \leqslant i \leqslant N$. Then the entries shown in table 1 are $\bar{E}_m(T) = \max\{|u_i(T)-\eta(i\Delta x, T)| : 0 \leqslant i \leqslant N\}$. The computations reported in table 1 a were made with $\eta_0 = 0.25$ (which was roughly the largest wave amplitude encountered in the

experiments), with x_0 chosen so that $\eta/\eta_0 = 0.1 \times 10^{-8}$ at $(0, 0)$, and with $\Delta t = \Delta x\ (\equiv \Delta)$. The choice of $\Delta t = \Delta x$ was made because preliminary tests suggested this was near the optimal choice, in terms of accuracy achieved for a given amount of work, and because it is sufficient to take $\Delta t/\Delta x =$ constant to check the convergence rate, if the error is proportional to $\Delta t^4 + \Delta x^4$. The domain used for these computations was approximately the same as that needed to make comparisons with the laboratory experiments.

It is seen in table 1 *a* that, apart from the smallest time quoted, the errors decreased at approximately the 16:1 ratio expected of the scheme when the mesh is halved. At $t = 19.2$ the wave crest had not yet emerged from the 'wavemaker', so that the wave amplitudes were quite small (cf. the value of $\max|\eta|$ quoted in the table) and the influence of the truncation of the input waveform is reflected by convergence rates being smaller than expected for the scheme. With $\eta(0, 0)/\eta_0$ chosen to be 0.1×10^{-9} the errors $\bar{E}_m$ (see table 1 *b*) were of a similar form to those given in table 1 *a*. Indeed, the difference between (*a*) and (*b*) is not as great as might appear from table 1. For example, when the error $\bar{E}_m$ was determined at the times at which the wave crest had reached $x = 45.70$, the errors for each experiment were nearly the same (for $\Delta = 0.15, 0.075$).

A similar test of the convergence of the numerical scheme was made by comparing solutions with $\eta(X, t)$ for X fixed. If $u^j(X)$ is the computed solution at position X and at time $t = j\Delta t$, $0 \leqslant j \leqslant M$, then we have calculated

$$E_1(X) = \sum_{j=0}^{M} |u^j(X) - \eta(X, j\Delta t)|\,\Delta t, \quad E_2(X) = \left\{\sum_{j=0}^{M} [u^j(X) - \eta(X, j\Delta t)]^2\,\Delta t\right\}^{\frac{1}{2}},$$

and

$$E_m(X) = \max\{|u^j(X) - \eta(X, j\Delta t)| : 0 \leqslant j \leqslant M\}.$$

The results of such a calculation for $\eta_0 = 0.25$ and $X = 36.0$ are given in table 2, and again a convergence order of about 4 was obtained.

TABLE 2. THE ERRORS E_1, E_2, E_m INDUCED IN INTEGRATING A SOLITARY WAVE (4.16) WITH $\eta_0 = 0.25$, $X = 36.0$

($M\Delta t = 180.0$. $\Delta = \Delta t = \Delta x$; $\alpha = 1, \beta = \frac{3}{2}, \mu = 0, \gamma = \frac{1}{6}$; $\eta(0, 0)/\eta_0 = 0.1 \times 10^{-8}$, $x_0 \approx 27.079$. An entry in a row labelled ratio is the ratio of the numbers above and below that entry.)

Δ	$E_1(X)$	$E_2(X)$	$E_m(X)$
0.6	0.286	0.178	0.165
ratio	15.6	13.6	13.8
0.3	0.183(−1)	0.131(−1)	0.120(−1)
ratio	16.6	15.1	15.5
0.15	0.110(−2)	0.870(−3)	0.774(−3)
ratio	16.9	15.8	16.0
0.075	0.651(−4)	0.551(−4)	0.485(−4)

With $\mu \neq 0$, we do not know of an exact solution to the continuous equation, so the convergence rate of the scheme was checked in a different way. To ascertain that the coding of the dissipative term was correct, experiments were run with the linear model (i.e. $\beta = 0$) with $h(t)$ chosen to be sinusoidal in time, and the decay rate of these waves was compared with that deduced from the dispersion relation. (The results of a test of this kind are described in §7.5.)

Having checked that the dissipative term had been correctly coded, the convergence rate for the full equation (with $\beta = 1.5$) was estimated by taking the 'exact' solution to be the results from a computation made with a small value of Δ (i.e. $\Delta = 0.0375$) and comparing this solution

with numerical solutions at larger values of Δ. Thus, using (4.16) at $x = 0$ for the boundary data $h(t)$, with $\eta(0,0)/\eta_0$ chosen to be 0.1×10^{-8} (i.e. $x_0 \approx 27.079$), and with $\mu = 0.014$ (the value used in the comparisons of §7.3) the convergence orders, as shown in table 3, were again found to be about 4.

TABLE 3. A CONVERGENCE TABLE FOR THE NUMERICAL SCHEME WITH $\mu \neq 0$

($M\Delta t = 75.0$; $\Delta = \Delta t = \Delta x$; $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$. An entry in a row labelled ratio is the ratio of the numbers above and below that entry.)

Δ	$E_1(15.0)$	$E_2(15.0)$	$E_m(15.0)$	$E_1(30.0)$	$E_2(30.0)$	$E_m(30.0)$
0.6	0.105	0.707(−1)	0.651(−1)	0.176	0.133	0.118
ratio	16.0	14.4	14.9	14.3	13.6	13.6
0.3	0.655(−2)	0.492(−2)	0.438(−2)	0.123(−1)	0.975(−2)	0.866(−2)
ratio	16.4	15.2	15.6	15.9	15.3	15.7
0.15	0.400(−3)	0.323(−3)	0.280(−3)	0.772(−3)	0.639(−3)	0.553(−3)
ratio	17.1	16.6	16.8	17.1	16.8	16.9
0.075	0.234(−4)	0.195(−4)	0.167(−4)	0.451(−4)	0.380(−4)	0.327(−4)

5. ERROR ESTIMATES FOR THE DISCRETE SCHEME

In this chapter we shall let c_i, $i = 1, 2, \ldots$, denote real constants. Also, we shall assume that $\Delta t, \Delta x \leqslant 1$ so that the dependence of constants on *positive* powers of Δt and Δx can be ignored. The notation is the same as that used in §§3, 4.

5.1. *Spatial discretization errors*

The error associated with the trapezoidal rule with derivative end correction, as given in (4.1), is as follows.

LEMMA 5.1. *If V has four bounded, continuous derivatives on the open interval $]j\Delta x, k\Delta x[$, then*

$$\left| \int_{j\Delta x}^{k\Delta x} V(y)\,\mathrm{d}y - I_{j,k}(V) \right| \leqslant \frac{\Delta x^4}{384} \int_{j\Delta x}^{k\Delta x} |V^{(4)}(y)|\,\mathrm{d}y.$$

This is a standard result (see, for example, Davis & Rabinowitz 1967).

The error arising from the use of the vector $\boldsymbol{F}^1$ can be estimated as the sum of a term proportional to Δx^4 and a term arising from the approximation used at the right-hand extremity of the interval:

LEMMA 5.2. *Suppose that v has four continuous derivatives on the interval $[0, N\Delta x]$. Let $\boldsymbol{v} = (v_0, \ldots, v_N)$, where $v_i \equiv v(i\Delta x)$. Then, for $i = 1, 2, \ldots, N$,*

$$\left| F_i(t, \boldsymbol{v}) - h'(t)\,\mathrm{e}^{-i\Delta x/\sqrt{\gamma}} - \int_0^{N\Delta x} \tilde{K}(i\Delta x, y)\, v(y)\,\mathrm{d}y \right|$$
$$\leqslant c_1 \Delta x^4 \max\{|v^{(j)}(x)| : x \in [0, N\Delta x], j = 0, \ldots, 4\} + c_2 \Delta x \max\{|v_{N-k}| : k = 0, 1, 2\},$$

where the constants c_1, c_2 depend only on γ.

Proof. By definition (see (4.1), (4.2), (4.7)) it follows, for $i = 1, \ldots, N-1$, that

$$F_i(t, \boldsymbol{v}) = h'(t)\,\mathrm{e}^{-i\Delta x/\sqrt{\gamma}} + I_{0,i}(\tilde{K}(i\Delta x, \cdot)\, v) + I_{i,N}(\tilde{K}(i\Delta x, \cdot)\, v)$$
$$- \frac{\Delta x^2}{12\gamma}\,\mathrm{e}^{-i\Delta x/\sqrt{\gamma}} \left[v'(0) - \frac{-v_2 + 4v_1 - 3v_0}{2\Delta x} \right] + \frac{\Delta x^2}{12\gamma} \left[v'(i\Delta x) - \frac{v_{i+1} - v_{i-1}}{2\Delta x} \right]$$
$$+ \tfrac{1}{2}\Delta x\, \tilde{K}(i\Delta x, N\Delta x)\, v_N + \tfrac{1}{12}\Delta x^2 (\tilde{K}(i\Delta x, \cdot)\, v)'|_{N\Delta x}.$$

The difference approximations in the fourth and fifth terms are less than

$$\tfrac{1}{3}\Delta x^2 \max\{|v^{(3)}(x)|: x\in[0, N\Delta x]\}.$$

The last term can be estimated as follows:

$$\begin{aligned}|(\tilde{K}(i\Delta x, \cdot)\, v)'|_{N\Delta x}| &= |\tilde{K}_y(i\Delta x, N\Delta x)\, v_N + \tilde{K}(i\Delta x, N\Delta x)\, v'(N\Delta x)|,\\ &\leqslant \gamma^{-\frac{3}{2}}|v_N| + \gamma^{-1}|v'(N\Delta x) - (v_{N-2} - 4v_{N-1} + 3v_N)/2\Delta x|\\ &\quad + (\gamma^{-1}/2\Delta x)|v_{N-2} - 4v_{N-1} + 3v_N|,\\ &\leqslant (\Delta x^2/3\gamma)\max\{|v^{(3)}(x)|: x\in[0, N\Delta x]\}\\ &\quad + (\gamma^{-\frac{3}{2}} + (2/\gamma\Delta x))\max\{|v_{N-k}|: k = 0, 1, 2\}.\end{aligned}$$

Combining these estimates, together with a direct estimate for the penultimate term, we have that

$$\begin{aligned}&|F_i(t, \boldsymbol{v}) - h'(t)\,\mathrm{e}^{-i\Delta x/\sqrt{\gamma}} - I_{0,i}(\tilde{K}(i\Delta x, \cdot)\, v) - I_{i,N}(\tilde{K}(i\Delta x, \cdot)\, v)|\\ &\quad\leqslant \frac{\Delta x^4}{12\gamma}\max\{|v^{(3)}(x)|: x\in[0, N\Delta x]\} + \Delta x\left(\frac{2}{3\gamma} + \frac{\Delta x}{12\gamma^{\frac{3}{2}}}\right)\max\{|v_{N-k}|: k = 0, 1, 2\}. \quad (5.1)\end{aligned}$$

But lemma 5.1 implies that

$$\begin{aligned}E &\equiv \left| I_{0,i}(\tilde{K}(i\Delta x, \cdot)\, v) + I_{i,N}(\tilde{K}(i\Delta x, \cdot)\, v) - \int_0^{N\Delta x} \tilde{K}(i\Delta x, y)\, v(y)\,\mathrm{d}y\right|\\ &\leqslant \frac{\Delta x^4}{384}\left[\int_0^{i\Delta x} |(\tilde{K}(i\Delta x, \cdot)\, v)^{(4)}(y)|\,\mathrm{d}y + \int_{i\Delta x}^{N\Delta x} |(\tilde{K}(i\Delta x, \cdot)\, v)^{(4)}(y)|\,\mathrm{d}y\right],\end{aligned}$$

which can be estimated further through the use of Leibnitz's rule (together with (3.6)) and the Hölder inequality. Thus, it follows that

$$E \leqslant c_3\Delta x^4 \max\{|v^{(j)}(x)|: x\in[0, N\Delta x], j = 0, \ldots, 4\},$$

where c_3 depends only on γ. Then, combining this estimate and (5.1), we have the required result for $i \neq N$. For $i = N$,

$$\begin{aligned}F_N(t, \boldsymbol{v}) &= h'(t)\,\mathrm{e}^{-N\Delta x/\sqrt{\gamma}} + I_{0,N}(\tilde{K}(N\Delta x, \cdot)\, v) + \tfrac{1}{2}\Delta x\,\tilde{K}(N\Delta x, N\Delta x-)\, v_N\\ &\quad - \frac{\Delta x^2}{12\gamma}\mathrm{e}^{-N\Delta x/\sqrt{\gamma}}\left[v'(0) - \frac{-v_2 + 4v_1 - 3v_0}{2\Delta x}\right] + \tfrac{1}{12}\Delta x^2(\tilde{K}(N\Delta x, \cdot)\, v)'\Big|_{N\Delta x-} + \frac{\Delta x}{24\gamma}v_{N-1}.\end{aligned}$$

The techniques used when $i \neq N$ apply in the same way in this case. (The constants here can be chosen to be the same as for $i \neq N$.)

A similar result can be established in relation to the vector $\boldsymbol{G}$.

LEMMA 5.3. *Suppose that v has four continuous derivatives on the interval $[0, N\Delta x]$. Let $\boldsymbol{v} = (v_0, \ldots, v_N)$, where $v_i \equiv v(i\Delta x)$. Then, for $i = 1, \ldots, N$,*

$$\begin{aligned}&\left|G_i(\boldsymbol{v}) - \gamma^{-1}v_i + \gamma^{-1}\mathrm{e}^{-i\Delta x/\sqrt{\gamma}}\, v_0 - \int_0^{N\Delta x}\tilde{H}(i\Delta x, y)\, v(y)\,\mathrm{d}y\right|\\ &\quad\leqslant c_4\Delta x^4\max\{|v^{(j)}(x)|: x\in[0, N\Delta x], j = 0, \ldots, 4\} + c_5\Delta x\max\{|v_{N-k}|: k = 0, 1, 2\},\end{aligned}$$

where the constants c_4, c_5 depend only on γ.

The proof of this lemma follows a similar pattern to that for lemma 5.2 and is therefore omitted.

The above lemmas can now be combined to give the following estimate.

COROLLARY 5.1. *Suppose that v has four continuous derivatives on the interval $[0, N\Delta x]$. Let $\boldsymbol{v} = (v_0, \ldots, v_N)$, where $v_i \equiv v(i\Delta x)$. Then, for $i = 1, 2, \ldots, N$,*

$$\begin{aligned}
&\left|\tilde{F}_i(t, \boldsymbol{v}) - h'(t)\, e^{-i\Delta x/\sqrt{\gamma}} - \int_0^{N\Delta x} \tilde{K}(i\Delta x, y)\, v(y)\, dy\right| \\
&\quad + \left|\tilde{G}_i(\boldsymbol{v}) - \gamma^{-1} v_i + \gamma^{-1} e^{-i\Delta x/\sqrt{\gamma}}\, v_0 - \int_0^{N\Delta x} \tilde{H}(i\Delta x, y)\, v(y)\, dy\right| \\
&\qquad \leqslant c_6 \Delta x^4 \max\{|v^{(j)}(x)| : x \in [0, N\Delta x], j = 0, \ldots, 4\} + c_7 \Delta x \max\{|v_{N-k}| : k = 0, 1, 2\} \\
&\qquad\quad + e^{-(N+1)\Delta x/\sqrt{\gamma}} \left[|h'(t)| + c_8 \Delta x \sum_{j=0}^{N} e^{j\Delta x/\sqrt{\gamma}} |v_j|\right],
\end{aligned}$$

where $c_6 = c_1 + c_4$, $c_7 = c_2 + c_5$ and c_8 is another constant depending only on γ.

Proof. Define

$$s_i = \sinh(i\Delta x/\sqrt{\gamma})/\sinh[(N+1)\Delta x/\sqrt{\gamma}]. \tag{5.2}$$

Recall from §4.2 that $\tilde{\boldsymbol{F}}^1$ and $\tilde{\boldsymbol{G}}^1$ respectively differ from $\boldsymbol{F}^1$ and $\boldsymbol{G}^1$ only because the terms involving F^1_{N+1} and G^1_{N+1} were not retained. Thus, it follows from the definitions (4.9)–(4.11) that

$$\tilde{F}_i(t, \boldsymbol{v}) = F_i(t, \boldsymbol{v}) - s_i F^1_{N+1}(t, \boldsymbol{v}), \quad \tilde{G}_i(\boldsymbol{v}) = G_i(\boldsymbol{v}) - s_i G^1_{N+1}(\boldsymbol{v}), \tag{5.3}$$

for $i = 1, 2, \ldots, N$. From the definition (4.7) of $\boldsymbol{F}^1$ we see that

$$\begin{aligned}
|F^1_{N+1}(t, \boldsymbol{v})| &\leqslant e^{-(N+1)\Delta x/\sqrt{\gamma}} \left\{|h'(t)| + \tfrac{1}{2}(\Delta x/\gamma)\left[\sum_{j=0}^{N} e^{-j\Delta x/\sqrt{\gamma}} |v_j| + \tfrac{1}{4}|v_0| + \tfrac{1}{3}|v_1| + \tfrac{1}{12}|v_2|\right]\right\} \\
&\quad + \Delta x \sum_{j=1}^{N} e^{(j-N-1)\Delta x/\sqrt{\gamma}} |v_j|, \\
&\leqslant e^{-(N+1)\Delta x/\sqrt{\gamma}} \left\{|h'(t)| + c_9 \Delta x \sum_{j=0}^{N} e^{j\Delta x/\sqrt{\gamma}} |v_j|\right\},
\end{aligned}$$

where $c_9 = 1 + \frac{2}{3}(\Delta x/\gamma)$. Similarly, it follows from (4.8) that

$$|G^1_{N+1}| \leqslant c_{10} \Delta x\, e^{-(N+1)\Delta x/\sqrt{\gamma}} \sum_{j=0}^{N} e^{j\Delta x/\sqrt{\gamma}} |v_j|,$$

where c_{10} depends only on γ. Therefore, on defining $c_8 = c_9 + c_{10}$, we have that

$$|\tilde{F}_i(t, \boldsymbol{v}) - F_i(t, \boldsymbol{v})| + |\tilde{G}_i(\boldsymbol{v}) - G_i(\boldsymbol{v})| \leqslant e^{-(N+1)\Delta x/\sqrt{\gamma}} \left\{|h'(t)| + c_8 \Delta x \sum_{j=0}^{N} e^{j\Delta x/\sqrt{\gamma}} |v_j|\right\},$$

and the result follows from lemmas 5.2 and 5.3.

5.2. *Lipschitz estimate for $\mathcal{F}$*

In this subsection $\|\boldsymbol{v}\|$ will be used to denote the l_∞ norm of $\boldsymbol{v}$; i.e. if we denote the ith component of $\boldsymbol{v}$ by v_i, then $\|\boldsymbol{v}\| = \max\{|v_i|\}$.

The map $\boldsymbol{v} = (v_0, \ldots, v_N) \mapsto \boldsymbol{F}^1(t, \boldsymbol{v})$ is an affine map, taking the form $\boldsymbol{F}^1(t, \boldsymbol{v}) = \boldsymbol{L}(t) + \boldsymbol{M}\boldsymbol{v}$, say. (Recall that $\boldsymbol{F}^1 = (F^1_1, \ldots, F^1_N)$ and similarly for the other $\boldsymbol{F}$ and $\boldsymbol{G}$ vectors.) The l_∞ operator norm of $\boldsymbol{M}$ (it is the maximum, absolute row sum of $\boldsymbol{M}$) can be estimated as

$$\begin{aligned}
\|\boldsymbol{M}\| &\leqslant \max_{1 \leqslant i \leqslant N+1} \left[e^{-i\Delta x/\sqrt{\gamma}} \frac{\Delta x}{2\gamma}\left(\sum_{j=0}^{N} e^{-j\Delta x/\sqrt{\gamma}} + \tfrac{2}{3}\right) + \frac{\Delta x}{2\gamma} \sum_{j=1}^{N} e^{-|i-j|\Delta x/\sqrt{\gamma}}\right], \\
&\leqslant \frac{3}{2}\frac{\Delta x}{\gamma} \sum_{j=0}^{N} e^{-j\Delta x/\sqrt{\gamma}} + \frac{1}{3}\frac{\Delta x}{\gamma}.
\end{aligned}$$

Since $\Delta x \sum_{j=1}^{N} e^{-j\Delta x/\sqrt{\gamma}} \leqslant \sqrt{\gamma}$, it follows that $\|M\| \leqslant \frac{3}{2}\gamma^{-\frac{1}{2}} + 2\gamma^{-1} \equiv \frac{1}{2}c_{11}$. Therefore $\|F^1(t, v) - F^1(t, w)\| = \|M(v-w)\| \leqslant \frac{1}{2}c_{11}\|v-w\|$. (Note that, here and below, the norm on $v-w$ is a norm on $(N+1)$-vectors whereas all other norms are taken on N-vectors.) Similarly, we have that

$$|F^1_{N+1}(t, v) - F^1_{N+1}(t, w)| \leqslant \tfrac{1}{2}c_{11}\|v-w\|,$$

and a Lipschitz estimate for $\tilde{F}^1$ can be obtained by using (5.2) and (5.3) thus:

$$\begin{aligned}\|\tilde{F}^1(t, v) - \tilde{F}^1(t, w)\| &\leqslant \|F^1(t, v) - F^1(t, w)\| + \|\tilde{F}^1(t, v) - F^1(t, v) - (\tilde{F}^1(t, w) - F^1(t, w))\|,\\ &\leqslant \|F^1(t, v) - F^1(t, w)\| + |F_{N+1}(t, v) - F_{N+1}(t, w)|,\\ &\leqslant c_{11}\|v-w\|.\end{aligned}$$

The map $v \mapsto F^2(v)$ is linear and its l_∞ operator norm is bounded by $\Delta x/12\gamma$ (see (4.7)), so that an estimate for $\tilde{F} = \tilde{F}^1 + F^2$ is

$$\|\tilde{F}(t, v) - \tilde{F}(t, w)\| \leqslant c_F\|v-w\|, \tag{5.4}$$

where $c_F = c_{11} + \frac{1}{12}\gamma$. A similar argument can be applied to the map $v \mapsto \tilde{G}(v)$ leading to an estimate of the form

$$\|\tilde{G}(v) - \tilde{G}(w)\| \leqslant c_G\|v-w\|, \tag{5.5}$$

where c_G depends only on γ.

A combination of these two estimates can be used to obtain a Lipschitz estimate for $\mathscr{F}$ (as defined in (4.12) and (4.13)). Let v denote the vector $(h(t), v_1, \ldots, v_N)$, let $\hat{v}$ denote $(v_1, \ldots, v_N)$ and let $v \circ v$ denote $(h^2(t), v_1^2, \ldots, v_N^2)$. Then, since $\tilde{F}(t, v)$ is affine in v,

$$\begin{aligned}\|\mathscr{F}(t, \hat{v}) - \mathscr{F}(t, \hat{w})\| &\leqslant \|\tilde{F}(t, \alpha v + \tfrac{1}{2}\beta v \circ v) - \tilde{F}(t, \alpha w + \tfrac{1}{2}\beta w \circ w)\| + \mu\|\tilde{G}(v) - \tilde{G}(w)\|,\\ &\leqslant \alpha\|\tilde{F}(t, v) - \tilde{F}(t, w)\| + \tfrac{1}{2}\beta\|\tilde{F}(t, v \circ v) - \tilde{F}(t, w \circ w)\| + \mu\|\tilde{G}(v) - \tilde{G}(w)\|,\\ &\leqslant c_F(\alpha\|v-w\| + \tfrac{1}{2}\beta\|v \circ v - w \circ w\|) + c_G\mu\|v-w\|,\\ &\leqslant [c_F(\alpha + \tfrac{1}{2}\beta\|\hat{v} + \hat{w}\|) + c_G\mu]\,\|\hat{v} - \hat{w}\|.\end{aligned}$$

Thus it follows that

$$\|\mathscr{F}(t, \hat{v}) - \mathscr{F}(t, \hat{w})\| \leqslant c_L(1 + \|\hat{v} + \hat{w}\|)\,\|\hat{v} - \hat{w}\|, \tag{5.6}$$

where c_L depends only on α, β, γ and μ. So $\mathscr{F}$ is uniformly Lipschitz continuous in v on bounded subsets of l_∞.

5.3. *Existence and bounds for the semi-discrete approximation*

Let $\eta(t)$ represent the vector function $\eta_i(t) \equiv \eta(i\Delta x, t)$, $i = 1, \ldots, N$, where η is the solution to (3.1). Then, from corollary (5.1) and lemmas (3.1), (3.2) it follows that

$$\begin{aligned}\|\dot{\eta}(t) - \mathscr{F}(t, \eta)\| &\leqslant c_{12}\Delta x^4 P(h_m^{(1)}(t), \sigma(t), t) + e^{-(N+1)\Delta x/\sqrt{\gamma}}\, h_m^{(1)}(t)\\ &\quad + \frac{c_{13}\Delta x}{C}\left(\frac{\mu}{\gamma}\sigma(t) + h_m^{(1)}(t)\right)\left[e^{Ct - rN\Delta x} + e^{-(N+1)\Delta x/\sqrt{\gamma}} \sum_{j=0}^{N} e^{j\Delta x/\sqrt{\gamma}}\, e^{Ct - rj\Delta x}\right],\end{aligned}$$

where $P(\xi, \sigma, \tau) = \max\{(1 + P_i(\xi, \sigma, \tau))\, P_j(\xi, \sigma, \tau) : 0 \leqslant i < j \leqslant 4\}$, $C = C(\sigma(t))$, c_{12} and c_{13} depend only on α, β, γ and μ; and $0 < r < \gamma^{-\frac{1}{2}}$ (cf. lemma 3.2). This expression can be simplified by the use of the inequality

$$\Delta x\, e^{-(N+1)\Delta x/\sqrt{\gamma}} \sum_{j=0}^{N} e^{j\Delta x/\sqrt{\gamma} + Ct - rj\Delta x} \leqslant (\gamma^{-\frac{1}{2}} - r)^{-1}\, e^{Ct - r(N+1)\Delta x},$$

so that

$$\left.\begin{aligned}\|\dot{\boldsymbol{\eta}}(t)-\mathscr{F}(t,\boldsymbol{\eta})\| &\leqslant c_{12}\Delta x^4 P(h_m^{(1)}(t),\sigma(t),t)+h_m^{(1)}(t)\,\mathrm{e}^{-(N+1)\Delta x/\sqrt{\gamma}}\\ &\quad+\frac{c_{14}}{C}\left[\frac{\mu}{\gamma}\sigma(t)+h_m^{(1)}(t)\right]\mathrm{e}^{Ct-rN\Delta x},\\ &\equiv e_1(t)\,[=e_1(h_m^{(1)}(t),\sigma(t),t,\Delta x,N\Delta x)],\end{aligned}\right\} \tag{5.7}$$

and c_{14} depends only on α, β, γ, μ and r. Note that, by definition, e_1 is an increasing function of t.

Under the assumption that $h\in\mathscr{C}^1$, it follows from §5.2 that $\mathscr{F}$ is locally Lipschitz continuous. Thus there is a unique solution $\boldsymbol{u}(t)$ to (4.12) for $t\in[0,t_0]$ for some $t_0>0$. Suppose that T_0 is given by

$$T_0=\sup\{t_0\geqslant 0:\boldsymbol{u}(t)\text{ exists and }\|\boldsymbol{u}(t)-\boldsymbol{\eta}(t)\|\leqslant 1\text{ for }t\in[0,t_0]\}. \tag{5.8}$$

Since $\boldsymbol{u}(0)=\boldsymbol{\eta}(0)=\boldsymbol{0}$, and both $\boldsymbol{u}$ and $\boldsymbol{\eta}$ are continuous, then $T_0>0$. We shall now obtain a lower bound for T_0 and show that $T_0\to\infty$ as $\Delta x\to 0$ and $N\Delta x\to\infty$. For $t\in[0,T_0]$ it follows from (5.6) and (5.7) that

$$\begin{aligned}\|\dot{\boldsymbol{u}}(t)-\dot{\boldsymbol{\eta}}(t)\| &=\|\mathscr{F}(t,\boldsymbol{u}(t))-\dot{\boldsymbol{\eta}}(t)\|\leqslant\|\mathscr{F}(t,\boldsymbol{u}(t))-\mathscr{F}(t,\boldsymbol{\eta}(t))\|+\|\mathscr{F}(t,\boldsymbol{\eta}(t))-\dot{\boldsymbol{\eta}}(t)\|,\\ &\leqslant c_L(1+\|\boldsymbol{u}(t)+\boldsymbol{\eta}(t)\|)\,\|\boldsymbol{u}(t)-\boldsymbol{\eta}(t)\|+e_1(t),\\ &\leqslant 2c_L(1+\sigma(t))\,\|\boldsymbol{u}(t)-\boldsymbol{\eta}(t)\|+e_1(t).\end{aligned} \tag{5.9}$$

Since $(\mathrm{d}/\mathrm{d}t)\,[\max\{|u_i-\eta_i|\}]\leqslant\max\{|(\mathrm{d}/\mathrm{d}t)\,(u_i-\eta_i)|\}$, except on a set of zero measure, it follows from Gronwall's lemma that

$$\begin{aligned}\|\boldsymbol{u}(t)-\boldsymbol{\eta}(t)\| &\leqslant e_1(t)\,[\mathrm{e}^{2c_L(1+\sigma(t))t}-1]/[2c_L(1+\sigma(t))],\\ &\equiv\psi(t),\end{aligned} \tag{5.10}$$

for $t\in[0,T_0]$.

However, if T_0 were such that $\psi(T_0)<1$, it would contradict the maximality in the definition (5.8), as follows. In this case $\boldsymbol{u}(t)$ is still defined for $t\in[T_0,T_0+t_1]$, $t_1>0$, because $\mathscr{F}$ is locally Lipschitz continuous; and $\|\boldsymbol{u}(t)-\boldsymbol{\eta}(t)\|\leqslant 1$, for $t\in[T_0,T_0+t_1]$, since $\boldsymbol{u}$ and $\boldsymbol{\eta}$ are continuous. Therefore $\psi(T_0)<1$ cannot hold.

Since $e_1(t)$ and $\sigma(t)$ are non-decreasing in t, it follows that $\psi(t)$ is strictly increasing in t, as soon as $e_1(t)>0$. Lemmas 3.1, 3.2 imply that σ is continuous and hence $\psi(t)$ is continuous. Also $\psi(t)\to\infty$ as $t\to\infty$. Thus it follows that $T_0\geqslant\overline{T}$, where $\overline{T}$ is the unique solution of

$$\psi(\overline{T})=1. \tag{5.11}$$

Note that, since $e_1(t)\to 0$ as $\Delta x\to 0$ and $N\Delta x\to\infty$ (with t fixed), $\overline{T}\to\infty$ as $\Delta x\to 0$ and $N\Delta x\to\infty$. Thus $\boldsymbol{u}$ exists on an interval $[0,T_0]$ that becomes arbitrarily large as $\Delta x\to 0$ and $N\Delta x\to\infty$. Moreover

$$\|\boldsymbol{u}(t)\|\leqslant\|\boldsymbol{u}(t)-\boldsymbol{\eta}(t)\|+\|\boldsymbol{\eta}(t)\|\leqslant 1+\sigma(t), \tag{5.12}$$

for $t\in[0,T_0]$. From now on we shall drop the distinction between T_0 and $\overline{T}$, and we shall think of $\overline{T}$ as the upper limit of the time interval over which the above estimates are valid. Although $\overline{T}\leqslant T_0$ the advantage of using $\overline{T}$ is that it is determined by the equation $\psi(\overline{T})=1$, whereas T_0 is not.

The above estimates are valid under the assumption that $h\in\mathscr{C}^1([0,\overline{T}])$. We shall now derive bounds for the temporal derivatives of $\boldsymbol{u}$ under the assumption that $h\in\mathscr{C}^k([0,\overline{T}])$, for some integer $k\geqslant 1$. These may be obtained directly from (4.13). Observe that $\mathscr{F}$ can be written in the form

$$\begin{aligned}\mathscr{F}(t,\boldsymbol{v}) &=\tilde{\boldsymbol{F}}[t,(\alpha h(t)+\tfrac{1}{2}\beta h^2(t),\alpha v_1+\tfrac{1}{2}\beta v_1^2,\ldots,\alpha v_N+\tfrac{1}{2}\beta v_N^2)]-\mu\tilde{\boldsymbol{G}}[h(t),v_1,\ldots,v_N]\\ &\equiv\boldsymbol{\Gamma}(t)+\boldsymbol{M}_F(\alpha\boldsymbol{v}+\tfrac{1}{2}\beta\boldsymbol{v}\circ\boldsymbol{v})-\mu\boldsymbol{M}_G\boldsymbol{v},\end{aligned} \tag{5.13}$$

where $\boldsymbol{\Gamma}(t) = \tilde{\boldsymbol{F}}[t, (\alpha h(t) + \frac{1}{2}\beta h^2(t), 0, \ldots, 0)] - \mu \tilde{\boldsymbol{G}}[h(t), 0, \ldots, 0]$ and $\boldsymbol{M}_F$, $\boldsymbol{M}_G$ are matrices such that $\|\boldsymbol{M}_F\| \leqslant c_F$ and $\|\boldsymbol{M}_G\| \leqslant c_G$, as defined in (5.4), (5.5). (We recall the notation for the product $\boldsymbol{u} \circ \boldsymbol{v}$, namely that $(u \circ v)_i = u_i v_i$, $i = 1, 2, \ldots, N$.) The vector $\boldsymbol{\Gamma}(t)$ is given by (cf. definition 5.2)

$$\Gamma_i(t) = s_{N+1-i}\left\{h'(t) + \frac{3\Delta x}{8\gamma}(\alpha + \tfrac{1}{2}\beta h(t)) - \mu\left[\gamma^{-1} + \frac{1}{12}\left(\frac{\Delta x}{\gamma}\right)^2\right]h(t)\right\} + \begin{cases} \dfrac{\Delta x}{24\gamma}(\alpha h(t) + \tfrac{1}{2}\beta h^2(t)), & \text{if } i = 1, \\ 0, & \text{if } i \geqslant 2. \end{cases}$$

Thus $\|(\mathrm{d}/\mathrm{d}t)^k \boldsymbol{\Gamma}(t)\| \leqslant q_k(h^{(0)}(t), \ldots, h^{(k)}(t))$, where q_k is a quadratic polynomial with coefficients that are polynomials in α, β, $\gamma^{-\frac{1}{2}}$, μ and Δx, with numerical coefficients.

From this partitioning of $\mathscr{F}$ we see that

$$\begin{aligned} \|\dot{\boldsymbol{u}}(t)\| &\leqslant q_1(h(t), h'(t)) + [c_F(\alpha + \tfrac{1}{2}\beta\|\boldsymbol{u}(t)\|) + \mu c_G]\,\|\boldsymbol{u}(t)\|, \\ &\equiv Q_1(h(t), h'(t), \|\boldsymbol{u}(t)\|), \end{aligned}$$

where Q_1 is quadratic. Then, on differentiating (5.13) we have

$$\ddot{\boldsymbol{u}} = \dot{\boldsymbol{\Gamma}} + \boldsymbol{M}_F(\alpha\dot{\boldsymbol{u}} + \beta\boldsymbol{u} \circ \dot{\boldsymbol{u}}) - \mu\boldsymbol{M}_G\dot{\boldsymbol{u}},$$

so that

$$\begin{aligned} \|\ddot{\boldsymbol{u}}\| &\leqslant q_2(h(t), h^{(1)}(t), h^{(2)}(t)) + [c_F(\alpha + \beta\|\boldsymbol{u}(t)\|) + \mu c_G]\, Q_1(h(t), h^{(1)}(t), \|\boldsymbol{u}(t)\|), \\ &\equiv Q_2(h(t), h^{(1)}(t), h^{(2)}(t), \|\boldsymbol{u}(t)\|). \end{aligned}$$

In this manner it can be shown inductively, together with the estimate (5.12) for $\|\boldsymbol{u}(t)\|$, that

$$\|(\mathrm{d}/\mathrm{d}t)^k \boldsymbol{u}(t)\| \leqslant Q_k(h^{(0)}(t), \ldots, h^{(k)}(t), 1 + \sigma(t)), \tag{5.14}$$

where Q_k is a polynomial of degree at most $k+1$, and $t \in [0, \overline{T}]$. Also, it follows that $(\mathrm{d}/\mathrm{d}t)^k \boldsymbol{u}(0) = \boldsymbol{0}$ if $h(0) = \ldots = h^{(k)}(0) = 0$, $k \geqslant 1$, and that

$$\max\{\|(\mathrm{d}/\mathrm{d}t)^k \boldsymbol{u}(t)\| : t \in [0, \overline{T}]\} \leqslant Q_k(h_m^{(0)}(\overline{T}), \ldots, h_m^{(k)}(\overline{T}), 1 + \sigma(\overline{T})).$$

Comment. Bounds for these temporal derivatives can also be obtained from (5.9) and (5.10). Proceeding from that starting point, estimates can be obtained showing that $\|(\mathrm{d}/\mathrm{d}t)^k(\boldsymbol{u} - \boldsymbol{\eta})(t)\| \to 0$ for any k, t, when $\Delta x \to 0$ and $N\Delta x \to \infty$.

5.4. *Bounds for the fully discrete problem*

Having shown in §5.3 that the semi-discrete approximation $\boldsymbol{u}$ is close to η we shall now consider the fully discrete approximation as effected by the prediction–correction method (4.15). The following proposition is a direct adaptation of the results given in Isaacson & Keller (1966, see p. 388 ff.).

PROPOSITION. *Let $\overline{T}$, $\Delta t > 0$ and let $\|\cdot\|$ be any norm on $\mathbb{R}^N$. Suppose that $\boldsymbol{y} = \boldsymbol{y}(t) \in \mathscr{C}^5([-3\Delta t, \overline{T}], \mathbb{R}^N)$ is such that $\dot{\boldsymbol{y}} = \boldsymbol{f}(t, \boldsymbol{y})$ on the interval $[-3\Delta t, \overline{T}]$, that $\boldsymbol{y} \equiv \boldsymbol{0}$ on $[-3\Delta t, 0]$ and that $\boldsymbol{f}$ is Lipschitz continuous in $\boldsymbol{y}$, with constant K, namely $\|\boldsymbol{f}(t, \boldsymbol{u}) - \boldsymbol{f}(t, \boldsymbol{v})\| \leqslant K\|\boldsymbol{u} - \boldsymbol{v}\|$, for $\boldsymbol{u}$, $\boldsymbol{v} \in \mathbb{R}^N$ and for $t \in [-3\Delta t, \overline{T}]$. Let $\boldsymbol{y}^n$, $n \geqslant 1$, be determined by*

$$\left.\begin{aligned} \bar{\boldsymbol{y}}^n &= \boldsymbol{y}^{n-1} + \tfrac{1}{24}\Delta t\,(55\boldsymbol{f}^{n-1} - 59\boldsymbol{f}^{n-2} + 37\boldsymbol{f}^{n-3} - 9\boldsymbol{f}^{n-4}) + \Delta t\,\bar{\theta}^n, \\ \boldsymbol{y}^n &= \boldsymbol{y}^{n-1} + \tfrac{1}{24}\Delta t\,(9\bar{\boldsymbol{f}}^n + 19\boldsymbol{f}^{n-1} - 5\boldsymbol{f}^{n-2} + \boldsymbol{f}^{n-3}) + \Delta t\,\theta^n, \end{aligned}\right\} \tag{5.15}$$

where $\boldsymbol{f}^j \equiv \boldsymbol{f}(j\Delta t, \boldsymbol{y}^j)$, $\bar{\boldsymbol{f}}^j \equiv \boldsymbol{f}(j\Delta t, \bar{\boldsymbol{y}}^j)$ *and* $\boldsymbol{y}^0 = \boldsymbol{y}^{-1} = \boldsymbol{y}^{-2} = \boldsymbol{y}^{-3} = \boldsymbol{0}$. *Suppose that the errors* $\boldsymbol{\theta}^n$ *and* $\bar{\boldsymbol{\theta}}^n$ *are such that*

$$\|\boldsymbol{\theta}^n\| + \tfrac{3}{8}K\Delta t\|\bar{\boldsymbol{\theta}}^n\| \leqslant \theta \quad \text{for} \quad n \leqslant \bar{T}/\Delta t.$$

Then, for all $n \leqslant \bar{T}/\Delta t$, *it follows that*

$$\|\boldsymbol{y}^n - \boldsymbol{y}(n\Delta t)\| \leqslant [(b_3 + b_4 K\Delta t)\,\Delta t^4 \sup_{t\in[0,\,T]} \|\boldsymbol{y}^{(5)}(t)\| + \theta]\,\frac{e^{c_d\bar{T}} - 1}{c_d},$$

with $b_3 = \frac{19}{720}$, $b_4 = \frac{251}{1920}$ *and* $c_d = \frac{1}{12}K(17 + 30K\Delta t)$.

To apply this proposition to the scheme (4.15) we shall use the l_∞ norm on $\mathbb{R}^N$ and let the errors $\bar{\boldsymbol{\theta}}^n$, $\boldsymbol{\theta}^n$ be

$$\left.\begin{aligned} \bar{\boldsymbol{\theta}}^n &= \left[\frac{1}{24}\sum_{j=1}^{4} \bar{a}_j h'((n-j)\,\Delta t) - \mathrm{d}h^{n-j}\right]\boldsymbol{z} \\ \text{and} \qquad \boldsymbol{\theta}^n &= \left[\frac{1}{24}\sum_{j=0}^{3} a_j h'((n-j)\,\Delta t) - \mathrm{d}h^{n-j}\right]\boldsymbol{z}, \end{aligned}\right\} \tag{5.16}$$

where $z_i = s_{N+1-i}$ (cf. (5.2)), $\mathrm{d}h^n$ is defined by (4.14),

$$(\bar{a}_1 = 55,\ \bar{a}_2 = -59,\ \bar{a}_3 = 37,\ \bar{a}_4 = -9) \quad \text{and} \quad (a_0 = 9,\ a_1 = 19,\ a_2 = -5,\ a_3 = 1).$$

Thus, $\boldsymbol{\theta}^n$ and $\bar{\boldsymbol{\theta}}^n$ can be estimated as

$$\|\boldsymbol{\theta}^n\|,\ \|\bar{\boldsymbol{\theta}}^n\| \leqslant c_{15}\Delta t^4 \sup\{|h^{(5)}(t)| : t/\Delta t \in [n-6, n+2]\},$$

and c_{15} is simply a numerical constant.

A necessary condition for $\boldsymbol{u}$ to be of class $\mathscr{C}^5$ is that $h \in \mathscr{C}^5$. Let us therefore assume that $h^{(0)}(0) = h^{(1)}(0) = \ldots = h^{(5)}(0) = 0$ and that $h \in \mathscr{C}^5([0, \bar{T} + 2\Delta t])$, where $\bar{T}$ is the solution of (5.11), and define $h(t) = 0$, $\boldsymbol{u}(t) = \boldsymbol{0}$ for $t < 0$. (The relationship of this assumption to the experimental situation is discussed in §6.2.) Then $\boldsymbol{u} \in \mathscr{C}^5(]-\infty, \bar{T}])$ and $\dot{\boldsymbol{u}}(t) = \mathscr{F}(t, \boldsymbol{u}(t))$ for all $t \in]-\infty, \bar{T}]$. Moreover,

$$\max\{\|\boldsymbol{\theta}^n\|, \|\bar{\boldsymbol{\theta}}^n\| : n \leqslant \bar{T}/\Delta t\} \leqslant c_{15}\Delta t^4 \sup_{t\in[0,\,\bar{T}+2\Delta t]} \{|h^{(5)}(t)|\}.$$

Since the Lipschitz estimate on $\mathscr{F}$ is not a global estimate the above proposition cannot be used directly for the scheme (4.15). But an argument similar to the one used to prove the existence of $\boldsymbol{u}$ in §5.3 can be used to show that the proposition is applicable to $\mathscr{F}$ over a time interval $[0, T_1]$ where $T_1 \to \infty$ as $\Delta t \to 0$. However, because we are interested in deriving *a posteriori* error estimates for σ we shall follow a different argument.

Let $\tilde{T} \leqslant \bar{T}$ and set $\tilde{\sigma}(\tilde{T}) = \max\{\|\boldsymbol{u}^n\|, \|\bar{\boldsymbol{u}}^n\| : n \leqslant \tilde{T}/\Delta t\}$. Note that $\tilde{\sigma}$ depends implicitly on Δt, Δx and N, but we shall view these as being fixed for now. Regard $\tilde{\sigma}$ as a quantity computed by the above method. Therefore $\tilde{\sigma}$ is known, at least *a posteriori*. Define

$$B(\tau) = \{\boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}\| \leqslant \max\{\tilde{\sigma}(\tau), 1 + \sigma(\tau)\}\}.$$

Thus, when $\tilde{T} < \bar{T}$, all the quantities $\boldsymbol{u}^n$, $\bar{\boldsymbol{u}}^n$ and $\boldsymbol{u}(t)$ belong to $B(\tilde{T})$ for t, $n\Delta t \in [0, \tilde{T}]$. Then, define $\boldsymbol{f}$ to be equal to $\mathscr{F}$ on $[0, \tilde{T}] \times B(\tilde{T})$ and such that $\boldsymbol{f}(t, \boldsymbol{v})$ is globally Lipschitz continuous in $\boldsymbol{v}$ (for $t \in [0, \tilde{T}]$), with a Lipschitz constant not exceeding that for $\mathscr{F}$ restricted to $B(\tilde{T})$. This is possible because the temporal dependence and the $\boldsymbol{v}$-dependence in $\mathscr{F}$ decouple (cf. (5.13)). In particular, a bound for the Lipschitz constant for $\boldsymbol{f}$ is afforded by

$$K(\tilde{T}) \equiv c_L(1 + 2\max\{\tilde{\sigma}(\tilde{T}), 1 + \sigma(\tilde{T})\}).$$

Since $\boldsymbol{u}^n$, $\tilde{\boldsymbol{u}}^n$ and $\boldsymbol{u}(t)$, for t, $n\Delta t \in [0, \tilde{T}]$, may be viewed equivalently as having been generated either by $\mathscr{F}$ or $\boldsymbol{f}$, the above proposition applies, yielding

$$\|\boldsymbol{u}^n - \boldsymbol{u}(n\Delta t)\| \leqslant c_{16}(1+\tfrac{3}{8}K(\tilde{T})\Delta t)\,\Delta t^4 \frac{e^{c_d\tilde{T}}-1}{c_d}\,[\sup_{t\in[0,\tilde{T}]}\|\boldsymbol{u}^{(5)}(t)\| + \sup_{t\in[0,\tilde{T}+2\Delta t]}|h^{(5)}(t)|],$$

for all $n \leqslant \tilde{T}/\Delta t$. Here, c_{16} is a numerical constant and

$$c_d = c_d(\sigma(\tilde{T}), \tilde{\sigma}(\tilde{T})) = \tfrac{1}{12}K(\tilde{T})[17 + 30K(\tilde{T})\Delta t].$$

Then, combining this estimate with (5.10) and (5.14), we have, for $0 \leqslant \tilde{T} \leqslant \overline{T}$ and for all $n \leqslant \overline{T}/\Delta t$, that

$$\|\boldsymbol{u}^n - \boldsymbol{\eta}(n\Delta t)\| \leqslant \psi(\tilde{T}) + c_{16}(1+\tfrac{3}{8}K(\tilde{T})\Delta t)\,\Delta t^4 \frac{e^{c_d\tilde{T}}-1}{c_d}$$
$$\times [Q_5(h_m^{(0)}(\tilde{T}), \ldots, h_m^{(5)}(\tilde{T}), 1+\sigma(\tilde{T})) + \sup_{t\in[0,\tilde{T}+2\Delta t]}|h^{(5)}(t)|] \equiv e_2. \quad (5.17)$$

The quantities $e_i(t) = e_i(h_m^{(0)}(t), \ldots, h_m^{(5)}(t), \sigma(t), \tilde{\sigma}(t), t, \Delta t, \Delta x, N\Delta x)$, $i \geqslant 2$, will be used to denote error expressions that tend to zero as Δt, $\Delta x \to 0$ and $N\Delta x \to \infty$. Thus, e_2 provides an estimate for the total error in the discrete scheme. In particular, for fixed $\tilde{T} > 0$, (5.17) shows that

$$|u_i^{\gamma_n} - \eta(i\Delta x, n\Delta t)| \leqslant c_T(\Delta t^4 + \Delta x^4 + e^{-rN\Delta x}),$$

for all $n \leqslant \tilde{T}/\Delta t$, for $i = 1, \ldots, N$ and for any r such that $0 < r < \gamma^{-\frac{1}{2}}$. The constant c_T is independent of Δt, Δx and N but depends on α, β, γ, μ, r and h, as well as on $\tilde{T}$, and a bound on $\tilde{\sigma}(\tilde{T})$ that is assumed to hold independently of Δt, Δx and N.

However, the above estimates have the shortcoming that the quantity $\sigma(\tilde{T})$ appears exponentially on the right-hand side and that the *a priori* bound (3.3) for σ allows the possibility of growth in time. To obviate the possibility of such large growth rates, we shall derive an *a posteriori* bound on σ based on our knowledge of $\tilde{\sigma}$. Estimate (5.17) and the mean-value theorem imply that

$$\begin{aligned}
&\max\{|\eta(x,t)| : x\in[0, N\Delta x], t\in[0,\tilde{T}]\}\\
&\quad\leqslant \max\{\|\boldsymbol{\eta}(n\Delta t)\| : 0 \leqslant n \leqslant \tilde{T}/\Delta t\}\\
&\qquad + \sqrt{2}\,(\Delta x + \Delta t)\max\{|\eta_x(x,t)| + |\eta_t(x,t)| : x\in[0, N\Delta x], t\in[0,\tilde{T}]\},\\
&\quad\leqslant \tilde{\sigma}(\tilde{T}) + e_2 + \sqrt{2}\,(\Delta x+\Delta t)\,[P_1(h_m^{(1)}(\tilde{T}), \sigma(\tilde{T}), \tilde{T}) + h_m^{(1)}(\tilde{T})\\
&\qquad + \gamma^{-\frac{1}{2}}(\alpha\sigma(\tilde{T}) + \tfrac{1}{2}\beta\sigma^2(\tilde{T})) + (3\mu/\gamma)\,\sigma(\tilde{T})],\\
&\quad\equiv \tilde{\sigma}(\tilde{T}) + e_3.
\end{aligned}$$

Then an upper bound for $|\eta(x,t)|$ for all $x \geqslant 0$ follows from lemma 3.2 and we have that

$$\begin{aligned}
\sigma(\tilde{T}) &\leqslant \tilde{\sigma}(\tilde{T}) + e_3 + ((\mu/\gamma)\,h_m^{(0)}(\tilde{T}) + h_m^{(1)}(\tilde{T}))\frac{e^{C\tilde{T}}-1}{C}\,e^{-rN\Delta x},\\
&= \tilde{\sigma}(\tilde{T}) + e_4. \qquad (5.18)
\end{aligned}$$

As in the definition (5.11) of $\overline{T}$, there is a unique $T_2 > 0$ such that

$$e_4(h_m^{(0)}(T_2), \ldots, h_m^{(5)}(T_2), 1+\tilde{\sigma}(T_2), \tilde{\sigma}(T_2), T_2, \Delta t, \Delta x, N\Delta x) = 1. \quad (5.19)$$

Furthermore, $T_2 \to \infty$ as Δt, $\Delta x \to 0$ and $N\Delta x \to \infty$, provided that $\tilde{\sigma}(t)$ remains bounded on bounded time intervals as these limits are approached. Thus, it follows from (5.18) that

$$\sigma(t) \leqslant 1 + \tilde{\sigma}(t) \quad \text{for} \quad t\in[0, T_2], \quad (5.20)$$

and $T_2 \leqslant \overline{T}$. Also we see that

$$|u_i^n - \eta(i\Delta x, n\Delta t)| \leqslant e_2(h_m^{(0)}(T), \ldots, h_m^{(5)}(T), 1+\tilde{\sigma}(T), \tilde{\sigma}(T), T, \Delta t, \Delta x, N\Delta x),$$

for $1 \leqslant i \leqslant N$, $n \leqslant T/\Delta t$ and $0 < T \leqslant T_2$, where e_2 is defined by (5.17) and T_2 is given by the solution to (5.19).

Thus, in summary, we have the following result.

THEOREM. *Let Δt and Δx be positive parameters not exceeding one. Let N be a positive integer and let $T > 0$. Suppose that $h^{(i)}(0) = 0$ for $i = 0, 1, \ldots, 5$, and that η is the solution to* (3.1). *Let $\boldsymbol{u}^n$ be the solution of* (4.15) *and let $\tilde{\sigma}(T) = \max\{|u_i^n|, |\bar{u}_i^n| : 1 \leqslant i \leqslant N, 1 \leqslant n \leqslant T/\Delta t\}$. If $T \leqslant T_2$, as defined by* (5.19), *then*

$$\begin{aligned}
&\max\{|\eta(i\Delta x, n\Delta t) - u_i^n| : 1 \leqslant i \leqslant N, 1 \leqslant n \leqslant T/\Delta t\} \\
&\quad \leqslant c_{12} P(h_m^{(1)}(T), 1+\tilde{\sigma}(T), T)\Delta x^4 \\
&\qquad + c_{14}(2+\tilde{\sigma}(T))\left[\frac{\mu}{\gamma}(1+\tilde{\sigma}(T)) + h_m^{(1)}(T)\right] \frac{e^{C(1+\tilde{\sigma}(T))T - rN\Delta x}}{C(1+\tilde{\sigma}(T))} \frac{e^{2c_L(2+\tilde{\sigma}(T))T} - 1}{2c_L(2+\tilde{\sigma}(T))} \\
&\qquad + c_{16}(1+\tfrac{3}{8}\tilde{K}(T)\Delta t) \frac{e^{\tilde{c}_d(T)T} - 1}{\tilde{c}_d(T)} [Q_5(h_m^{(0)}(T), \ldots, h_m^{(5)}(T), 2+\tilde{\sigma}(T)) + \sup_{t \in [0, T+2\Delta t]} |h^{(5)}(t)|] \Delta t^4,
\end{aligned}$$

where $\tilde{K}(T) = c_L(5+2\tilde{\sigma}(T))$ and $\tilde{c}_d(T) = \frac{1}{12}\tilde{K}(T)(17+30\tilde{K}(T)\Delta t)$. Here, $0 < r < \gamma^{-\frac{1}{2}}$; c_L, c_{12} c_{14} and c_{16} are constants (introduced previously) that depend only on α, β, γ, μ and r; P is defined in the proof of lemma 3.1 *and in* § 5.3; *C is defined in lemma* 3.2; *$h_m^{(i)}$, $i = 0, \ldots, 5$, are defined by* (3.10); *Q_5 is defined in* § 5.3 (*cf.* (5.14)).

Remarks. (i) The effects of round-off error can be incorporated into the above theorem as follows. Let the errors $\boldsymbol{\theta}^n$, $\bar{\boldsymbol{\theta}}^n$ of (5.15) include the rounding error associated with the computation of f^n, y^n, etc., at each time step. Suppose this additional error is bounded by θ_R (which will depend on N, Δx etc.). Then, from the proposition as stated, the final estimate in our theorem is modified simply by the addition of the term $\theta_R[e^{\tilde{c}_d(T)T} - 1]/\tilde{c}_d(T)$.

(ii) A consequence of the *a posteriori* estimate is that we can replace the bound σ by $1+\tilde{\sigma}$ wherever it appears in the preceding estimates. However, σ and $\tilde{\sigma}$ may be small with respect to one, say $O(\epsilon)$, and this replacement might not be a particularly good one. If we were to define $\overline{T}$, T_2 by the unique solutions to $\psi(\overline{T}) = \epsilon$ and $e_4(T_2) = \epsilon$ respectively, then $\sigma \leqslant \tilde{\sigma} + \epsilon$ on $[0, T_2]$ so that σ may be replaced by $\tilde{\sigma} + \epsilon$ wherever it occurs. In fact, we could define $\epsilon = \max_t \tilde{\sigma}(t)$ and then σ can be replaced by 2ϵ on $[0, T_2]$. Note that, regardless of the size of $\epsilon > 0$, $\overline{T}$ and $\overline{T}_2$ tend to infinity as Δx, $\Delta t \to 0$ and $N\Delta x \to \infty$.

6. EXPERIMENTAL APPARATUS AND PROCEDURE

6.1. *Experimental apparatus*

The experiments were made in a uniform channel of length 5.5 m and width 30 cm. One end of the channel was fitted with a plane beach of slope 1 in 10; at the other end there was a rigid plane flap which was used to generate the waves. The gap between the flap and the sides and bed of the channel was packed with foam plastic to restrict leakage past the wavemaker. In its rest position the flap was vertical and normal to the walls of the channel. It was supported by a horizontal shaft, the axis of which was normal to the walls of the channel at a height of about 1 m above the bed of the channel. The shaft was free to rotate about this axis. Since the water depth in the

channel was only 3 cm for these experiments, the action of the paddle was effectively equivalent to that of a plane piston. The paddle was forced in an oscillatory motion by a long crank attached to an eccentric on the shaft of a synchronous motor. Thus, the frequency and amplitude of the paddle motion were fixed for any given experiment and the arrangement was such that the paddle could be set oscillating almost instantaneously under these conditions.

The walls and bed of the channel were made from plate glass. The width of the channel was uniform to within 0.01 cm and the bed was levelled so that it deviated from a mean horizontal plane by no more than 0.040 cm. (The r.m.s. variation in depth from the mean was 0.020 cm.) The levelling of the tank can be important, as any unevenness in the bed gives rise to reflected waves, and systematic variations in depth lead to phase speeds different from those expected for a uniform channel. The walls of the channel were lined with an absorbent bandage to provide even wetting at the waterline.

Wave heights were measured by means of proximity transducers placed near the surface of the water. (Briefly, the principle of the instrument is that these transducers form one plate of a capacitor, the liquid surface being the second plate. By determining the capacitance it is possible to infer the distance of the water surface from the transducer.) The output from these transducers was relayed to an ultraviolet chart recorder, giving a continuous record of the surface elevation. The frequency response of the system extended from d.c. to about 1 kHz. Since the sensitivity and range of a given transducer is related to its area, we have, by choosing the appropriate transducer, recorded wave amplitudes ranging between 0.005 cm and 0.5 cm with about the same relative accuracy over the entire range. The wave heights thus determined were accurate to within about 2 % of the maximum recorded amplitude in any given run.

6.2. *Experimental procedure*

The tank was filled with water to roughly the desired depth, and surface films were skimmed off. The water was then topped up until the level was within 0.001 cm of a reference level, set by the tip of a pointer gauge. For all the experiments to be described the mean water depth was 3.00 cm (the main uncertainty deriving from the unevenness in the bed, see above). Several transducers (usually four) were then positioned along the channel, the distance of each transducer from the mean position of the wavemaker being known to within about 1 mm. Typically, the first transducer was placed about 15 to 20 cm from the wavemaker. On the basis of linear wavemaker theory (see Havelock 1929), we judged this distance to be well beyond the extent of the parasitic field of the wavemaker. The other transducers were then placed at distances of about 120, 220 and 320 cm from the wavemaker.

When the surface of the water in the tank was free of disturbances the wavemaker was set in motion, executing sinusoidal oscillations at a fixed amplitude and frequency, and the water elevation at each of the transducers was recorded. Because the first transducer was located well away from the wavemaker, any possible discontinuities in the wavefield arising from the start-up procedure were found to have been well 'smoothed out' at that position (cf. §7.3, figures 2 and 10). The experiment was stopped when the wavefront reached the beach at the far end of the channel. All experiments to be described here were made at a fixed period of 0.6930 s (i.e. $\omega_0 = 0.5014$) for the motion of the paddle, but the amplitude of the motion was changed from experiment to experiment by adjusting the throw on the driving crank.

Under the above conditions the theoretical wavelength of infinitesimal waves is 36.00 cm, giving a wavelength:depth ratio of 12:1. (The reasons for this choice are outlined in §2.5.) It is

instructive, then, to examine typical experimental conditions in relation to some of the theoretical assumptions for the model equations, as described in §2.1.

(*a*) The wave amplitude, ϵ, took values ranging between 0.002 and 0.2.

(*b*) The wavenumber, k_0, was nominally 0.5234. The main reason for requiring that k be 'small' is that the dispersion relations for the model equations should be good approximations to the dispersion relation derived from the full linear theory (see equations (2.4)). For $k = 0.5234$ the phase speeds ω/k for the three models are

model	exact	(KdV)	(M)
ω/k	0.9580	0.9543	0.9562

so that the error in the phase speed for infinitesimal waves, arising from the use of model M, is less than 0.2 % (but cf. the discussion in §7.6).

(*c*) The parameter S $(= \epsilon(\lambda/d)^2)$ took values between 0.4 and 36.

(*d*) The influence of surface tension is to increase the phase speeds by about 0.1 % (see Whitham 1974, p. 403), which is smaller than the differences indicated in (*b*) above.

6.3. *Comparison procedure*

The analogue data representing the wave profiles were recorded at a chart speed of 300 mm s^{-1} so that, in one period of the wavemaker, roughly 200 mm of chart paper moved past the marking beam. A discretization of this signal was made by measuring the wave amplitudes at 4 mm intervals. The peak-to-trough amplitude of the trace on the chart paper was adjusted to be about 60 to 70 mm (by suitably amplifying the output from the proximity gauge), and the displacement of the trace from its undisturbed position was measured to within about $\pm$ 0.3 mm. The above discretization corresponded to a temporal step of 0.2401 but preliminary tests suggested this would be too coarse for the degree of accuracy we would like for the numerical solutions. So, to use a time step of half this value, a (second-order) interpolation was made of the data obtained from the transducer nearest the wavemaker, and the resulting data set was then used as the boundary datum h for the numerical computation. The initial datum g was taken to be zero for all experiments.

If the theoretical solution at the location X is given by $\eta(X, t)$, $t \in [0, T]$, and the observed wave amplitude at the same position and over the same time interval is denoted by $v(X, t)$, let us define an error $E(j\Delta t)$, $j \in \mathbb{Z}$, between these two functions by

$$E(j\Delta t) = \sum_{i=j_+}^{M-j_-} |\eta(X, i\Delta t - j\Delta t) - v(X, i\Delta t)| \Big/ \sum_{i=j_+}^{M-j_-} |v(X, i\Delta t)|, \tag{6.1}$$

where $$j_+ = \max\{j, 0\}, \quad j_- = -\min\{j, 0\} \quad \text{and} \quad M\Delta t = T.$$

The local minimum of $E(j\Delta t)$ closest to zero was determined, say at $j = j_0$. The errors $E(j\Delta t)$ were then interpolated by a quartic polynomial $\tilde{E}(\tau)$, $\tau \in \mathbb{R}$, at the points $\tau = (j_0 + i)\,\Delta t$, $|i| \leqslant 2$. A minimum of $\tilde{E}$, denoted $\inf\{\tilde{E}\}$, was sought by Newton's method, the iteration being started at $\tau = j_0\Delta t$. (This procedure was successful in all cases.) This minimum value of $\tilde{E}$ gives essentially a measure of the difference in shape between the functions η and v, whereas the value of τ that realizes the minimum is effectively a phase error and can be used to provide a measure of the difference in speed of propagation of the two waveforms. Thus, η and v could have a very similar 'shape' but give a relatively large value for $E(0)$ by virtue of only a small 'phase' error. So, in making comparisons between theoretical and experimental data, it is useful to evaluate $E(0)$, $\inf\{\tilde{E}\}$ and the 'phase' error.

The numerical solutions used for the comparisons to be described in §7 were obtained on a Cyber 175 computer. With $\Delta t = 0.12005$ and $\Delta x = 0.15$, as used in the computations, the difference $E(0)$ between an exact solitary-wave solution of the model equation and the computed solution, under conditions comparable with those of the experiment, was about 0.1 %.

7. Experimental results

7.1. *Damping coefficient*

A determination of the damping along the channel was made from the steady wavefield established after the wavemaker had been working for a long time. Although this situation greatly simplifies measurements of the wavefield, it adds the complication of our having to identify the incident and reflected wave components. However, such a separation can be made without too much difficulty if there are no nonlinear effects present and if the waves are monochromatic. Indeed, for the same conditions as those used in the present experiments, Mahony & Pritchard (1980) measured a lapse rate $\eta^{-1}\eta_x$ of 0.38×10^{-2} at a wave amplitude of about 0.009. Before the present study another measurement of the decay rate was made at an amplitude of about 0.003 and this gave the same result as that found previously. Such a decay rate leads to a value for μ in (M*) of 0.014. (Recall that all the physical quantities are given in dimensionless form. Note also that the rather small numerical value of μ is merely a reflexion of the property that the term η_{xx} is larger than the terms η_{xxt} and $\eta\eta_x$ by a factor of order $\epsilon^{-\frac{1}{2}}$.)

7.2. *Two-dimensionality of the wavefield*

The magnitude of the cross-channel variations of the wavefield were measured to see by how much the assumption of two-dimensionality of the wave motions was violated. This measurement was made by placing two transducers at different positions across the channel, but at the same distance along the channel from the wavemaker, and the difference between the signals from each of the transducers was formed.

The most important cross-channel structure was that of a transverse wave motion, an example of which is given in figure 1. The waveform observed at the centre of the tank, at a distance $46.3d$ from the paddle, is shown in figure 1*a*, and the difference between the wave in the centre and that at a distance 5.9 cm from the side of the tank is shown in figure 1*b*. The transverse wave is seen to have an amplitude of about 4 % of that of the longitudinal wave and a frequency twice that of the forcing frequency of the wavemaker. This is roughly the scale of the transverse motions at each of the observation points, at all amplitude settings. By moving one transducer across the tank relative to the other, it was also found to be the scale representative of the size of the cross-channel variations. At the smaller wave amplitudes used in the experiments (ϵ less than about 0.01) a transverse motion was also evident, but the voltage differences between the two transducers were so small that they were only comparable with the noise level and it was therefore difficult to make any definitive statements about the frequency content of the transverse wave. However, the relative size of the transverse waves was certainly no greater than at the larger amplitudes.

The structure of the cross-channel motions was evidently complicated, being forced by the meniscus on the side walls or through the second harmonics of the longitudinal waves. We would expect (see, for example, Madsen 1974) the transverse motions to consist mainly of a mixture of wave modes of the form $\cos(m\pi y/b)$, where y is the cross-channel coordinate, b is the width of

the channel and m is a positive integer. Since waves at a frequency $2\omega_0$ satisfy the dispersion relation $\omega = (k \tanh k)^{\frac{1}{2}}$ at a value of $k \approx 1.2043$, corresponding to a wavelength of 15.65 cm here, it would appear that the modes most easily excited should have been those with $m = 3$ or 4. Waves with $m = 3$ would have been able to radiate along the tank, whereas those with $m = 4$ would have been decaying modes.

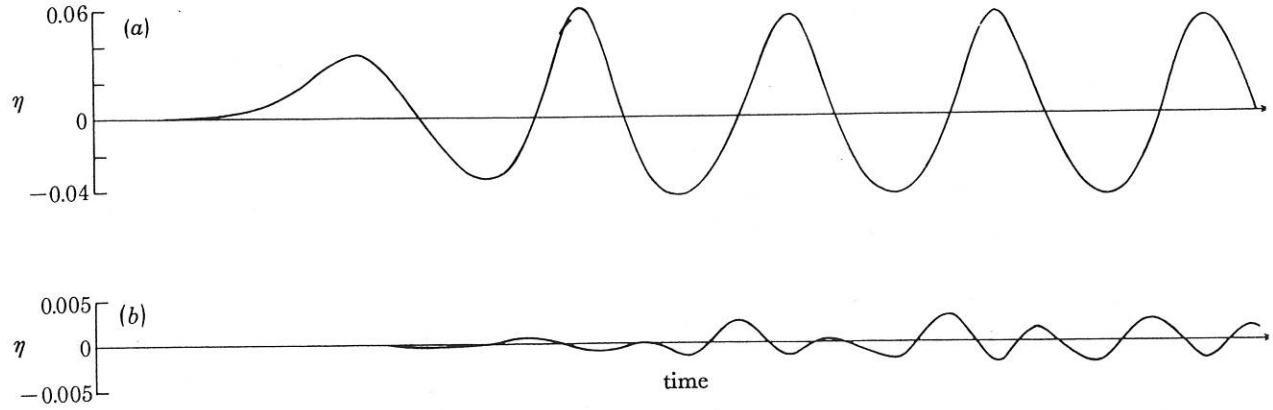

FIGURE 1. A tracing of the transducer voltage recorded at a distance $46.3d$ from the paddle. The scaling for the ordinate has been made dimensionless; the frequency of the paddle was 0.6930 s ($\omega_0 = 0.5401$). (*a*) The wave profile at the centre of the channel. (*b*) The difference between a transducer at the centre and one placed at a distance 5.9 cm from the side of the channel.

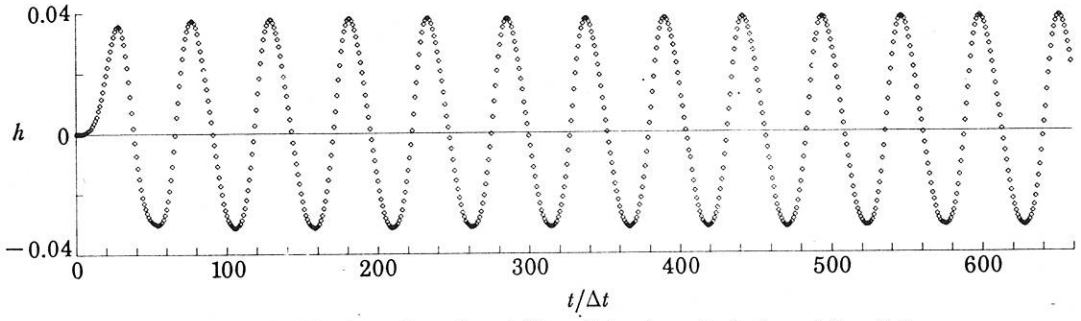

FIGURE 2. The boundary data $h(t)$ used for the calculation at $S = 5.5$.

7.3. *The main comparisons*

The main results of this study are summarized in table 4 and illustrated in figures 2–14. Several different kinds of tests have been made, as indicated in the table, but only a selection of the results are shown graphically. The eight experiments described in the table are defined by the parameter S, which ranged between 0.38 and 36. The 'stations' A, B, C are used to reference the locations of the transducer relative to the one used for the determination of $h(t)$, the actual distances between the two transducers being given in the column headed 'x'. The wave amplitude ϵ is taken to be $\sup\{|h|\}$. The column headed $-\ln x/\ln \epsilon_0$ indicates the position of the station expressed as a power of ϵ_0, where $\epsilon_0 = \max\{\epsilon, \delta^2\}$. Henceforth we shall refer to the station at which the boundary data $h(t)$ were measured as the 'boundary station'.

The comparisons given in columns I–III are the differences E and $\tilde{E}$, as defined in §6. The upper left-hand entry at each station is the difference $E(0)$ and the entry below that is $\inf\{\tilde{E}(\tau)\}$. The entry to the right indicates the 'phase error' τ at which the infimum of $\tilde{E}$ was realized, the error being expressed as a percentage of the time taken for a wave of speed 1.0 to reach the

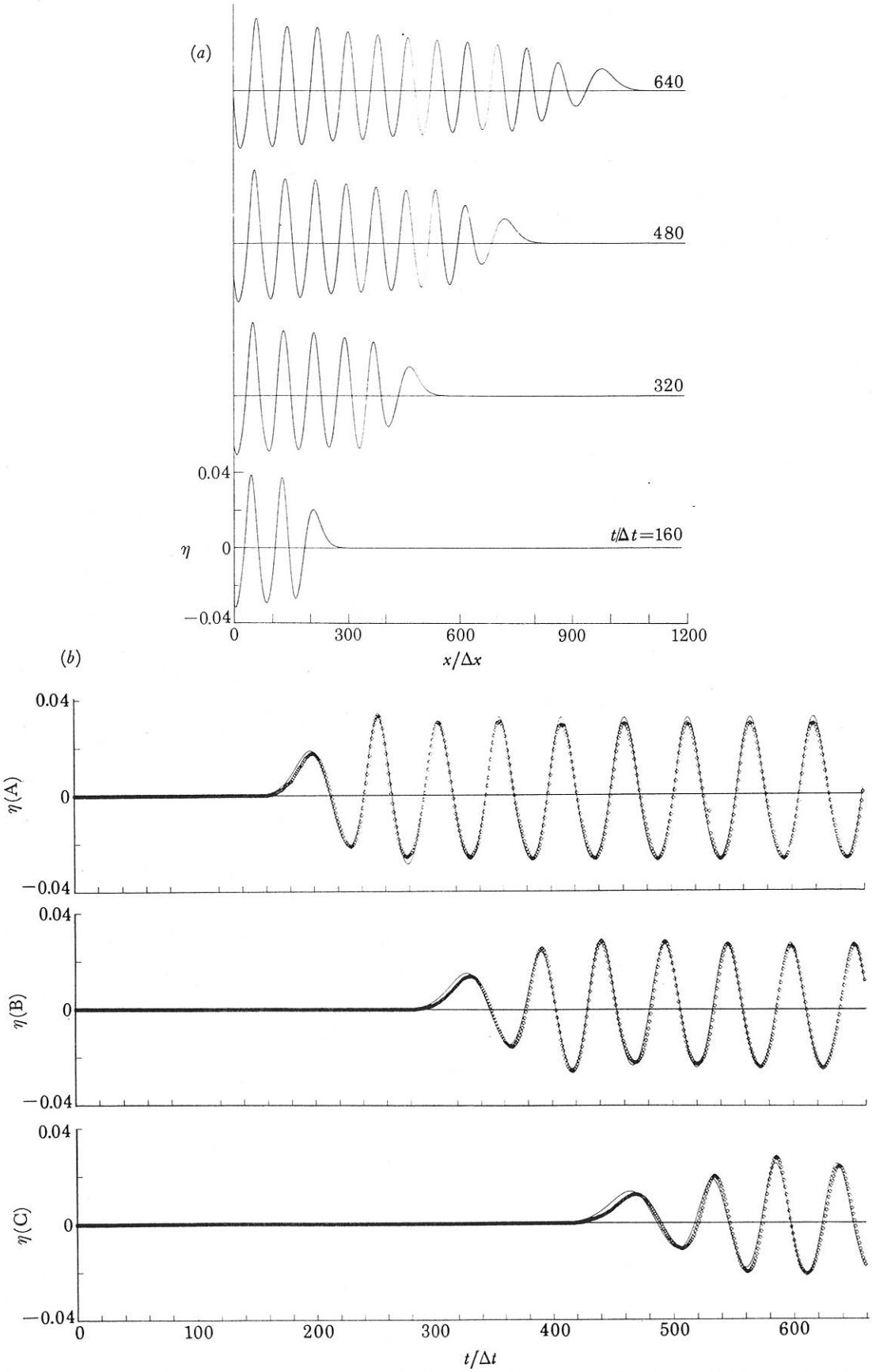

FIGURE 3. The experiment at $S = 5.5$ is compared with (M*) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$. (*a*) Computed amplitudes as a function of x. (*b*) Temporal comparisons at stations A, B, C.

TABLE 4. DETAILED SUMMARY OF COMPARISONS MADE BY USING THE MODEL (M*)

$\eta_t+\alpha\eta_x+\beta\eta\eta_x-\mu\eta_{xx}-\gamma\eta_{xxt}=0$, SUBJECT TO $\eta(0,t)=h(t)$, $\eta(x,0)=0$, $x\geqslant 0$, $t\geqslant 0$

The entries for each column are defined at the foot of the table. The errors in columns I–V, defined according to (6.1), are a comparison between the two sets of data indicated at the top of the appropriate column, with the abbreviations taking the following meanings:

Expt: experimental data; dissip. model: $\alpha=1$, $\beta=\frac{3}{2}$, $\mu=0.014$, $\gamma=\frac{1}{6}$; inviscid model: $\alpha=1$, $\beta=\frac{3}{2}$, $\mu=0$, $\gamma=\frac{1}{6}$; linear model: $\alpha=1$, $\beta=0$, $\mu=0.014$, $\gamma=\frac{1}{6}$.

The quantity p represents the error in the 'phase' speed given as a percentage of the 'long-wave' speed, $\sqrt{gd}$; $p>0$ means the computed speed exceeds the experimental value; $\epsilon_0=\max\{\epsilon,\delta^2\}$.

S	station	x	$\sup\{\lvert\eta\rvert\}$	$-\frac{\ln x}{\ln \epsilon_0}$	$\int_0^T \lvert\eta\rvert$	I expt / dissip. model		II expt / inviscid model		III expt / linear model		IV dissip. (M*) / inviscid (M*)	V (M*) / linear (M*)
0.38	B	70.9	0.0019	0.86	0.131	0.324 0.079	0.57	0.479 0.329	0.54	0.320 0.079	0.57	0.331	0.007
	C	103.8	0.0017	0.93	0.076	0.451 0.095	0.51	0.746 0.530	0.51	0.444 0.094	0.50	0.495	0.009
0.95	A	40.1	0.0053	0.74	0.336	0.100 0.098	0.06	0.287 0.286	0.04	0.101 0.097	0.06	0.185	0.017
	B	71.5	0.0045	0.86	0.208	0.103 0.092	0.06	0.429 0.426	0.07	0.099 0.092	0.06	0.338	0.017
	C	104.4	0.0043	0.94	0.108	0.191 0.106	0.22	0.549 0.509	0.22	0.183 0.110	0.20	0.480	0.022
4.5	A	40.1	0.0274	1.07	1.77	0.137 0.084	0.31	0.300 0.274	0.27	0.139 0.123	0.21	0.187	0.092
	B	71.5	0.0229	1.23	1.12	0.189 0.084	0.29	0.464 0.412	0.31	0.153 0.121	0.16	0.339	0.096
	C	104.5	0.0216	1.34	0.606	0.327 0.095	0.38	0.657 0.529	0.43	0.271 0.130	0.29	0.488	0.116
5.5	A	40.2	0.0335	1.13	1.98	0.109 0.080	0.26	0.262 0.247	0.19	0.132 0.131	0.06	0.186	0.111
	B	71.6	0.0285	1.31	1.27	0.113 0.081	0.14	0.379 0.360	0.15	0.115 0.114	−0.03	0.337	0.127
	C	104.4	0.0271	1.42	0.601	0.212 0.121	0.22	0.538 0.477	0.24	0.132 0.118	0.06	0.471	0.174
$\epsilon(\lambda/d)^2$	definition	x	$\sup_t\{\lvert\eta\rvert\}$	$-\frac{\ln x}{\ln \epsilon_0}$	$\int_0^T \lvert\eta\rvert$	$E(0)$ $\inf\{\tilde{E}\}$	p (%)	$E(0)$ $\inf\{\tilde{E}\}$	p (%)	$E(0)$ $\inf\{\tilde{E}\}$	p (%)	$E(0)$	$E(0)$

S	station	x	sup $\{\|\eta\|\}$	$-\frac{\ln x}{\ln \epsilon_0}$	$\int_0^T \|\eta\|$	I expt dissip.	I model	II expt inviscid	II model	III expt linear	III model	IV dissip. (M*) inviscid (M*)	V (M*) linear (M*)
11.8	A	39.3	0.0687	1.47	2.15	0.144		0.274		0.223		0.208	0.292
						0.097	0.32	0.230	0.34	0.210	−0.27		
	B	70.9	0.0590	1.70	0.954	0.186		0.431		0.255		0.357	0.398
						0.120	0.27	0.373	0.19	0.193	−0.34		
	C	103.8	0.0527	1.85	1.70	0.420		0.845		0.200		0.522	0.489
						0.107	0.51	0.546	0.66	0.189	−0.08		
18.1	A	39.3	0.105	1.77	2.88	0.164		0.279		0.431		0.236	0.368
						0.142	−0.23	0.272	−0.15	0.318	−1.01		
	B	70.9	0.0917	2.06	1.40	0.162		0.376		0.510		0.393	0.456
						0.162	−0.01	0.333	0.30	0.310	−0.80		
	C	103.8	0.0747	2.24	1.62	0.205		0.596		0.641		0.560	0.596
						0.188	−0.09	0.507	0.30	0.334	−0.66		
26.3	A	13.4	0.209	1.53	10.8	0.352		0.402		0.760		0.152	0.475
						0.346	−0.51	0.384	−0.73	0.645	−5.22		
	B	26.7	0.149	1.93	10.8	0.197		0.375		0.470		0.245	0.403
						0.193	−0.20	0.373	0.20	0.265	−1.94		
	C	39.6	0.153	2.16	8.92	0.302		0.452		0.642		0.284	0.551
						0.273	−0.47	0.449	−0.17	0.475	−1.70		
35.9	A	39.3	0.201	2.64	4.91	0.235		0.401		0.851		0.268	0.951
						0.221	0.21	0.318	0.63	0.523	−2.52		
	B	70.9	0.155	3.07	2.82	0.331		0.880		1.17		0.544	1.35
						0.128	0.53	0.358	1.18	0.447	−2.26		
	C	103.8	0.124	3.34	2.92	0.512		1.51		1.37		0.944	1.57
						0.136	0.63	0.580	1.46	0.540	−1.86		
$\epsilon(\lambda/d)^2$	definition	x	$\sup_t \{\|\eta\|\}$	$-\frac{\ln x}{\ln \epsilon_0}$	$\int_0^T \|\eta\|$	$E(0)$ inf $\{\tilde{E}\}$	p (%)	$E(0)$ inf $\{\tilde{E}\}$	p (%)	$E(0)$ inf $\{\tilde{E}\}$	p (%)	$E(0)$	$E(0)$

station. It is taken to be positive when the computed speed exceeded the experimental value. Column I shows the comparison between the experimental results and the wave amplitudes predicted by (M*). The second column shows the same kind of comparison, but no dissipative effects were included in the computations. For the results in the third column the nonlinear corrections were not included but dissipation was retained in the theoretical model.

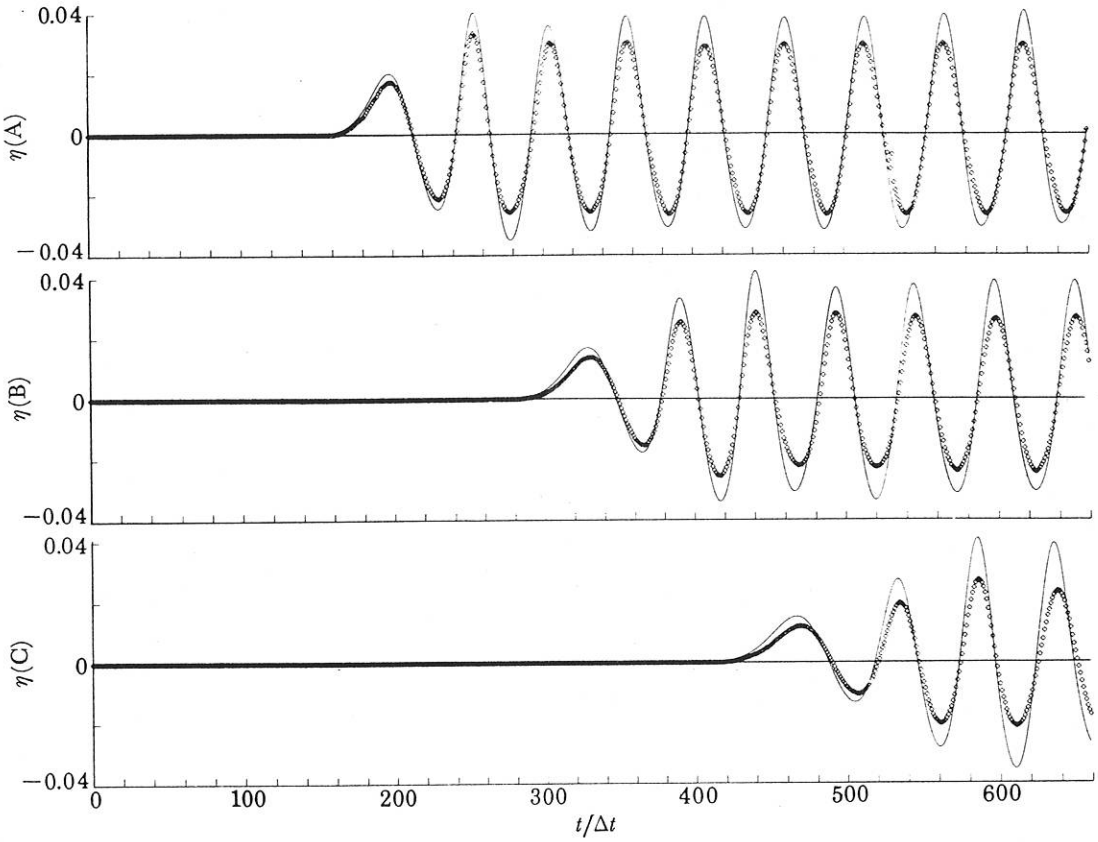

FIGURE 4. The experiment at $S = 5.5$ is compared with the inviscid version of (M*) ($\alpha = 1, \beta = \frac{3}{2}, \mu = 0, \gamma = \frac{1}{6}$).

The final two columns of the table show comparisons between different mathematical models so only the difference $E(0)$ is given. Thus, the penultimate column shows the difference between the solutions with and without dissipative effects included and the last column gives an indication of the importance of the nonlinear term.

In the figures the unit used for the temporal axes is the time step Δt and that used for spatial coordinates is Δx. The diamond-shaped symbols represent the experimental data, in their discretized form, and the continuous curves are piecewise linear segments linking the computed values of the wave amplitudes at the mesh points.

The graph shown in figure 2 is the discretized form of the function h used for the experiment at $S = 5.5$. The various comparisons for this experiment are shown in figures 3–5. Figure 3 shows computed wave amplitudes as a function of x at four different times, and figure 3 *b* gives the comparisons between the nonlinear, dissipative version of (M*) and the experimental results. As given in table 4, the relative differences $E(0)$ between the two functions were approximately 11 %, 11 % and 21 % at stations A, B, C respectively but, after allowances had been

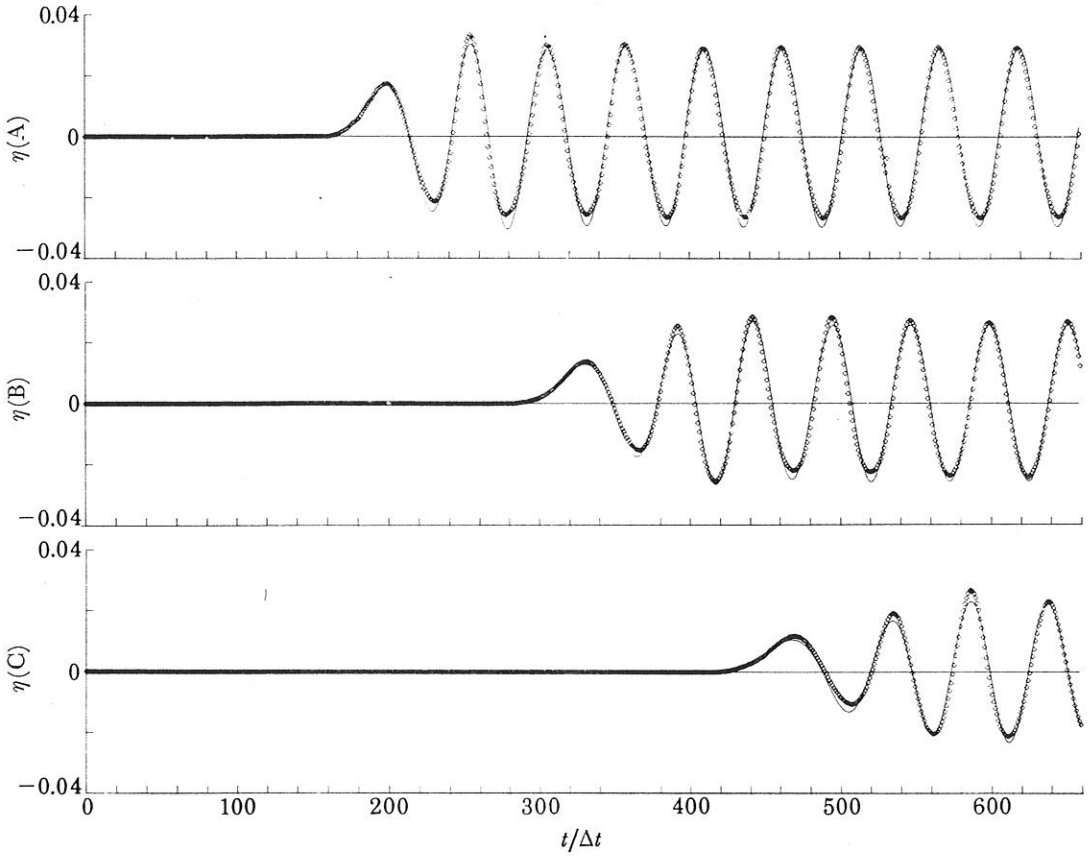

FIGURE 5. The experiment at $S = 5.5$ is compared with the linear version of (M*) ($\alpha = 1$, $\beta = 0$, $\mu = 0.014$, $\gamma = \frac{1}{6}$).

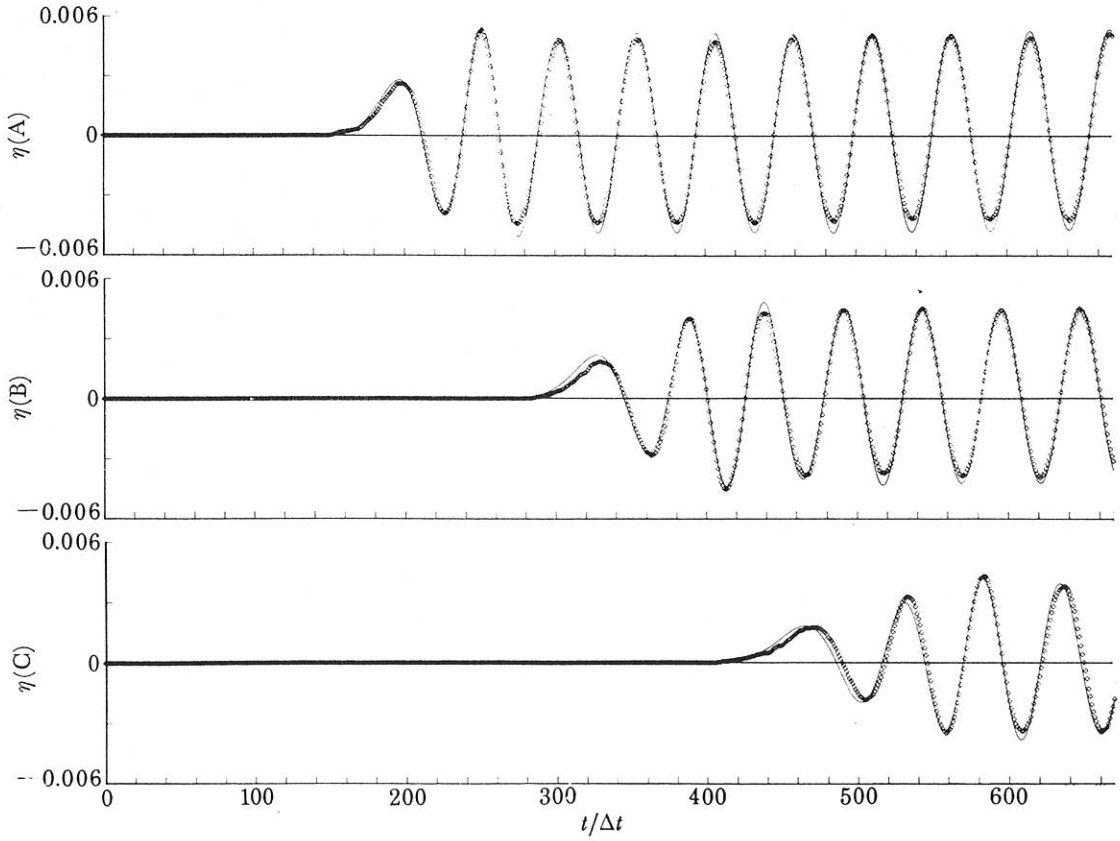

FIGURE 6. The experiment at $S = 0.95$ is compared with (M*) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$.

made for small phase-speed corrections of about 0.2 % these differences were reduced to about 8 %, 8 % and 12 % respectively. The importance of including the dissipative term is indicated by the results in figure 4, where the numerical solution is seen to differ markedly from the experimental results (cf. columns II and IV of table 4). On the other hand, the inclusion of the nonlinear term at this value of S is not so important, as shown in figure 5 (and see columns III, V of the table).

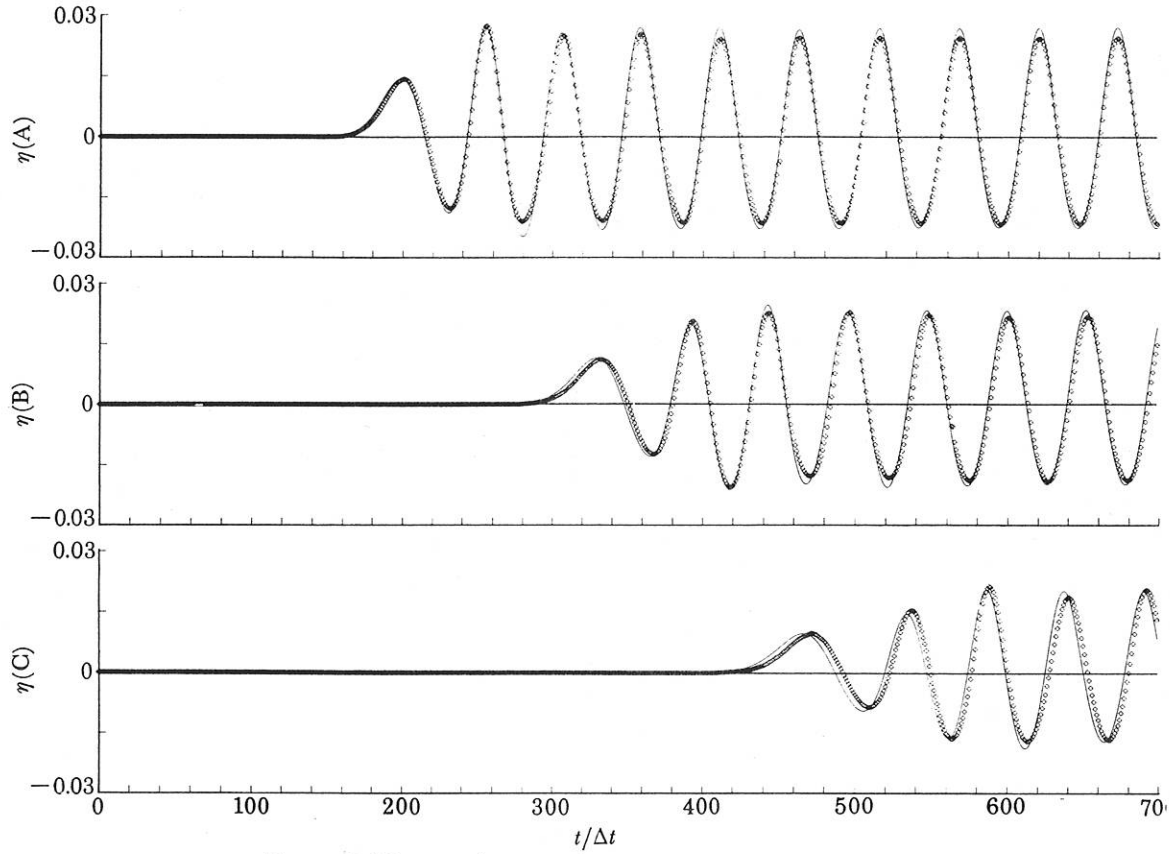

FIGURE 7. The experiment at $S = 4.5$ is compared with (M*) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$.

Note that, under these conditions, namely $S = 5.5$, the nonlinearity had the effect of modifying the waveform by about 17 % at a 'distance' $\epsilon^{-1.4}$ from the boundary station, whereas the dissipative effects had modified the waveform by 47 % at the same distance along the channel. The nonlinear effects seem to have brought about only a slight flattening of the wave troughs and sharpening of the crests, a feature that can be seen by comparing figures 3*b* and 5.

An experiment for which the nonlinear effects were of only very minor importance is shown in figure 6. In this experiment $S = 0.95$ and the nonlinear term affected the waveform by only about 2 %. The agreement between the theoretical prediction of (M*) and the observed waveform is not quite as good as for the results at $S = 5.5$, the main discrepancies apparently arising at the crests and troughs of the waves. Similar comparisons are shown in figure 7 (for $S = 4.5$), in figure 8 ($S = 11.8$) and in figure 9 (for the case $S = 18.1$). The experiment at $S = 11.8$ showed roughly the same kind of agreement as at the smaller values of S and this was confirmed by the

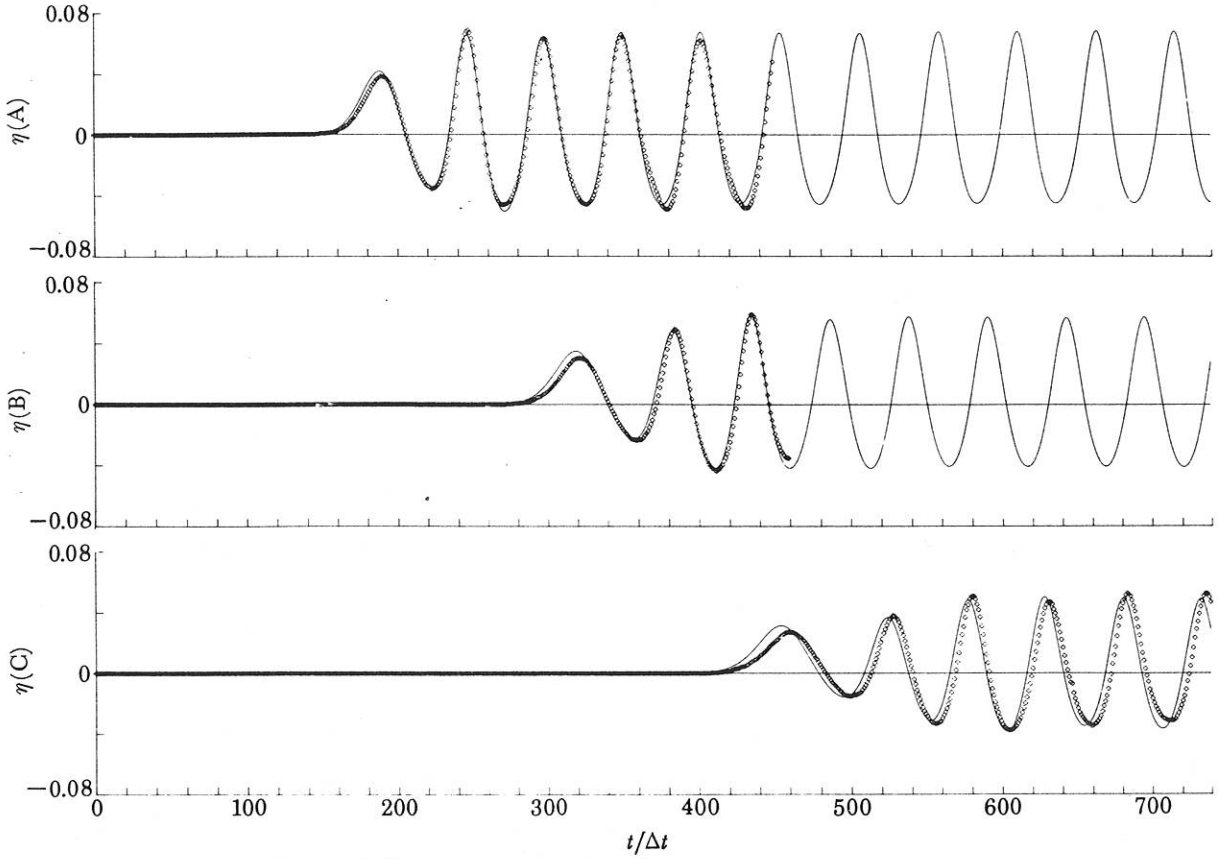

FIGURE 8. The experiment at $S = 11.8$ is compared with (M*) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$.

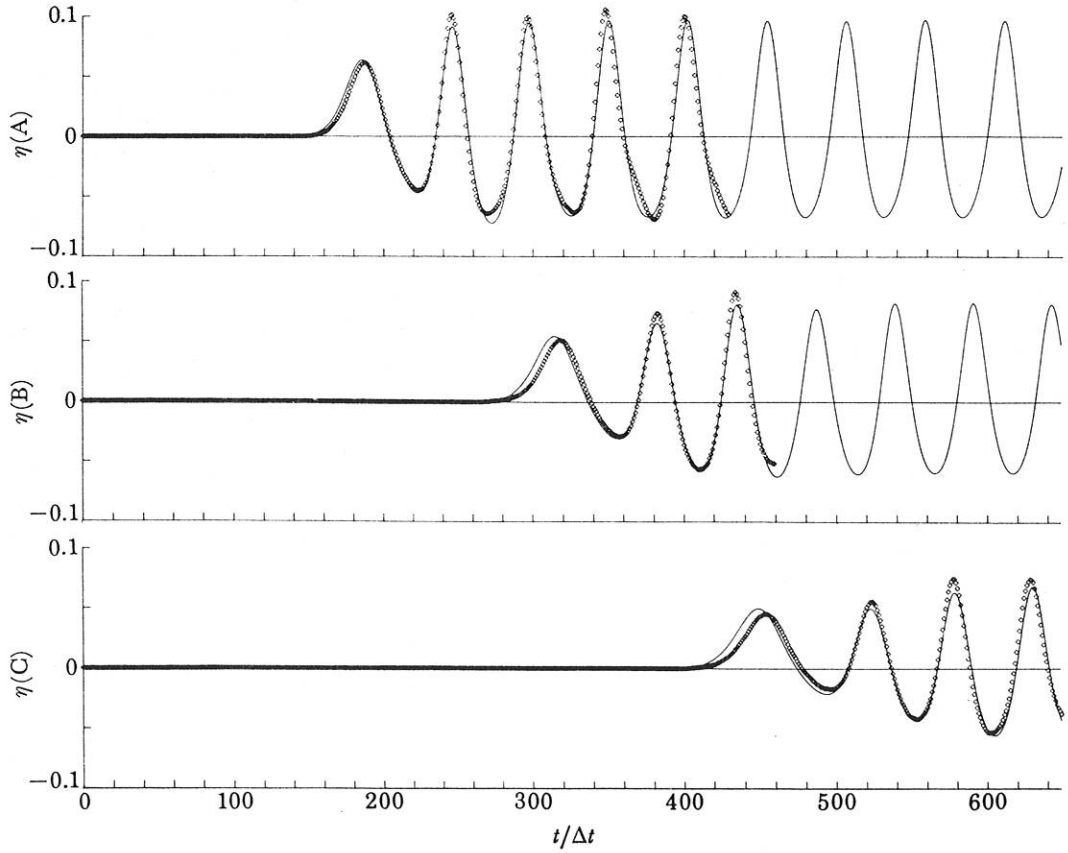

FIGURE 9. The experiment at $S = 18.1$ is compared with (M*) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$.

quantitative comparisons. For the experiments at $S = 4.5$ and 5.5 the nonlinear term had had only a small beneficial effect on the theoretical prediction of the observations but, at $S = 11.8$, the inclusion of the nonlinear term provided a significantly better model than the linear dissipative theory (cf. columns I, III of the table). On the other hand, the *inviscid* model gave a very poor representation of all these experiments. Thus, while there was some advantage to be gained from retaining the nonlinear term under these conditions, it was far more important that the dissipative effects be taken into account.

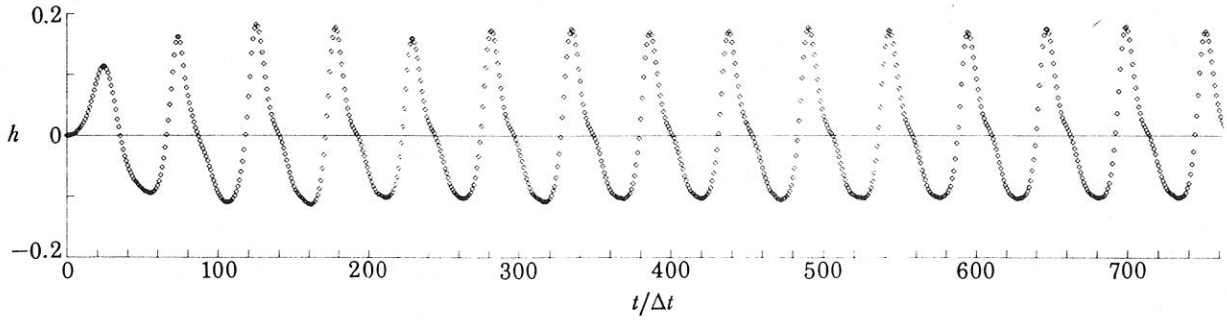

FIGURE 10. The boundary data $h(t)$ used for the calculation at $S = 26.3$.

The theoretical prediction of the experimental results at $S = 18.1$ was significantly worse than in the earlier cases. Whereas for all the previous experiments the difference $\inf\{\tilde{E}\}$ was less than about 10 %, it was about 15 % for the conditions at $S = 18.1$. One of the main reasons for the poorer agreement at $S = 18.1$ is that the theoretical speed of the leading wave appears to have been too large (see figure 9), with the result that the phase correction needed to minimize $\tilde{E}(\tau)$ was quite different from that found for the earlier experiments. The contribution from the nonlinear terms at $S = 18.1$, which was quite large, is indicated in column V of the table.

Two experiments at yet larger amplitudes were made, one at $S = 26.3$ and the other at $S = 35.9$. For the experiment at $S = 26.3$ the stations A, B, C were located much nearer the boundary station than in the other experiments so that they would not lie beyond the (formal) range of validity of the model equation. The form of the boundary data $h(t)$ for this experiment is given in figure 10, and the computed structure of the wavefield along the channel at four times is shown in figure 11*a*. The comparisons between the numerical solutions and the observed waveforms are shown in figures 11*b*–13. As indicated in the table, the agreement between the theoretical predictions and the experiment was not very close and the reason for this is apparent from the graphs. The experimental results indicate the presence of a substantial amount of second-harmonic component which is not nearly so strongly evident in the theoretical solutions of figure 11*b*. (In retrospect, this property is also evident in the results shown in figure 9 ($S = 18.1$), and figure 8 ($S = 11.8$).) At station B the agreement is seemingly much better than at the other stations, but the reason for this appears to be that the phase of the second harmonic is such that it reinforces the trough and diminishes the crest of the observed waveform and so the agreement is probably fortuitous.

The experiment at $S = 35.9$ gave similar kinds of comparisons (see figure 14) to those shown for the experiment at $S = 26.3$.

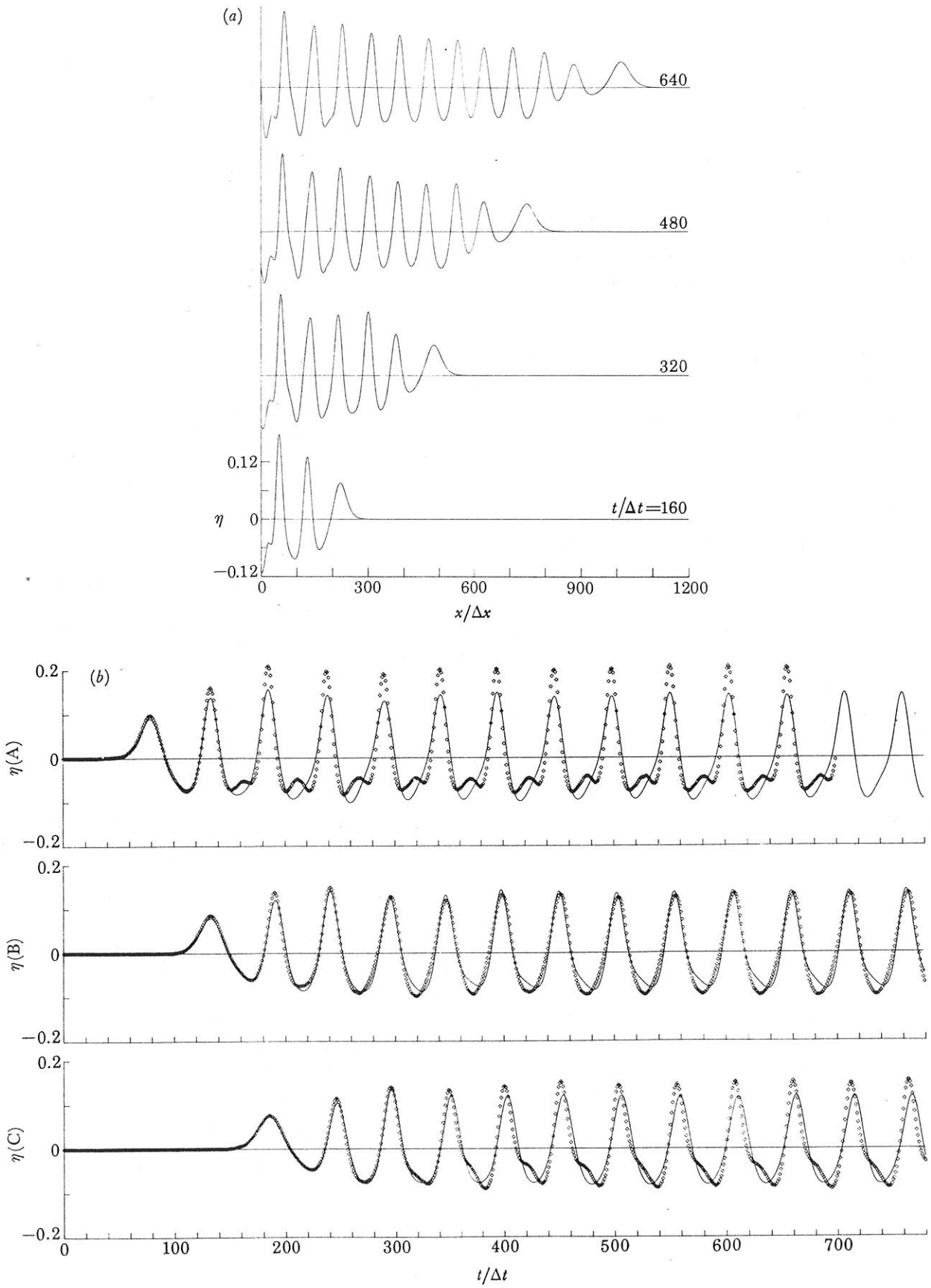

FIGURE 11. The experiment at $S = 26.3$ is compared with (M*) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$. (*a*) Computed amplitudes as a function of x. (*b*) Temporal comparisons.

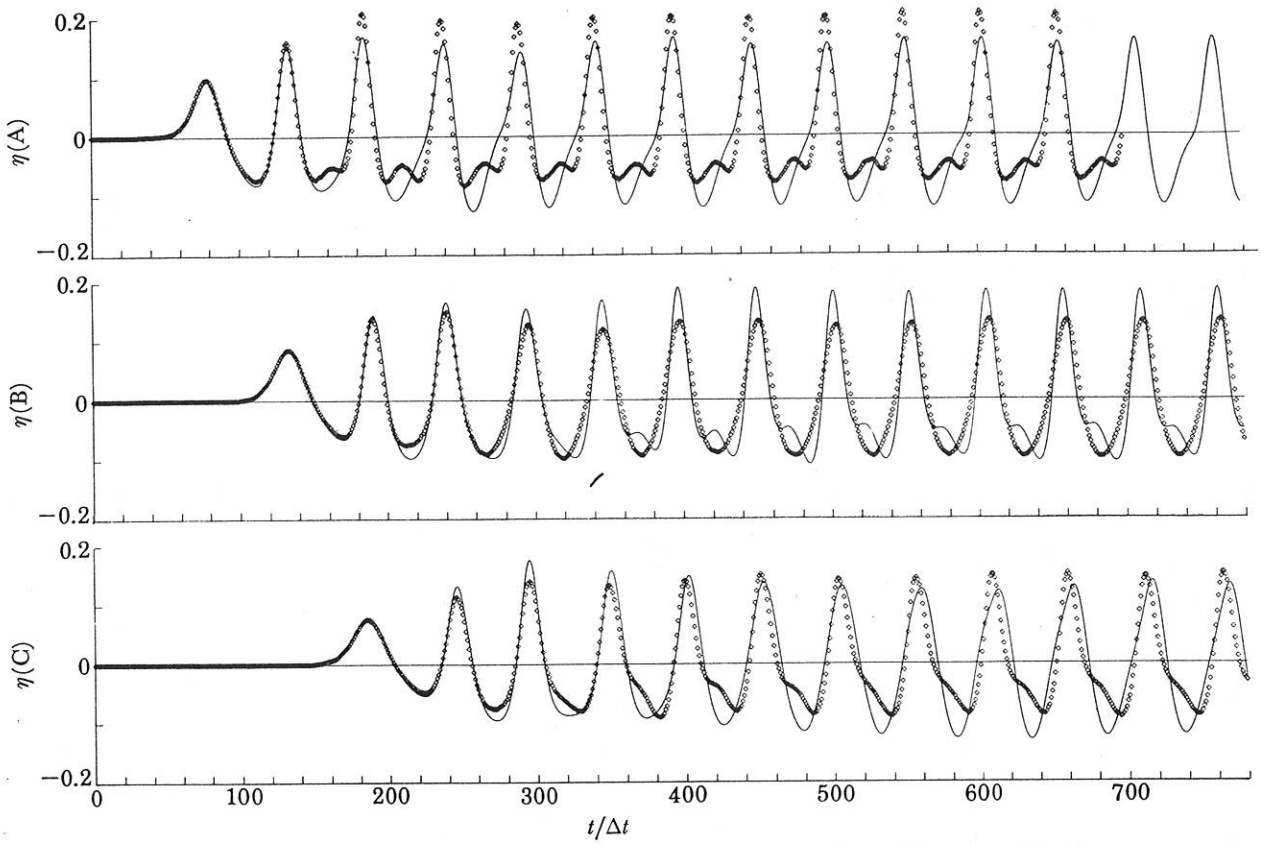

FIGURE 12. The experiment at $S = 26.3$ is compared with the inviscid version of (M*) ($\alpha = 1, \beta = \frac{3}{2}, \mu = 0, \gamma = \frac{1}{6}$).

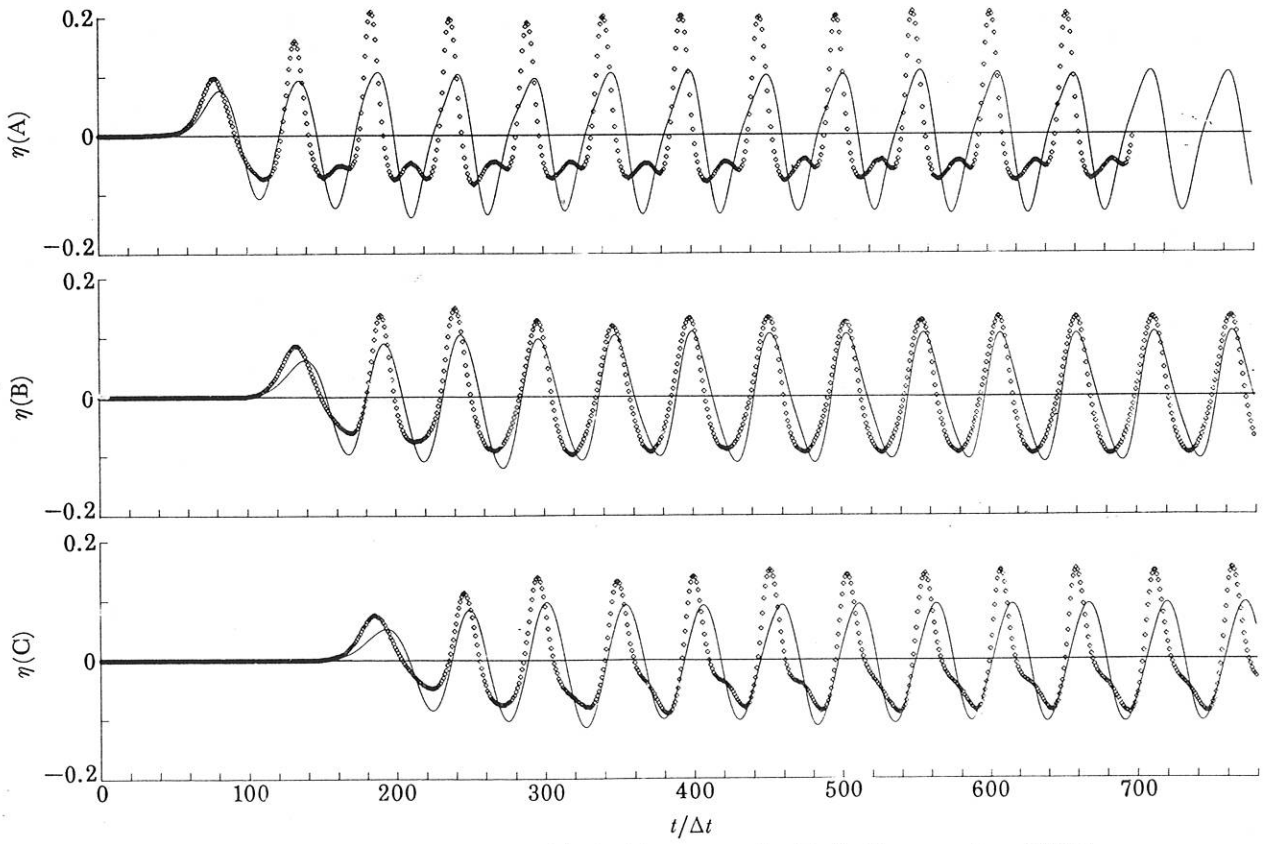

FIGURE 13. The experiment at $S = 26.3$ is compared with the linear version of (M*) ($\alpha = 1, \beta = 0, \mu = 0.014, \gamma = \frac{1}{6}$).

7.4. *Assessment*

The model appears to have given a fairly good description of the experiments at the smaller values of S, the differences being about 8–10 %. To give more meaning to these comparisons it is worth while to examine some of the sources of error. There are two kinds of error involved: one arising from uncertainties in making the physical measurements and the other from not matching accurately the assumptions on which the model is based. For the present experiments, uncertainties in the physical measurements were not more than 2 %, but since quantitative estimates of the other errors are not so easily made we shall attempt only a rough assessment of them.

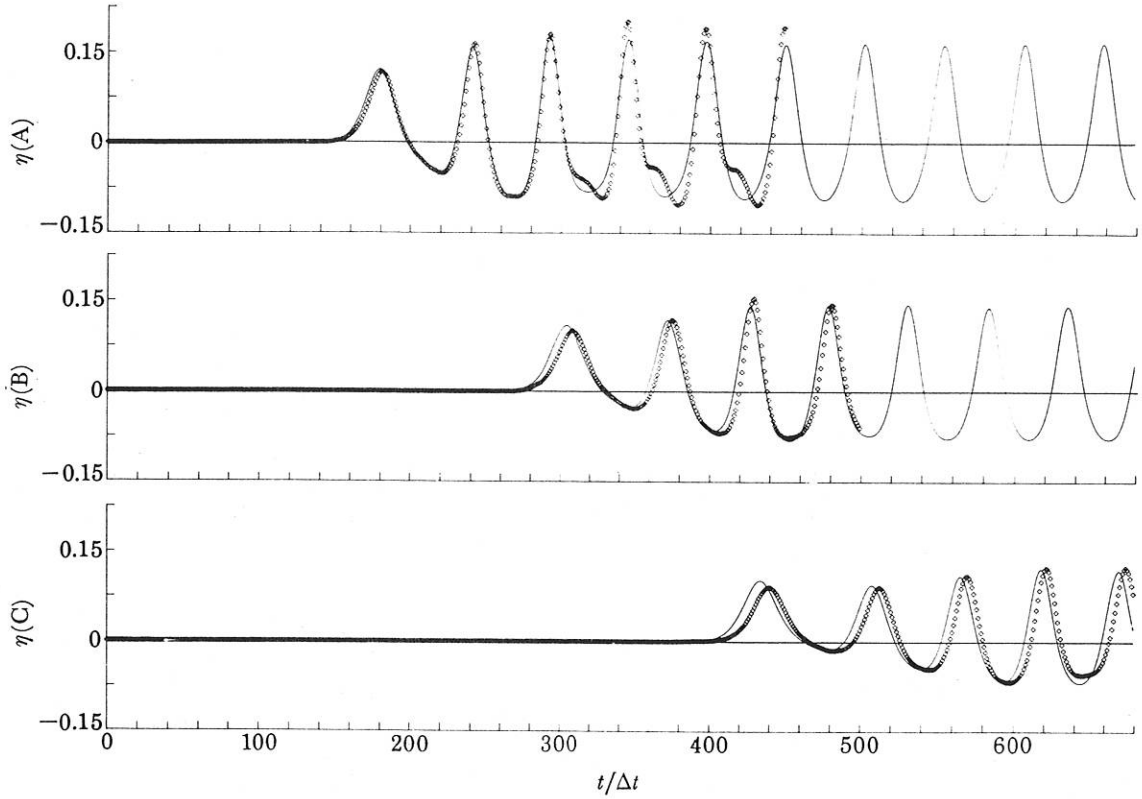

FIGURE 14. The experiment at $S = 35.9$ is compared with (M*) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\mu = 0.014$, $\gamma = \frac{1}{6}$.

The non-uniformity of the waves across the channel (cf. §7.2) was of the order of 4 or 5 % of the wave amplitude. This feature could influence the results both through the inaccuracy of representing the initial data $h(t)$ and through the error in making the comparisons at each of the stations A, B, C. In addition, there are uncertainties in the representation of the dissipative effects and deficiencies arising from the use of a one-dimensional model. Thus it does not seem as though we could expect closer agreement than the 8–10 % found at the smaller values of S.

However, as S was increased, both the quantitative and the qualitative agreement between the experiments and the theory deteriorated, and it is of interest to ascertain why this should have been so. There appeared to be three possible causes for the discrepancy.

(i) The dissipative effects were poorly modelled.

(ii) The presence of a non-negligible transverse-wave component.

(iii) The dispersion relation $\omega = (k \tanh k)^{\frac{1}{2}}$ was not very closely approximated by (M*) for wavenumbers near k_1, where k_1 is the wavenumber corresponding to $2\omega_0$. Thus, although the phase speeds of waves with wavenumbers near k_0 were closely approximated, the phase speeds of the shorter wavelengths evident in the experimental results were inaccurately represented by the model, and this feature could account for some of the disparities.

Without developing new theory or undertaking new experiments, it is not easy to account for (i) and (ii). We have, however, tried to make an appraisal of our modelling of the dissipation (see §7.5) and it is our view that this was not the main source for the discrepancies. It is, on the other hand, relatively straightforward to test the importance of (iii) (see §7.6), and the tests suggest that this was a major source of weakness of the model with regard to the present experiments. A discussion of (ii) has been given in §7.2.

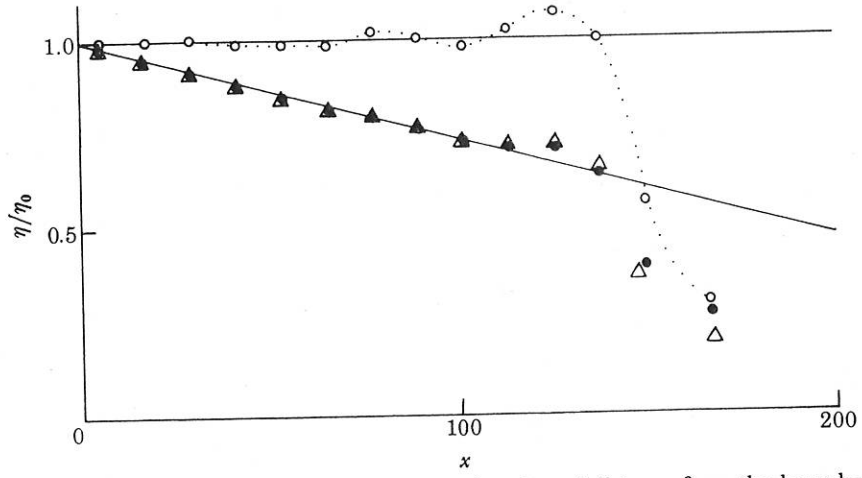

FIGURE 15. Computed amplitudes of wave crests as a function of distance from the boundary station for linear models ($\beta = 0$), with boundary data $h(t) = \eta_0 \sin \omega_0 t$. Time $t = 172.8$. ○, Inviscid model ($\nu = 0, \mu = 0$); ●, $\nu = 0$, $\mu = 0.01$; △, $\nu = 0.24 \times 10^{-2}$, $\mu = 0.11 \times 10^{-2}$. The slopes of the straight lines are $-(\nu + \mu k_0^2)$. The computations were made with $\Delta t = \Delta x = 0.15$.

7.5. *Modelling the dissipation*

The comparisons described in §7.3 indicate that the inclusion of dissipation is crucial if the model is to give a reasonable description of the experimental results. Therefore, in view of the discussion of §2.3, it seemed propitious to examine the sensitivity of the theory to different ways of modelling the dissipative effects.

In the comparisons of §7.3 the theoretical solutions gave a reasonably good account of the experimental results at small values of S, but at larger values of S the agreement was not so good. A possible explanation of this is that wavenumbers different from k_0 were being dissipated at an incorrect rate and, in particular, the harmonics were likely to have been considerably overdamped because the dissipation was taken to be proportional to k^2. This would certainly be the case if all the damping occurred in the boundary layers (cf. equation (2.5)). To test the sensitivity of the model to the way the dissipative effects were represented we have examined the consequences of using some alternative models for the damping. For this purpose we shall work from the *Ansatz* that the entire damping at wavenumber k is proportional to $|k|^{\frac{1}{2}}$, as suggested

by the boundary-layer theory, with the constant of proportionality, ρ_0, chosen to match the experimental decay rate at $k = k_0$. However it appears, at present, to be rather complicated to implement a numerical scheme to solve the initial- and boundary-value problem when the model equation includes the pseudo-differential operator whose symbol is $|k|^{\frac{1}{2}}$. Thus, for the purposes of this study, we have chosen to interpolate the function $\rho_0|k|^{\frac{1}{2}}$ by the polynomial $\nu+\mu k^2$. The interpolant used in §7.3, to be referred to as the $(0, k_0)$ interpolant, matched the magnitude of $\rho_0|k|^{\frac{1}{2}}$ at $k = 0$ and at $k = k_0$. But since this interpolant will dissipate waves with

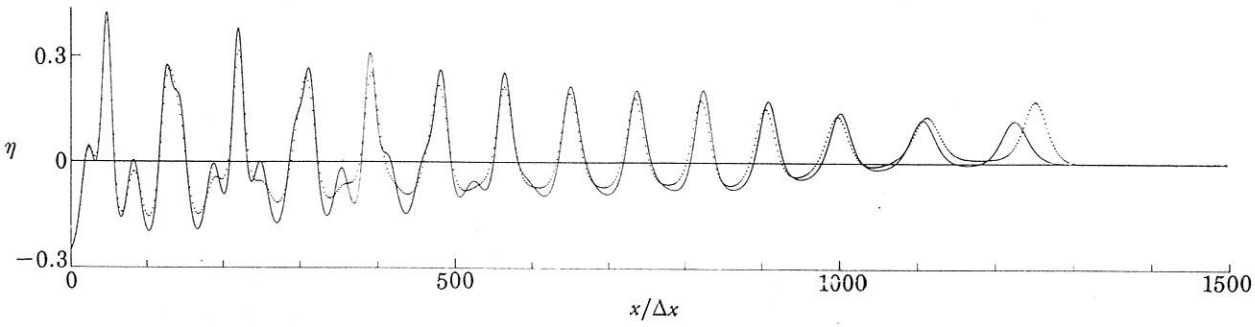

FIGURE 16. Computed wavefields at time $t = 172.8$ for nonlinear models ($\beta = \frac{3}{2}$) with boundary data $h(t) = 0.25 \sin \omega_0 t$. ····, $\nu = 0, \mu = 0.014$ ($(0, k_0)$ interpolation of $\rho_0|k|^{\frac{1}{2}}$); ——, $\nu = 0.340 \times 10^{-2}, \mu = 0.168 \times 10^{-2}$ ((k_0, k_1) interpolation). The computations were made with $\Delta t = \Delta x = 0.15$.

wavenumbers $k > k_0$ much faster than implied by $\rho_0|k|^{\frac{1}{2}}$, we have also considered (k_0, k_1)-interpolation of $\rho_0|k|^{\frac{1}{2}}$ where k_1 is the wavenumber corresponding to the frequency $2\omega_0$. This was done to provide a different representation of the damping of the wavemodes at the frequency $2\omega_0$ evident in the experimental results (for example, see figure 11*b*). (We have, incidentally, also examined the consequences of using Hermite interpolation of $\rho_0|k|^{\frac{1}{2}}$ by the function $\nu+\mu k^2$ at $k = k_0$; i.e. the magnitude and derivative of the functions were matched at k_0. But since the results were similar to those for the (k_0, k_1)-interpolation we shall not describe them here.) Note that the terms $\nu+\mu k^2$ in the dispersion relation correspond to the terms $\nu\eta - \mu\eta_{xx}$ in the differential equation.

A series of numerical experiments were made with boundary data of the form $h(t) = \eta_0 \sin \omega_0 t$. To check that the dissipative terms had been correctly coded, a preliminary test was made, with the linear model ($\beta = 0$) for which the decay rate is known theoretically. To estimate the decay rate along the channel from the computed solutions, the amplitudes of the wave crests were found at a given time ($t = 172.8$) and were plotted as a function of their distance from the boundary station. This graph (figure 15) shows that, except for a few crests near the front of the wavetrain, the amplitudes of the crests decreased at roughly the rate expected from the dispersion relation and that the two forms of dissipation gave similar results. For comparison, we have also included in the graph the results of the same experiment with no dissipative effects (i.e. $\nu = 0$, $\mu = 0$).

However, with $\beta = \frac{3}{2}$, the computed solutions differed significantly under the various representations of the dissipation, as illustrated in figure 16. This graph shows the computed wavefields, for $\eta_0 = 0.25$, at a time $t = 172.8$ corresponding roughly to the duration of a laboratory experiment. The comparison shown in this graph is that between the solutions obtained with the $(0, k_0)$-interpolation of $\rho_0|k|^{\frac{1}{2}}$ (dotted line) and with the (k_0, k_1)-interpolation of $\rho_0|k|^{\frac{1}{2}}$ (full line).

The substantial differences between these two solutions suggest that it could be very important to model the dissipative effects accurately.

To quantify the differences between these solutions we have evaluated the quantity $\mathscr{E} = \sum_{j=0}^{N} |\eta_1(j\Delta x, t) - \eta_2(j\Delta x, t)|\Delta x / \sum_{j=0}^{N} |\eta_1(j\Delta x, t)|\Delta x$, where η_1, η_2 are the functions being compared. Thus, for the comparison in figure 16, the difference $\mathscr{E} = 0.313$. A more complete list of comparisons, at various values of η_0 and at various times, is given in table 5. At small values of η_0 the differences were not too large, but with $\eta_0 = 0.1$ the differences had risen to about 10 %.

The solutions given in figure 16 suggest that the form of dissipation used for the comparisons of §7.3 probably dampened the larger wavenumbers too rapidly, which could account for the

TABLE 5. VALUES OF $\mathscr{E}$ WHEN $(0, k_0)$ INTERPOLATION OF $\rho_0|k|^{\frac{1}{2}}$ WAS COMPARED WITH (k_0, k_1) INTERPOLATION OF $\rho_0|k|^{\frac{1}{2}}$, FOR $h(t) = \eta_0 \sin \omega_0 t$

The computations were made with $\Delta t = \Delta x = 0.15$.

η_0	time 57.6	115.2	172.8
0.005	0.034	0.039	0.041
0.050	0.048	0.068	0.078
0.100	0.065	0.098	0.119
0.250	0.136	0.248	0.313

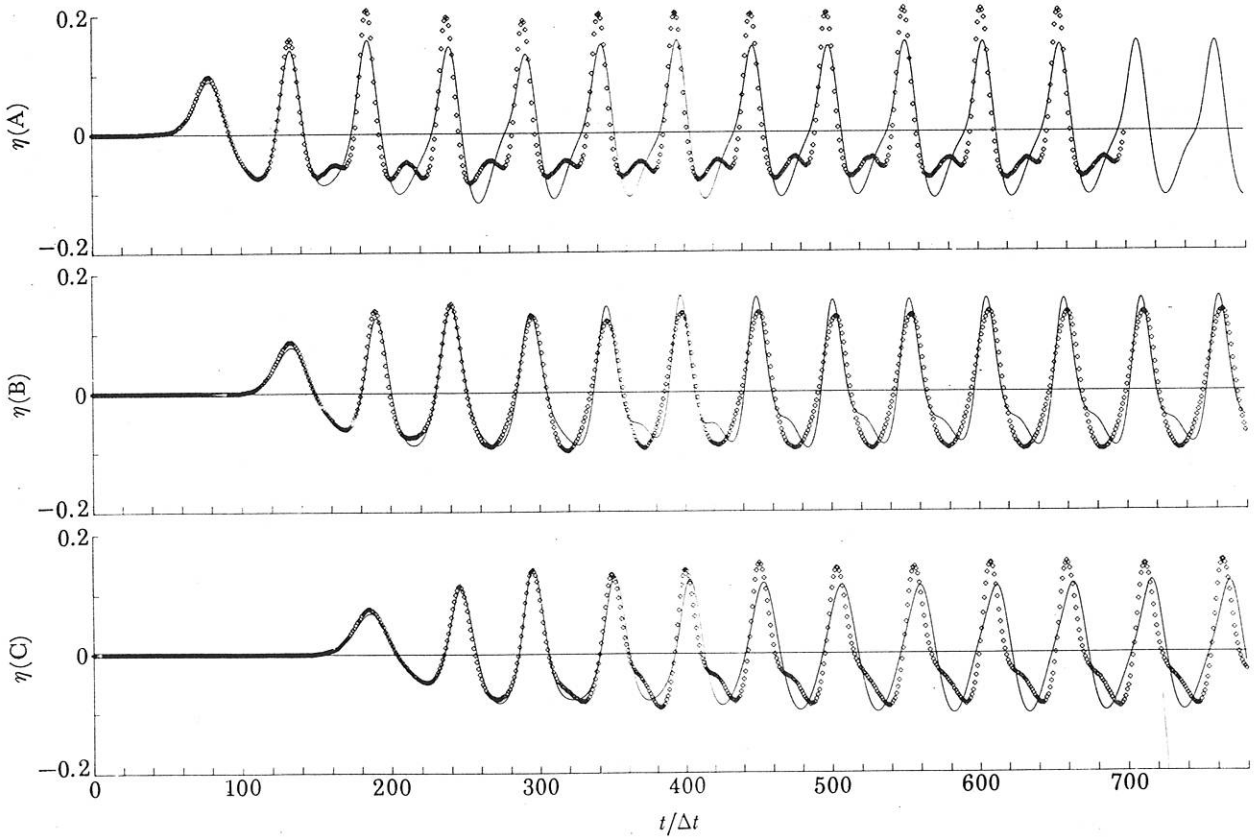

FIGURE 17. The experiment at $S = 26.3$ is compared with (M†) (see §7.6) when $\alpha = 1$, $\beta = \frac{3}{2}$, $\nu = 0.340 \times 10^{-2}$, $\mu = 0.168 \times 10^{-2}$, $\gamma = \frac{1}{6}$.

theoretical solutions not yielding the shorter wavelength components apparent in the experimental results. Such a possibility was checked, for the experiment at $S = 26.3$, by using the (k_0, k_1)-interpolation of $\rho_0 |k|^{\frac{1}{2}}$ to model the dissipative effects. A graph of the comparison is given in figure 17, the error $E(0)$ being 0.386, 0.283, 0.378 at the stations A, B, C respectively. These errors could be reduced to 0.368, 0.283, 0.363 with phase corrections of 0.84 %, 0.07 % and 0.90 % at the respective stations. The agreement is slightly worse here than in § 7.3. At stations A, C the amplitudes at the crests of the computed solutions were much smaller than those observed experimentally, whereas at station B the computed amplitudes of the crests were too large. But similar features to this were also evident in the solutions with no damping, i.e. $\nu = \mu = 0$ (see figure 12), and for $\nu = 0$, $\mu = 0.014$ (see figure 11 *b*), which suggested to us that the inaccurate model for the damping of the larger wavenumbers was probably not the main source of these discrepancies.

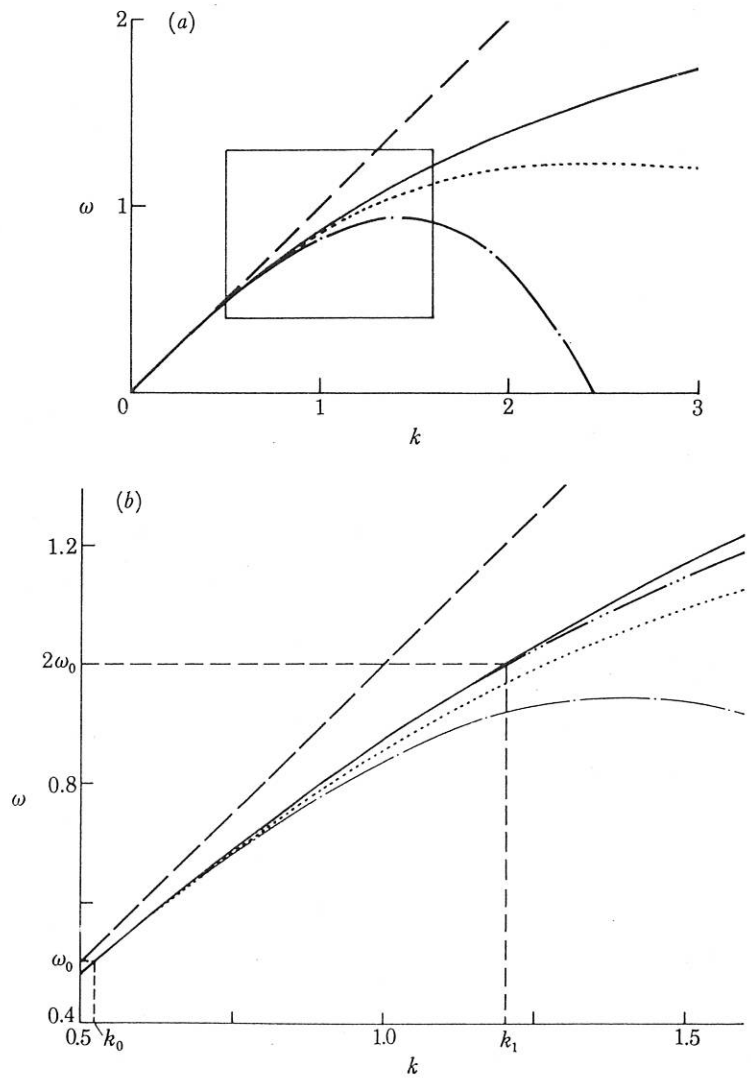

FIGURE 18. Graphs of the linear dispersion relations for various models. (*a*) – – –, Shallow-water model, $\omega = k$; ——, 'exact' relation, $\omega = (k \tanh k)^{\frac{1}{2}}$; - - - - -, model (M), $\omega = k/(1+\frac{1}{6}k^2)$; –·–·–, (KdV), $\omega = k(1-\frac{1}{6}k^2)$. (*b*) Magnified version of (*a*): – – –, $\omega = k$; ——, $\omega = (k \tanh k)^{\frac{1}{2}}$; –· · ·–, $\omega = 0.9898k/(1+0.1325k^2)$; ·····, $\omega = k/(1+\frac{1}{6}k^2)$; ——·——, $\omega = k(1-\frac{1}{6}k^2)$.

7.6. *The approximation to the dispersion relation*

First we examine the dispersion relations for the various models. These are shown graphically in figure 18 where the shallow-water model ($\omega = k$) and (M) (as well as (KdV)) are compared with the 'exact' relation $\omega = (k \tanh k)^{\frac{1}{2}}$. By construction, these relations are all close at small values of k; at $k = 0.5$ it is evident that the shallow-water approximation is a poor model and at $k = 1$ all three models give a poor approximation to $\omega = (k \tanh k)^{\frac{1}{2}}$.

However, for the wavenumbers arising in our experiments, the equation

$$\eta_t + \alpha\eta_x + \beta\eta\eta_x + \nu\eta - \mu\eta_{xx} - \gamma\eta_{xxt} = 0 \qquad (7.1)\ (\mathrm{M}\dagger)$$

can be used to provide a better interpolation of the 'exact' dispersion relation than that afforded by (M). This is achieved through a suitable choice of the parameters α, γ. Since the dominant wavenumbers appear to have been those corresponding to the frequencies ω_0 and $2\omega_0$, we have chosen α, γ so that the phase speeds for the linear form of (M†) (i.e. $\beta = 0$) coincided with those for the 'exact' theory at the wavenumbers k_0 and k_1. However, the displacement effects of the boundary layer lead not only to a damping of the waves but also to a correction in the phase speed of a wavemode (cf. equation (2.5)). Therefore, taking the boundary-layer correction to the dispersion relation to be of the form suggested by the theory of Kakutani & Matsuuchi (1975), we have chosen to interpolate the dispersion relation

$$\omega = (k \tanh k)^{\frac{1}{2}} - \rho_0(-1+i)\,|k|^{\frac{1}{2}},$$

with ρ_0 taken to be the empirical constant used for the comparisons in §7.3.

Under the conditions of the present experiments this interpolation gives $\alpha = 0.9898$, $\gamma = 0.1325$, for which values the real part of the dispersion relation for (M†) is shown in figure 18*b*, together with that for some of the other models. The theoretical solutions that result from the use of (M†) for the experiment at $S = 26.3$ are shown in figure 19. The spatial form of the wavetrain, which is given in figure 19*a*, shows a number of qualitative differences from that obtained with (M*) (cf. figure 11*a*), and the comparison between (M†) and the experimental results is given in figure 19*b*. This comparison also shows a qualitative improvement in the prediction of the experimental results over the comparisons given in figures 11*b* and 17. The quantitative comparisons for (M†), which gave differences for this experiment of 14 %, 16 % and 18 % at the stations A, B, C respectively, are summarized in table 6. Indeed, (M†) represents all the experimental results to within about 8 % except for those for $S = 26.3$ and $S = 35.9$.

A graph of the comparison at $S = 35.9$ is given in figure 20. The leading wave at each station is represented very well by the model (cf. the results of figure 14 for (M*), where this was not so), but the subsequent oscillations were modelled less accurately. To illustrate further the significant improvement obtained through the use of the form (M†) to describe the experimental results, additional comparisons made with the model are given in figures 21–25. Figure 21 shows the comparison at $S = 0.95$ (cf. figure 6). The comparison at $S = 5.5$ is given in figure 22 (cf. figure 3*b* for (M*)), and the comparison at $S = 18.1$ is given in figure 23. The influence of the different representations of dissipation that we have considered are shown in figures 24 and 25, for the experiment at $S = 26.3$. The theoretical solution shown in figure 24 is that for (M†) ($\alpha = 0.9898$, $\gamma = 0.1325$) with no dissipation, i.e. $\nu = 0$, $\mu = 0$, and the one given in figure 25 has $\nu = 0$, $\mu = 0.014$ (cf. figures 11*b*, 17 and 19*b*).

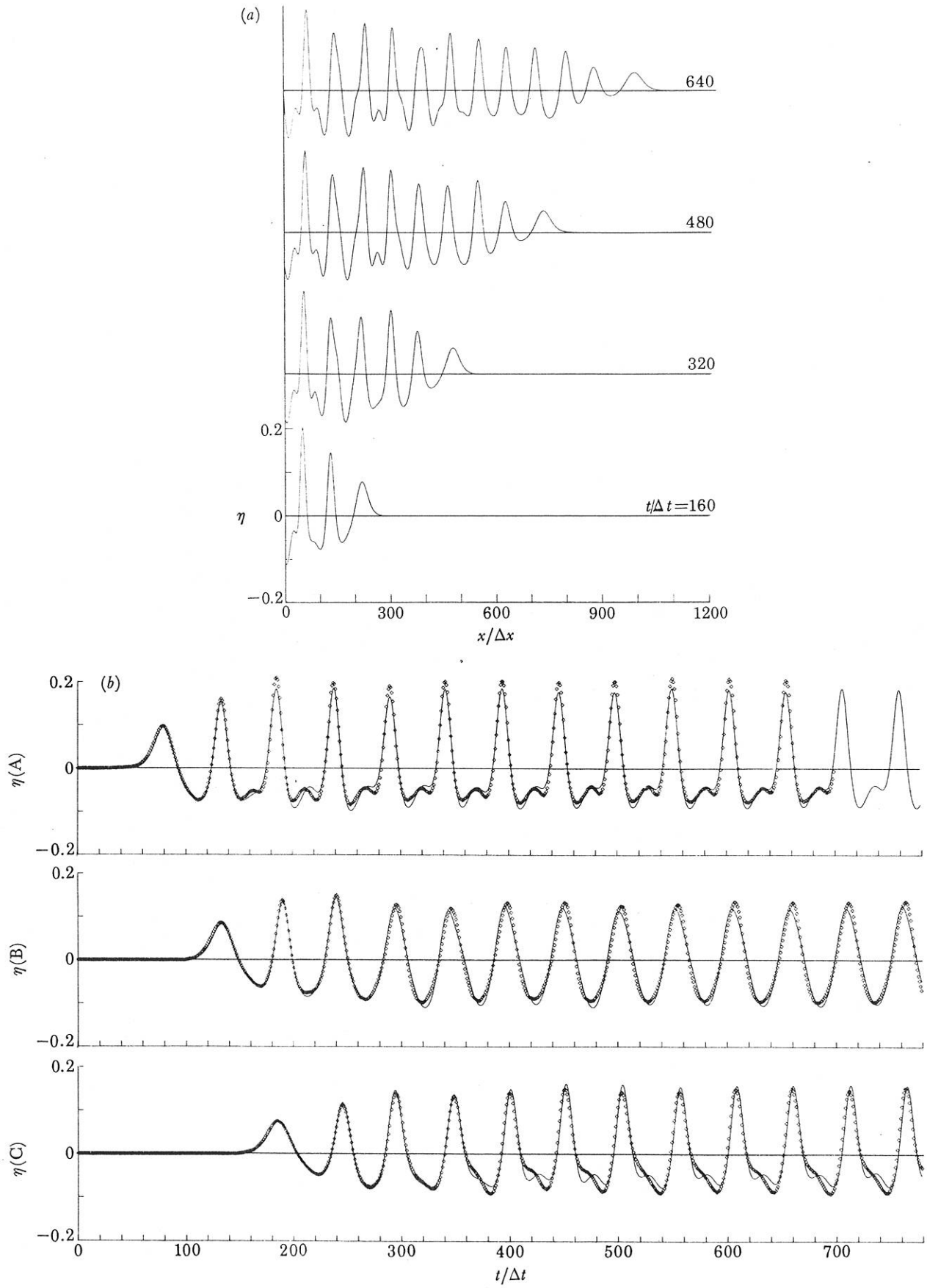

FIGURE 19. The experiment at $S = 26.3$ is compared with (M†) when $\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0.340 \times 10^{-2}$, $\mu = 0.168 \times 10^{-2}$, $\gamma = 0.1325$. (*a*) Computed amplitudes as a function of x. (*b*) Temporal comparisons.

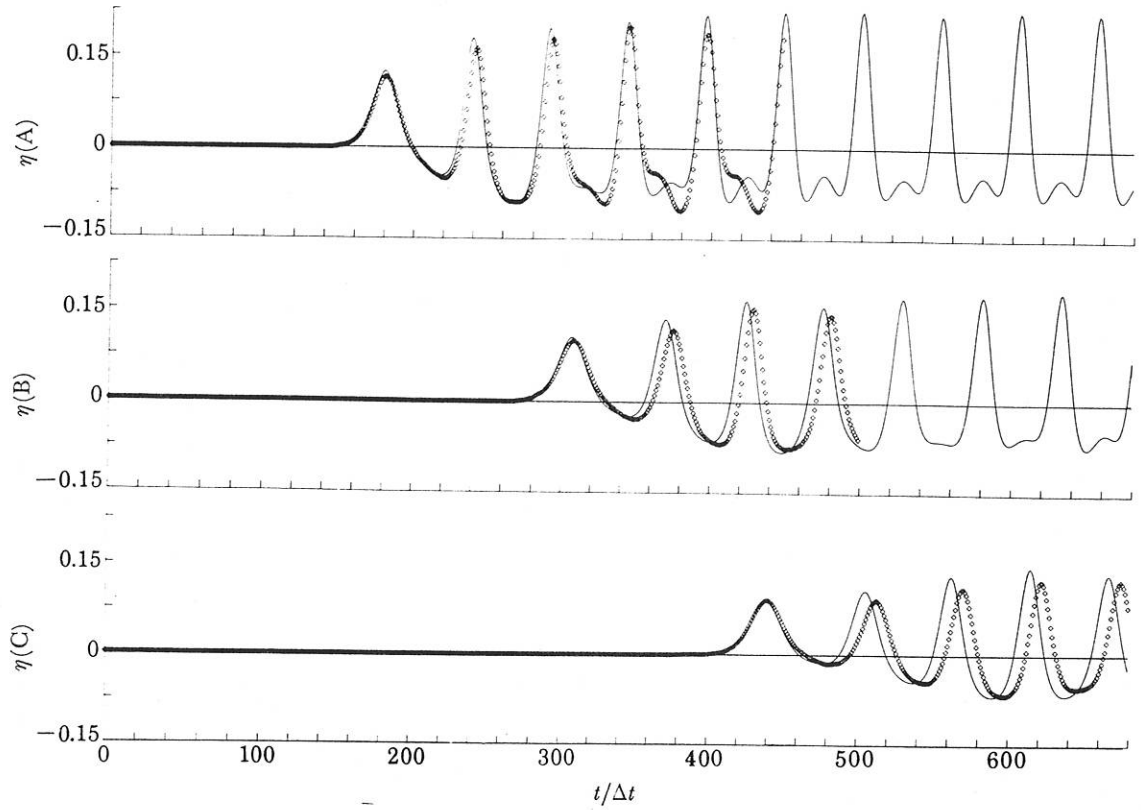

FIGURE 20. The experiment at $S = 35.9$ is compared with (M†) when $\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0.340 \times 10^{-2}$, $\mu = 0.168 \times 10^{-2}$, $\gamma = 0.1325$.

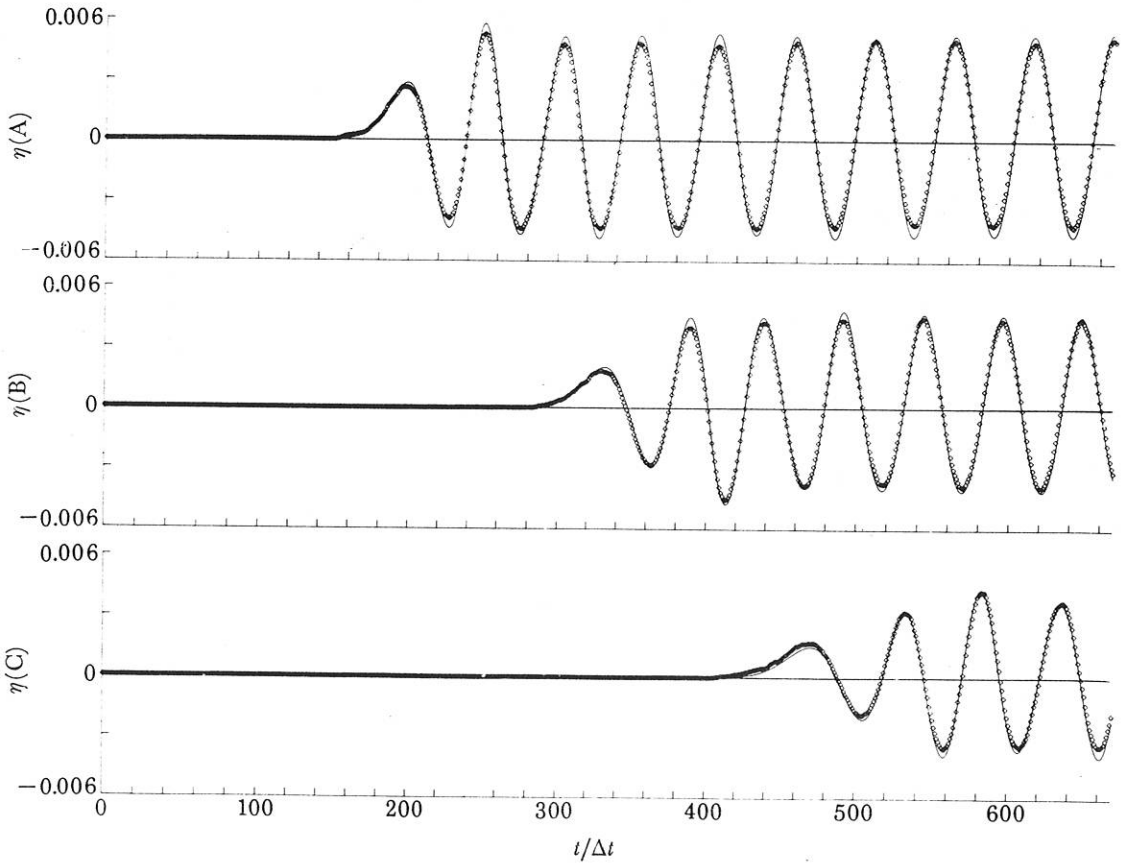

FIGURE 21. The experiment at $S = 0.95$ is compared with (M†) when $\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0.340 \times 10^{-2}$, $\mu = 0.168 \times 10^{-2}$, $\gamma = 0.1325$.

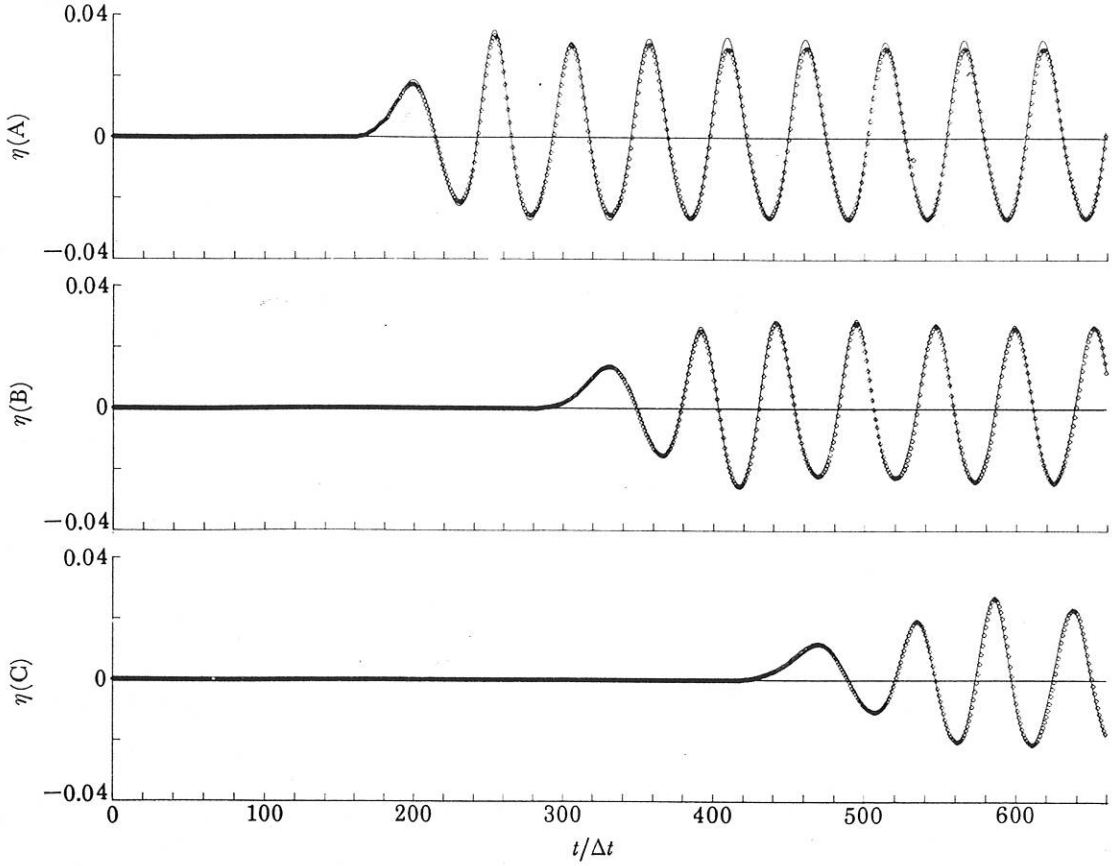

FIGURE 22. The experiment at $S = 5.5$ is compared with (M†) when $\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0.340 \times 10^{-2}$, $\mu = 0.168 \times 10^{-2}$, $\gamma = 0.1325$.

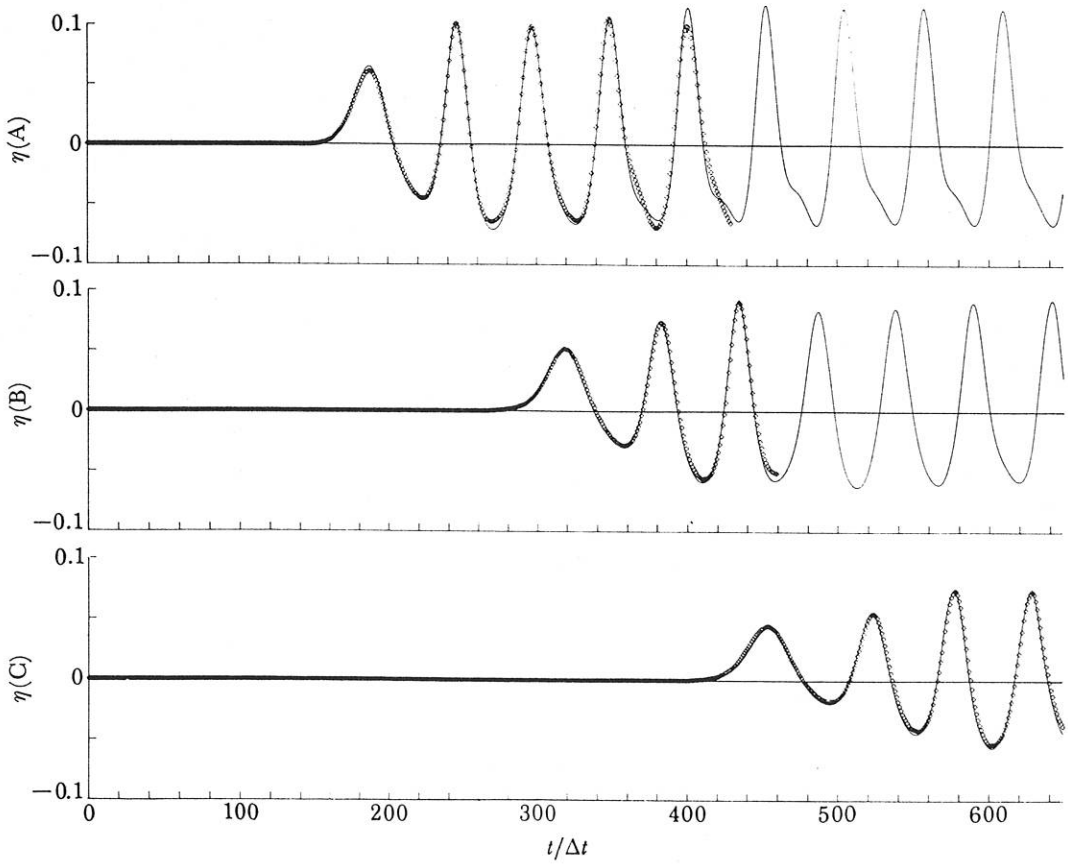

FIGURE 23. The experiment at $S = 18.1$ is compared with (M†) when $\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0.340 \times 10^{-2}$, $\mu = 0.168 \times 10^{-2}$, $\gamma = 0.1325$.

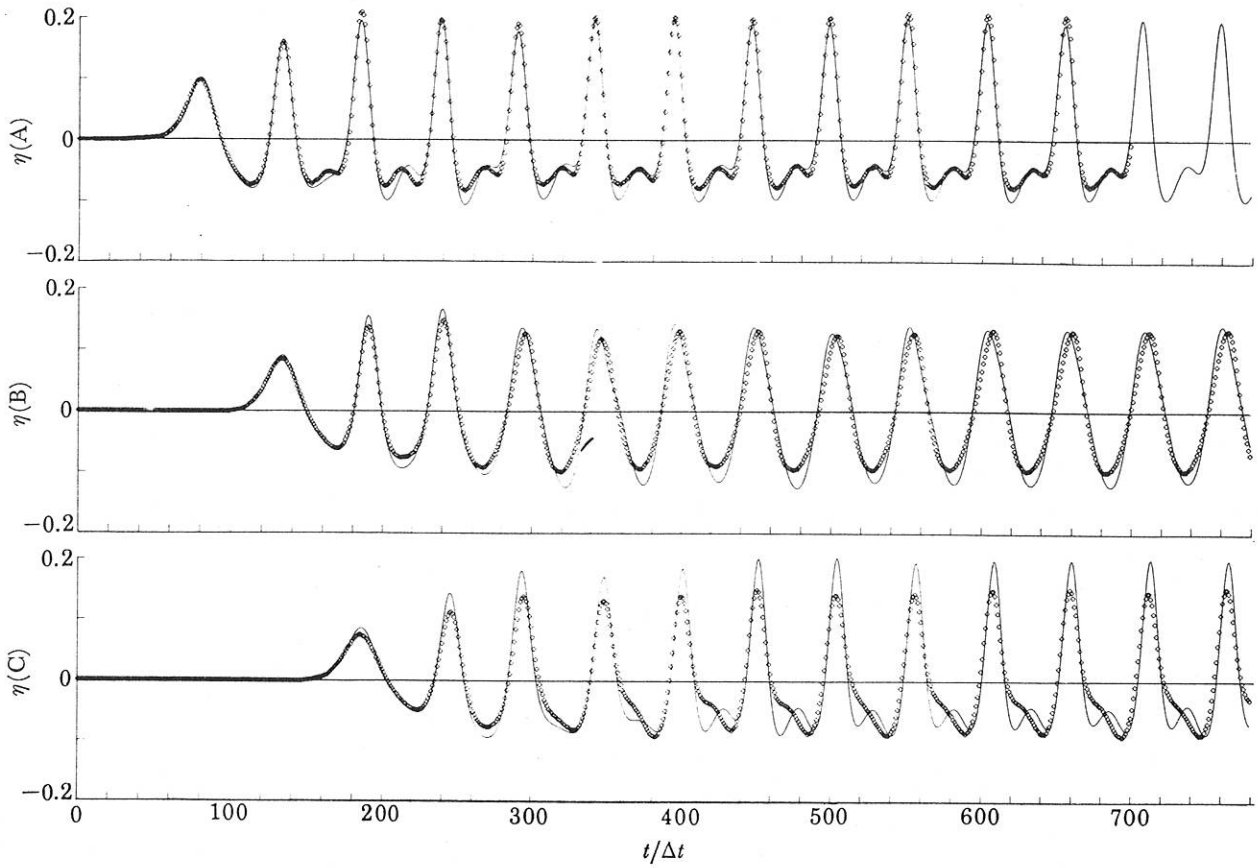

FIGURE 24. The experiment at $S = 26.3$ is compared with the inviscid version of (M†) ($\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0$, $\mu = 0$, $\gamma = 0.1325$).

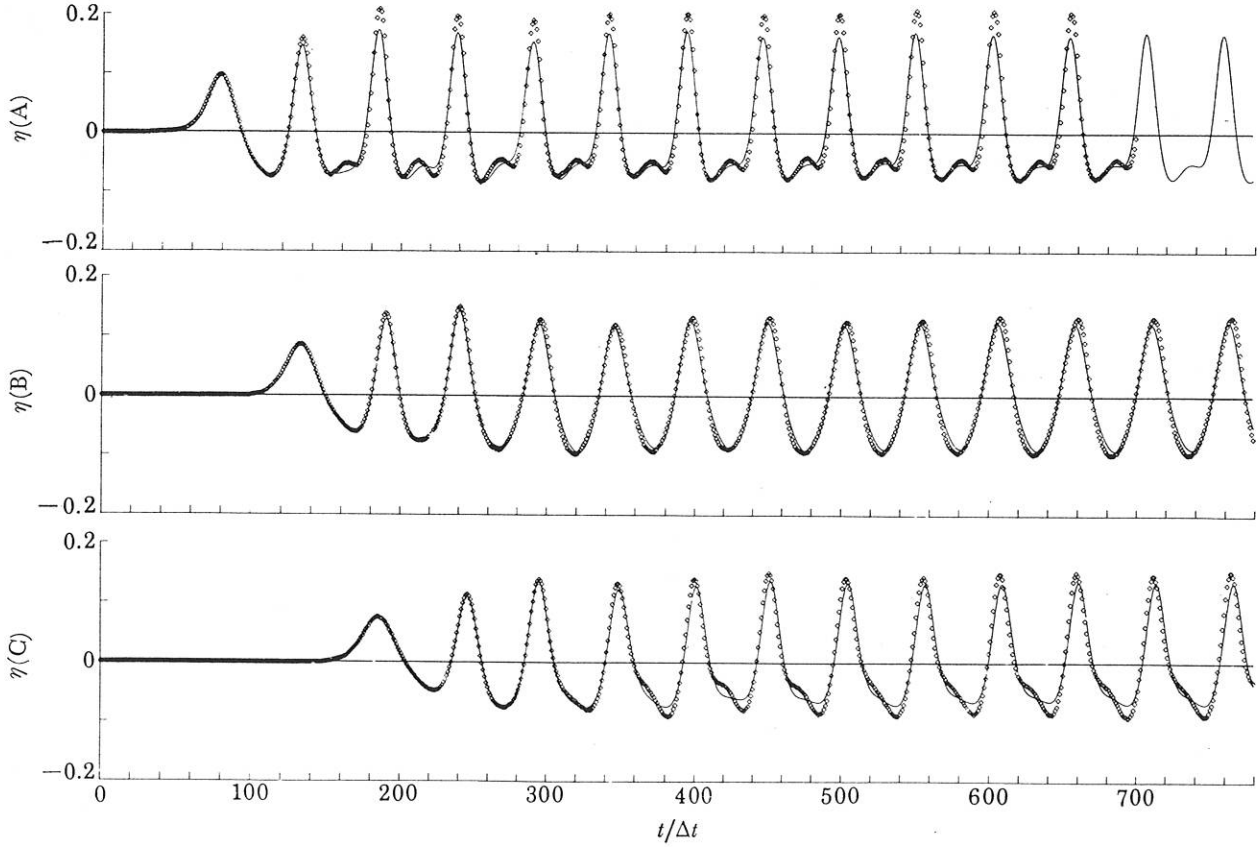

FIGURE 25. The experiment at $S = 26.3$ is compared with (M†) when $\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0$, $\mu = 0.014$, $\gamma = 0.1325$.

Thus, it would appear that some of the major discrepancies between the predictions of the model and the experimental results originated in the poor theoretical representation of the phase speeds of some of the larger wavenumbers arising in the experiments.

TABLE 6. THE COMPARISONS BETWEEN THE EXPERIMENTAL RESULTS AND (M†) WITH $\alpha = 0.9898$, $\beta = \frac{3}{2}$, $\nu = 0.340 \times 10^{-2}$, $\mu = 0.168 \times 10^{-2}$, $\gamma = 0.1325$

(The scheme is the same as for column I in table 4.)

station	S	0.38		0.95		4.5		5.5	
A		—	—	0.098	−0.03	0.117	0.24	0.082	0.12
				0.098		0.091		0.076	
B		0.225		0.092		0.143		0.063	
		0.064	0.46	0.090	−0.05	0.088	0.20	0.059	0.05
C		0.326		0.114		0.244		0.103	
		0.061	0.41	0.103	0.07	0.100	0.28	0.069	0.11

station	S	11.8		18.1		26.3		35.9	
A		0.133		0.075		0.149		0.313	
		0.077	0.35	0.075	−0.03	0.141	0.38	0.192	0.66
B		0.157		0.091		0.156		0.514	
		0.104	0.26	0.048	0.17	0.156	0.05	0.161	1.01
C		0.367		0.107		0.194		0.745	
		0.090	0.47	0.077	0.11	0.177	−0.21	0.221	1.07

8. RÉSUMÉ

The theoretical model predicted the experimental results, to as good an accuracy as could be expected, for the experiments made at values of S ranging up to 11.8. For these five experiments it was found that the inclusion of a dissipative term was much more important than the inclusion of the nonlinear term, although the inclusion of the nonlinear term was undoubtedly beneficial in describing the observations.

At larger values of S there were features of the experiments that were not predicted by the model. These features appear mainly to have been associated with harmonics (generated through nonlinear properties of the fundamental wavefield) having wavenumbers too large to be well represented by the small-wavenumber model. But by introducing a modification to the basic model that represented more accurately the phase speed of these harmonics, the description of the experimental results was significantly improved. This improvement suggests that a modification of the nonlinear effects, to allow a better representation of their influence at the larger wavenumbers, might also be helpful. Thus, in short, we feel that the original model would provide a good description of experiments in which the dominant wavenumber is much smaller than that used here, over a fairly wide range of values of the parameter S.

Appendix A. Deficiencies in an approximate procedure based on the pure initial-value problem

We wish to solve the pure initial-value problem

$$\eta_t+\eta_x+\tfrac{3}{2}\eta\eta_x-\tfrac{1}{6}\eta_{xxt}=0, \quad x\in\mathbb{R}, \tag{M \textit{bis}}$$

with the initial condition $\eta(x,0)=g(x)$. However, the initial datum g, to be determined empirically, is not easily obtained. Instead, a measurement of data $\eta(0,t)=\tilde{g}(t)$, $t\geqslant 0$, is made and, to recover the intended problem, the function $\tilde{g}(t)$ is transformed to an 'equivalent' spatial representation $\tilde{g}(x)$ by the leading-order approximation $\eta_t+\eta_x=0$ to (M). This transformation generates a small error, of order ϵ, in the representation $\tilde{g}(x)$ of the initial data $g(x)$, which would seem to be unimportant but, in the present example, is equivalent to the introduction of a forcing term on the right-hand side of (M) of size comparable with that of the nonlinear and the dispersive terms.

To illustrate the kinds of error that can arise in a practical case, let us consider the solitary-wave solution of (M), namely

$$\eta(x,t)=\eta_0\operatorname{sech}^2\{[3\eta_0/(4+2\eta_0)]^{\frac{1}{2}}[x+x_0-(1+\tfrac{1}{2}\eta_0)t]\}, \tag{A1}$$

where η_0 is the (maximum) wave amplitude and x_0 is a constant. Suppose that the measured data $\tilde{g}(t)$ are given by

$$\tilde{g}(t)=\eta_0\operatorname{sech}^2\{[3\eta_0/(4+2\eta_0)]^{\frac{1}{2}}[x_0-(1+\tfrac{1}{2}\eta_0)t]\},$$

then, by choosing x_0 large enough, the solution to the initial- and boundary-value problem for (M) with

$$\eta(x,0)=0, \quad \eta(0,t)=\tilde{g}(t)$$

(taking $\eta(0,0)/\eta_0=0.1\times 10^{-8}$), is a close approximation to (A1) for $x,t>0$ (cf. §3, table 1).

If $\tilde{g}(t)$ is now transformed to an 'equivalent' spatial form, it follows that

$$\tilde{g}(x)=\eta_0\operatorname{sech}^2\{[3\eta_0/(4+2\eta_0)]^{\frac{1}{2}}[x_0+(1+\tfrac{1}{2}\eta_0)x]\},$$

from which we see that $\tilde{g}(x)$ differs from the 'exact' form of g (the solution (A1) at time $t=0$) in both its shape and phase, the phase difference between $g(x)$ and $\tilde{g}(x)$ being

$$x_1=\tfrac{1}{2}\eta_0 x_0/(1+\tfrac{1}{2}\eta_0).$$

Notwithstanding the phase error, let us, for the time being, investigate the importance of the 'shape' error in $\tilde{g}$ by using $\tilde{g}(x-x_1)$ as the initial data for (M).

Thus, using a scheme similar to that described in §3 (which can also be analysed similarly), we have solved numerically the pure initial-value problem (M) with $\eta(x,0)=\tilde{g}(x-x_1)$, which solution we denote by $\tilde{\eta}(x,t)$, and have compared $\tilde{\eta}$ with the 'exact' solution (A1).

The kinds of error that can arise in practice are shown in figure A1. Here the wave amplitude $\eta_0=0.25$ was chosen to correspond approximately to the largest amplitudes used in the laboratory experiments and the integration was carried on to about the same time as that occurring in the experiments. In figure A1 the dotted line represents the function $\tilde{\eta}$, and the full line represents the solitary-wave solution. Let

$$\mathscr{E}=\sum_{j=0}^{N}|\eta(j\Delta x,t)-\tilde{\eta}(j\Delta x,t)|\Delta x\Big/\sum_{j=0}^{N}|\eta(j\Delta x,t)|\Delta x$$

measure the difference between $\tilde{\eta}$ and the 'exact' solution. The initial difference between $\tilde{g}$ and g was approximately 0.111. At time $t = 32.0$ the approximate solution $\tilde{\eta}$ had developed a distinct oscillatory tail, and the difference $\mathscr{E}$ had increased to 0.254. This difference then continued to increase with time, roughly linearly, taking values of 0.561 at $t = 96.0$ and 0.898 at $t = 192.0$. (The error $\mathscr{E}$ in integrating numerically a solitary wave of amplitude 0.25 under the conditions of this experiment was less than 0.55×10^{-3} at $t = 192.0$.) The figure shows how the tail developed by $\tilde{\eta}$ gradually separated from the leading wave. The speed of the leading crest of

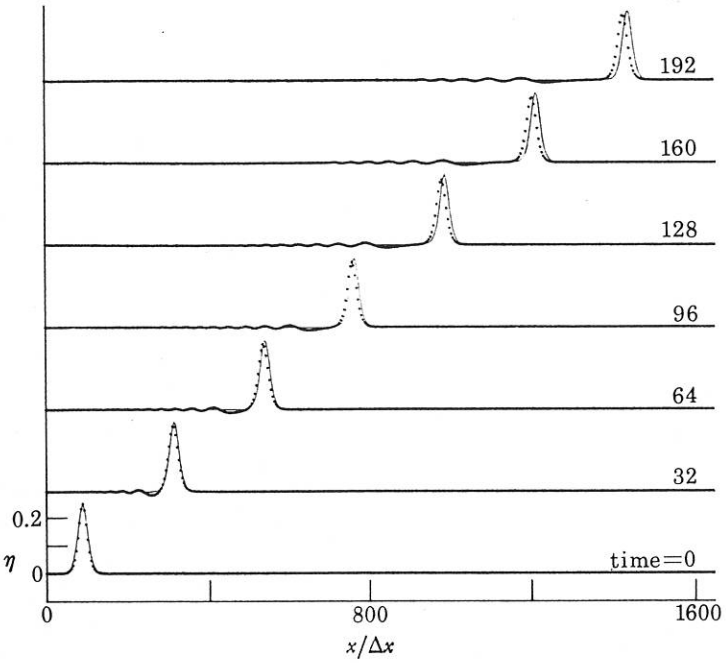

FIGURE A1. The solitary-wave solution of (M) (full line) is compared with the solution to (M) with the initial data $\eta(x, 0) = \tilde{g}(x-x_1)$ (dotted lines). The amplitude $\eta_0 = 0.25$; the computations were made with $\Delta t = \Delta x = 0.16$.

the oscillatory tail was approximately 0.9719 at $t = 192.0$. On the other hand, the leading wave of $\tilde{\eta}$ appeared to be evolving towards a solitary wave of the form (A1) with an amplitude of approximately 0.2139, as determined from a fourth-order interpolation of the discretized solution. For example, the speed of this wave differed from a solitary-wave solution (A1) of the same amplitude by less than 0.28×10^{-6} at $t = 192.0$, and the difference $\mathscr{E}$ between the two waveforms was less than 0.27×10^{-3}. (For this latter comparison the crest of the solitary-wave profile was chosen to coincide with that of the leading wave of $\tilde{\eta}$, and the domain for the comparison was terminated at a distance a from the crest, where a was chosen so that the solitary waveform had decayed to 0.1×10^{-4} of its maximum amplitude.)

Similar results were obtained with $\eta_0 = 0.1$, except that the initial error $\mathscr{E}$ at $t = 0$ was 0.048 and this degraded to 0.095 at $t = 96.0$ and 0.155 at $t = 192.0$. At $t = 192.0$ the amplitude of the leading wave of $\tilde{\eta}$ was 0.0971, but the wave was still undergoing significant modifications.

It should be noted that the above comparisons underestimate considerably the actual errors arising with this method because we have removed the initial phase error induced by the approximate transformation of the data. (For example, with $\eta_0 = 0.1$ and $x_0 = 15.44$ the error $\mathscr{E}$ between $g(x)$ and $\tilde{g}(x)$ was 0.197; cf. the difference of only 0.048 between $g(x)$ and $\tilde{g}(x-x_1)$.) With more general data it would not be easy to eliminate this initial phase error.

J. L. B. is grateful for support from the Science Research Council (U.K.) and for the hospitality of the Fluid Mechanics Research Institute, University of Essex. W. G. P. and L. R. S. are grateful for support received during the course of this study from I.C.A.S.E. (at the N.A.S.A. Langley Research Center, Hampton, Virginia) under contract no. NAS1–14101. Part of the work was also done at the Brookhaven National Laboratory under U.S. Dept. of Energy contract EY–76–C–02–0016, and at the Mathematics Research Center, University of Wisconsin, Madison, under U.S. Army contract DAAG29–80–C–0041. In addition J. L. B. was supported by the National Science Foundation (U.S.A.) and L. R. S. by U.S. Air Force contract F49620–79–C–0149.

References

Barnard, B. J. S., Mahony, J. J. & Pritchard, W. G. 1977 The excitation of surface waves near a cut-off frequency. *Phil. Trans. R. Soc. Lond.* A **286**, 87.

Benjamin, T. B., Bona, J. L. & Mahony, J. J. 1972 Model equations for long waves in nonlinear dispersive systems. *Phil. Trans. R. Soc. Lond.* A **272**, 47.

Bona, J. L. & Bryant, P. J. 1973 A mathematical model for long waves generated by wavemakers in nonlinear dispersive systems. *Proc. Camb. phil. Soc.* **73**, 391.

Bona, J. L. & Smith, R. 1975 The initial-value problem for the Korteweg–de Vries equation. *Phil. Trans. R. Soc. Lond.* A **278**, 555.

Bona, J. L. & Winther, R. 1981 The Korteweg–de Vries equation, posed in a quarter plane. University of Wisconsin, Mathematics Research Center, Report no. 2258.

Davis, P. J. & Rabinowitz, P. 1967 *Numerical Integration.* Waltham, Massachusetts: Blaisdell.

Hammack, J. L. 1973 A note on tsunamis: their generation and propagation in an ocean of uniform depth. *J. Fluid Mech.* **60**, 769.

Hammack, J. L. & Segur, H. 1974 The Korteweg–de Vries equation and water waves. Part 2. Comparison with experiments. *J. Fluid Mech.* **65**, 289.

Havelock, T. H. 1929 Forced surface waves on water. *Phil. Mag.* F **8**, 569.

Isaacson, E. & Keller, H. B. 1966 *Analysis of numerical methods.* New York: John Wiley.

Kakutani, T. & Matsuuchi, K. 1975 Effect of viscosity on long gravity waves. *J. phys. Soc. Japan* **39**, 237.

Keulegan, G. H. 1948 Gradual damping of solitary waves. *J. Res. natn. Bur. Stand.* **40**, 487.

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

Madsen, O. S. 1974 A three dimensional wave maker, its theory and application. *J. hydraul. Res.* **12**, 205.

Mahony, J. J. & Pritchard, W. G. 1980 Wave reflexion from beaches. *J. Fluid Mech.* **101**, 809.

Mei, C. C. & Liu, L. F. 1973 The damping of surface gravity waves in a bounded liquid. *J. Fluid Mech.* **59**, 239.

Meyer, R. E. 1972 Note on the longwave equations. University of Essex, Fluid Mech. Res. Inst. Rep. no. 23.

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

Peregrine, D. H. 1966 Calculations of the development of an undular bore. *J. Fluid Mech.* **25**, 321.

Whitham, G. B. 1974 *Linear and nonlinear waves.* New York: John Wiley.

Zabusky, N. J. & Galvin, C. J. 1971 Shallow-water waves, the Korteweg–de Vries equation and solitons. *J. Fluid Mech.* **47**, 811.