JOURNAL OF COMPUTATIONAL PHYSICS 75, 31–53 (1988)

Grid Resonances, Focusing and Benjamin–Feir Instabilities in Leapfrog Time Discretizations

A. CLOOT AND B. M. HERBST

Department of Applied Mathematics, University of the Orange Free State, Bloemfontein 9300, South Africa

Received September 3, 1986; revised April 3, 1987

W. L. Briggs, A. C. Newell, and T. Sarie (*J. Comput. Phys.* **50**, 83 (1983)) studied the solutions of a class of discretizations of $u_t + au_x + uu_x = 0$, where "a" is a constant. They discovered that the numerical solution of the discretization becomes unbounded despite the fact that theoretically the solution should remain bounded. They attributed this anomalous behavior to a focusing mechanism. In this paper we make use of a multiple scales analysis and show that the instabilities are caused by a resonance effect introduced by the discretizations. The contributions from the space and time discretizations are analysed separately and in detail. Thus, the structure of the focusing mechanism becomes transparent.

1. INTRODUCTION

The stability of nonlinear partial difference equations must be ranked as one of the more important questions in numerical analysis. In this paper we are concerned with difference schemes arising from discretizations of

$$u_t + uu_x = 0. \tag{1}$$

Despite significant progress on partial difference equations arising from (1) by several people, including Richtmyer and Morton [7], Fornberg [2], Newell [6], Briggs *et al.* [1], Sanz-Serna [8, 11], and Sloan and Mitchell [10], it is safe to say that the stability of these difference schemes is not yet completely understood.

The earlier investigators already recognized that solutions of (1) which oscillate around zero are particularly susceptible to instabilities. Let us now consider solutions of (1) which are perturbations around the constant solution $u = a$; i.e., let us consider

$$u_t + au_x + uu_x = 0, \tag{2}$$

where u is regarded as the perturbation and therefore considered to be small. In addition we only consider 2π-periodic solutions of (2), i.e.,

$$u(x + 2\pi, t) = u(x, t).$$

0021-9991/88 $3.00

581/75/1-3

Introducing a uniform grid with grid length

$$h := 2\pi/N,$$

we follow Briggs *et al.* [1] and semi-discretize (2) by

$$\dot{U}_j + \frac{a}{2h}(U_{j+1} - U_{j-1}) + \frac{\theta}{4h}(U_{j+1}^2 - U_{j-1}^2) + \frac{1-\theta}{2h} U_j(U_{j+1} - U_{j-1}) = 0, \quad (3)$$

where

$$0 \leqslant \theta \leqslant 1.$$

In order to ensure periodic solutions we impose the boundary conditions

$$U_0 = U_N, \quad U_1 = U_{N+1}.$$

For large values of U_j in (3) the nonlinear terms will dominate which may result in an explosive growth of the solution. If, on the other hand, U_j is small, then solutions of (3) are expected to behave rather like solutions of the linear poblem

$$\dot{U}_j + \frac{a}{2h}(U_{j+1} - U_{j-1}) = 0 \quad (4)$$

which allows solutions of the form

$$U_j = e^{i(kx_j - \omega t)}$$

provided ω satisfies the dispersion relation

$$\omega_k = \frac{a}{h}\sin(kh). \quad (5)$$

Since we assume 2π-periodic solutions, k is an integer.

Briggs *et al.* [1] considered solutions of (3) of the form

$$U_j = A(t)\, e^{ikx_j} + \text{c.c.} \quad (6)$$

where c.c. indicates the complex conjugate of the preceeding terms. For $k = \frac{1}{3}N$, $A(t)$ and its complex conjugate $A^*(t)$ are given by

$$\dot{A}(t) + i\omega_k A(t) = \frac{i}{2a}\omega_k(2 - 3\theta)\, A^*(t)^2. \quad (7)$$

Briggs *et al.* solved (7) by the leapfrog or explicit midpoint rule for $\theta \neq \frac{2}{3}$ and obtained a threshold on the magnitude of the solution, if the solution is initially below the threshold the solution remains bounded. However, if for some reason the

solution exceeds this value the nonlinear processes dominate and the solution rapidly becomes unbounded. Consult Sloan and Mitchell [10] for the connection with the earlier work of Fornberg [2].

The situation becomes more complicated when we attempt to solve (3) with the leapfrog scheme. Rather surprisingly the numerical experiments of Briggs *et al.* showed that initial conditions which lead to bounded solutions in the case of (7) may lead to unbounded solutions in the case of the partial difference scheme obtained from (3). Their numerical experiments indicate that the Fourier modes in the vicinity of the original or fundamental mode, k, first start to grow and from these modes energy is fed into more modes until most or all modes are excited. In physical space this causes the solution to develop sharp peaks at isolated positions. As soon as a peak exceeds the threshold value the solution quickly becomes unbounded. In their investigations into the origins of the instability Sloan and Mitchell [10] did a Benjamin–Feir side-band analysis which led to necessary conditions for the onset of the instability.

In this paper we are also concerned with the processes up to the stage where the full nonlinear effects take over and the solution rapidly becomes unbounded. Although the results of our investigations agree entirely with that of Sloan and Mitchell, within the frameworks of the various limitations imposed by the techniques employed, we believe that our investigations shed new light on the structure of the mechanism responsible for the instability. The technique used in this paper was inspired and resembles that used by Moore [5].

This paper is organized in two parts. In the first part we investigate the semi-discrete system (3). The basic ideas needed for the second part are developed and as a result we show that the semi-discrete system (3) does not become unstable. In the second part we consider the leapfrog discretization of (3) and using the ideas discussed in the first part we explain in some detail why the leapfrog time discretization can become unstable.

I. Semi-Discretization

2. *Wave-Train Solutions*

In the previous section it was mentioned that (3) may admit solutions of the form

$$U_j = \varepsilon(e^{i(kx_j - \omega_k t)} + \text{c.c.}) + O(\varepsilon^2), \tag{8}$$

where ω_k satisfies the dispersion relation and $\varepsilon \ll 1$. More precisely, assume a steady wave solution of (3) of the form

$$U_j = \frac{1}{2} \sum_{l=1}^{\infty} a_l [e^{il(kx_j - ct)} + \text{c.c.}], \tag{9}$$

where c and the a_l are unknown.

We emphasize that it is reasonable to expect solutions of this form only for small U_j. If we substitute (9) into (3) and gather coefficients of $e^{il(kx_j - ct)}$ we obtain the following system of equations:

for $l=1$,

$$-ca_1+\frac{a}{h}\sin(kh)\,a_1+\frac{\theta}{2h}\sin(kh)\sum_{s=1}^{\infty}a_s a_{s+1}$$
$$+\frac{1-\theta}{2h}\sum_{s=1}^{\infty}(\sin((s+1)\,kh)-\sin(skh))\,a_s a_{s+1}=0; \tag{10a}$$

for $l>1$,

$$-cla_l+\frac{a}{h}\sin(lkh)\,a_l+\frac{\theta}{4h}\sin(lkh)\left[\sum_{s=1}^{l-1}a_s a_{l-s}+2\sum_{s=1}^{\infty}a_s a_{l+s}\right]$$
$$+\frac{1-\theta}{2h}\left[\sum_{s=1}^{l-1}\sin(skh)\,a_s a_{l-s}+\sum_{s=1}^{\infty}(\sin((l+s)\,kh)-\sin(skh))\,a_s a_{l+s}\right]=0. \tag{10b}$$

These equations may be solved by expanding a_l and c in Stokes fashion (Whitham [12], Moore [5]),

$$a_1=\varepsilon$$
$$a_l=\sum_{s=l}^{\infty}\varepsilon^s A_{ls}, \qquad l>1 \tag{11}$$
$$c=\sum_{s=0}^{\infty}\varepsilon^s c_s.$$

This may now be used to calculate the solution (9). In particular this yields

$$c=\omega_k+O(\varepsilon^2) \tag{12}$$

with ω_k given by (5). However, we have no proof that (10) can actually be solved. Henceforth we assume that (3) possesses a solution of the form (8) for $\varepsilon \ll 1$. For any particular value of k this assumption should be verified numerically and an upper bound on ε established.

3. *Grid Resonances*

Consider a solution $\tilde{U}_j$ of (3) of the form (8) and let ϕ_j be a perturbation of $\tilde{U}_j$, i.e.,

$$U_j=\tilde{U}_j+\phi_j, \tag{13}$$

where

$$\phi_j = \xi \sum_{n=0}^{(1,2)N} (\alpha_n(t)\, e^{i nx_j} + \alpha_n^*(t)\, e^{-i nx_j}), \qquad \xi \ll \varepsilon. \tag{14}$$

If (13) is substituted into (3) and linear terms in ϕ_j retained, it follows that

$$\dot{\phi}_j + \frac{a}{2h}(\phi_{j+1} - \phi_{j-1}) + \frac{\theta}{2h}(\tilde{U}_{j+1}\phi_{j+1} - \tilde{U}_{j-1}\phi_{j-1})$$
$$+ \frac{1-\theta}{2h}((\tilde{U}_{j+1} - \tilde{U}_{j-1})\phi_j + \tilde{U}_j(\phi_{j+1} - \phi_{j-1})) = 0. \tag{15}$$

If we substitute (8) for $\tilde{U}_j$ in (15) and (14) for ϕ_j and collect coefficients of $e^{i sx_j}$, we obtain

$$\dot{\alpha}_s + i\omega_s\alpha_s + i\varepsilon C_s e^{-i\omega_k t}(\alpha_{s-k} + \alpha_{k-s}^*)$$
$$+ i\varepsilon D_s e^{i\omega_k t}(\alpha_{s+k} + \alpha_{N-(k+s)}^*) = 0 \tag{16}$$

where

$$C_s = \frac{1}{h}(\theta \sin(sh) + (1-\theta)(\sin(kh) + \sin((s-k)h))) \tag{17a}$$

and

$$D_s = \frac{1}{h}(\theta \sin(sh) - (1-\theta)(\sin(kh) - \sin((s+k)h))). \tag{17b}$$

We observe that only those terms α_n, α_n^* appear in (16) whose index n satisfies

$$0 \leqslant n \leqslant \tfrac{1}{2}N.$$

For small values of ε we may follow the reasoning of Moore [5] and consider the $O(\varepsilon)$ terms in (16) to be source terms for the equation.

$$\dot{\alpha}_s + i\omega_s\alpha_s = 0. \tag{18}$$

Thus, for those values of s in (16) for which either

$$\omega_s = \omega_k + \omega_{s-k} \tag{19a}$$

or

$$\omega_s = \omega_{s+k} - \omega_k \tag{19b}$$

the $O(\varepsilon)$ terms in (16) will be solutions of (18), i.e., α_s grows linearly in time. It is

obvious that (19a) will be satisfied for $s=k$. For values of s in the vicinity of k (19a) will be approximately satisfied, i.e.,

$$\omega_s - \omega_k - \omega_{s-k}$$

may be small and it is possible that this resonance will be strong enough to cause linear growth in time. In the next section this situation will be analyzed, by using the method of multiple scales in a similar fashion as Moore [5] did.

4. *A Multiple Scales Analysis*

Assume that for a certain value of s,

$$\omega_s - \omega_k - \omega_{s-k} = \varepsilon\delta, \tag{20}$$

where $\delta = O(1)$. We are particularly interested in values of s close to k, i.e.,

$$s = k + r \qquad \text{and} \qquad s = k - r,$$

where r is a small positive integer. If δ in (20) is defined by

$$\omega_{k+r} = \omega_k + \omega_r + \varepsilon\delta \tag{21a}$$

then, if $rh \ll 1$, it follows from (5) and (20) that

$$\omega_{k-r} = \omega_k - \omega_r - \varepsilon\delta. \tag{21b}$$

The following relations follow from (21) and are needed in the subsequent analysis,

$$\omega_k = \tfrac{1}{2}(\omega_{k+r} + \omega_{k-r}) \tag{22a}$$

$$\omega_r = \tfrac{1}{2}(\omega_{k+r} - \omega_{k-r}) - \varepsilon\delta \tag{22b}$$

$$\omega_{k+r} = \omega_k + \tfrac{1}{2}(\omega_{k+r} - \omega_{k-r}) \tag{22c}$$

$$\omega_{k-r} = \omega_k - \tfrac{1}{2}(\omega_{k+r} - \omega_{k-r}). \tag{22d}$$

According to the method of multiple scales, (see, e.g., Jeffrey and Kawahara [3])

$$\alpha_s = \sum_{j=1}^{\infty} \varepsilon^j \alpha_s^j(T_0, T_1, \ldots), \tag{23}$$

where the slow time variables T_k are given by

$$T_k = \varepsilon^k t, \qquad k = 0, 1, 2, \ldots . \tag{24}$$

If (23) and (24) are substituted into (16) and use is made of (22), we obtain from $O(\varepsilon)$ terms:

$$\alpha^1_{k+r} = A_{k+r}(T_1, T_2, \ldots)\, e^{-i(\omega_k+\omega_r)T_0} \tag{25a}$$

$$\alpha^1_{k-r} = A_{k-r}(T_1, T_2, \ldots)\, e^{-i(\omega_k-\omega_r)T_0} \tag{25b}$$

$$\alpha^1_r = A_r(T_1, T_2, \ldots)\, e^{-i\omega_r T_0}. \tag{25c}$$

The $O(\varepsilon^2)$ contribution is

$$\frac{d\alpha^2_{k+r}}{dT_0} + i(\omega_k+\omega_r)\,\alpha^2_{k+r} + \frac{d\alpha^1_{k+r}}{dT_1} + i\delta\alpha^1_{k+r} + iC_{k+r}e^{-i\omega_k T_0}\alpha^1_r + iD_{k+r}e^{i\omega_k T_0}\alpha^{1*}_{N-(2k+r)} = 0, \tag{26a}$$

$$\frac{d\alpha^2_{k-r}}{dT_0} + i(\omega_k-\omega_r)\,\alpha^2_{k-r} + \frac{d\alpha^1_{k-r}}{dT_0} - i\delta\alpha^1_{k-r} + iC_{k-r}e^{-i\omega_k T_0}\alpha^{1*}_r + iD_{k-r}e^{i\omega_k T_0}\alpha^{1*}_{N-(2k-r)} = 0 \tag{26b}$$

$$\frac{d\alpha^2_r}{dT_0} + i\tfrac{1}{2}(\omega_{k+r}-\omega_{k-r})\,\alpha^2_r + \frac{d\alpha^1_r}{dT_1} - i\delta\alpha^1_r + iC_r e^{-i\omega_k T_0}\alpha^{1*}_{k-r} + iD_r e^{i\omega_k T_0}\alpha^1_{k+r} = 0. \tag{26c}$$

Recall that a "*" denotes the complex conjugate and we assumed that k is such that

$$0 \leqslant N-(2k\pm r) \leqslant \tfrac{1}{2}N.$$

There is no loss of generality by this assumption.

If (25) is substituted into (26) the secular terms are removed by

$$\frac{dA_{k+r}}{dT_1} + i\delta A_{k+r} + iC_{k+r}A_r = 0 \tag{27a}$$

$$\frac{dA_{k-r}}{dT_1} - i\delta A_{k-r} + iC_{k-r}A^*_r = 0 \tag{27b}$$

$$\frac{dA_r}{dT_1} - i\delta A_r + iC_r A^*_{k-r} + iD_r A_{k+r} = 0. \tag{27c}$$

It follows readily from (27) that A_{k+r} and A_{k-r} are of the form

$$A_{k\pm r} \sim e^{\lambda T_1},$$

where λ is given by

$$\lambda^2 = \begin{cases} -(\delta^2+A) \\ -\delta^2, \end{cases} \tag{28}$$

where

$$\Delta = C_{k+r}D_r - C_{k-r}C_r. \tag{29}$$

From (28) it is clear that a necessary condition for the side bands to grow exponentially on the T_1 time scale, i.e., linearly on the t time scale, is

$$\delta^2 + \Delta < 0. \tag{30}$$

This condition can only be satisfied if $\Delta < 0$ and together with (30) this provides an upper bound on the value of δ. From (21) follows that an upper bound on δ implies a lower bound on ε. Thus, in order to have instability on the T_1 time scales the resonance needs to be strong enough and this is only possible if ε, i.e., the amplitude of the fundamental wave-train (8) is big enough. On the other hand, in Section 2 we argued that ε needs to be small in order to have a wave-train solution. It is therefore possible, that one condition may exclude the other, and as we show in subsequent sections this indeed appears to be the case.

5. *The Stability Condition*

In order to analyze the meaning of the stability conditions (28) and (30), assume

$$rh \ll 1$$

in which case

$$\Delta = r^2[(2\theta + 1)(1 - \theta)\cos^2(kh) + 2(2\theta^2 - 2\theta + 1)\cos(kh) + (2\theta - 1)(1 - \theta)]. \tag{31}$$

From (31) follows that (30) cannot be satisfied for small values of k. In fact all values of k such that

$$2\pi k \leqslant N \arccos(\sqrt{2} - 1) \tag{32}$$

have $\Delta \geqslant 0$ for all $0 \leqslant \theta \leqslant 1$. For $kh \ll 1$ in (5), the numerical dispersion relation is a good approximation of the corresponding linear one. In this case, (32) guarantees a stable solution. The optimal value of θ, which according to (31) allows the least number of potentially unstable modes is $\theta \approx 0.7$. This value is close to the value $\theta = \frac{2}{3}$ used by Briggs *et al.* [1] for different reasons.

6. *Examples*

It is not easy to see if the two opposing conditions on ε, namely ε smaller than a threshold value to ensure a steady wave-train solution and ε bigger than a certain value to provide strong enough resonances for instability, have any region of overlap. In the present section we investigate specific examples.

6.1. *One Mode Solution*

We return to the example of Section 1. Ignoring for the moment the obvious choice $\theta = \frac{2}{3}$ we follow Sloan and Mitchell [10] and set in (7)

$$A(t) = X(t) + i\,Y(t)$$

to obtain

$$\dot{X} = \omega Y + \frac{\omega}{a}(2-3\theta)\,XY \tag{33a}$$

$$\dot{Y} = -\omega X + \frac{\omega}{2a}(2-3\theta)(X^2 - Y^2), \tag{33b}$$

where we used the simplified notation ω instead of $\omega_{(1\cdot 3),N}$.

Following standard practice, let

$$X(t) = \varepsilon(t)\cos(\eta(t)) \tag{34a}$$

$$Y(t) = \varepsilon(t)\sin(\eta(t)), \tag{34b}$$

i.e., we transform in phase-space to polar coordinates. It is now readily shown that

$$\dot{\varepsilon}(t) = \frac{\omega}{2a}(2-3\theta)\,\varepsilon(t)^2[3\sin(\eta(t)) - 4\sin^3(\eta(t))] \tag{35}$$

and

$$\dot{\eta}(t) = -\omega + \frac{\omega}{2a}(2-3\theta)\,\varepsilon(t)[4\cos^3(\eta(t)) - 3\cos(\eta(t))]. \tag{36}$$

Making use of an averaging procedure (Jordan and Smith [4]), it follows that

$$\varepsilon(t) = \varepsilon_0 + O(\varepsilon^3) \tag{37a}$$

and

$$\eta(t) = -\omega t + O(\varepsilon^2), \tag{37b}$$

where ε_0 is a small constant. Thus a solution of the form (8) with $k = \frac{1}{3}N$ has been established.

We note that Sloan and Mitchell [10] proved

$$\varepsilon_0 < \frac{a}{|2-3\theta|} \tag{38}$$

to be necessary and sufficient to prevent nonlinear instabilities in this particular

case. Clearly $\theta = \frac{2}{3}$ places no restriction on ε_0. This is, of course, the choice for which (33) becomes linear. In order to test the side-band instability conditions, choose

$$N = 120, \qquad k = 40, \qquad \theta = 0, \qquad \text{and} \qquad a = 0.5.$$

For $r = 1$ we obtain $\Delta = -1.749$ which implies $\varepsilon > 0.58$. This value of ε is prevented by the nonlinear stability condition (38). Thus we conclude that the resonance allowed by the nonlinear stability condition is not strong enough to cause side-band instabilities.

For choices of k other than $\frac{1}{3}N$ we do not have an equivalent for (38) and it is therefore difficult to tell if strong enough resonances are allowed within the non-linear stability limit. In Table I we choose

$$N = 120, \qquad \theta = 0, \qquad a = 0.5, \qquad \text{and} \qquad r = 1$$

and calculate the minimum value of ε required for resonance for various values of k.

From Table I it follows that fairly large values of ε are required to provide the necessary resonance for instability. The numerical evidence we have for the threshold required to prevent nonlinear instability appears to require values of ε less than those given in Table I.

The same applies to other choices of θ. For instance, $\theta = \frac{2}{3}$ and $k = \frac{1}{3}N$ require no restriction on ε to prevent nonlinear instability. However, in this case a value of $\varepsilon > 1.5$ is required for side-band instabilities. Clearly this value falls outside the range for which a multiple scales analysis applies, which forms the basis of the estimate.

6.2. *Two-Mode Solution*

Aliasing may also be used to obtain a solution of the form

$$U_j(t) = A(t)\, e^{i(1/4)Nx_j} + A^*(t)\, e^{-i(1/4)Nx_j} + B(t)\, e^{i(1/2)Nx_j}, \tag{39}$$

where the complex function $A(t)$ and the real function $B(t)$ satisfy

$$\dot{A}(t) + i\frac{a}{h} A(t) = \frac{i}{h}(1 - 2\theta)\, A^*(t)\, B(t) \tag{40a}$$

$$\dot{B}(t) = \frac{i}{h}(\theta - 1)(A^2(t) - A^{*2}(t)). \tag{40b}$$

TABLE I

Resonance Requirements for Various Values of k

k	25	30	35	40	45	50	55	60
$\varepsilon_{\min}$	0.59	0.51	0.53	0.58	0.62	0.67	0.70	0.71

We note that (39) is not of the form (8), because of the presence of the $\frac{1}{2}N$ mode. Also no theoretical results providing a threshold on the magnitude of the solution (39) are available. However, a naive perturbation analysis gives

$$A(t) = \varepsilon(e^{i((1/4)Nx_j - \omega_{(1/4)N}t)} + \text{c.c.}) + O(\varepsilon^2), \tag{41a}$$

$$B(t) = \varepsilon(e^{i((1/2)Nx_j - \omega_{(1/2)N}t)} + \text{c.c.}) + O(\varepsilon^2). \tag{41b}$$

As pointed out by Sloan and Mitchell [10], (40) becomes linear if $\theta = 1$ which implies that the $O(\varepsilon^2)$ terms in (41) disappear. Assuming a steady wave-train of the form

$$U_j(t) = \varepsilon(e^{i(k_1 x - \omega_{k_1}t)} + \text{c.c.}) + \varepsilon e^{i(k_2 x - \omega_{k_2}t)} + O(\varepsilon^2),$$
$$k_1 = \tfrac{1}{4}N, \qquad k_2 = \tfrac{1}{2}N \tag{42}$$

for small values of ε, the arguments leading to (28) may be repeated. Hence, we find that the side bands in the vicinity of k_1 or k_2 will grow under the same conditions as before. The general remarks made in connection with the previous example therefore also apply to the present one.

II. Discretizing the Time Variable

7. *The Leapfrog Scheme*

We now proceed to discretize (3) by the leapfrog scheme,

$$U_j^{n+1} - U_j^{n-1} + \alpha(U_{j+1}^n - U_{j-1}^n) + \tfrac{1}{2}\gamma\theta((U_{j+1}^n)^2 - (U_{j-1}^n)^2)$$
$$+ \gamma(1-\theta)\, U_j^n(U_{j+1}^n - U_{j-1}^n) = 0, \tag{43}$$

where

$$\gamma := \Delta t/h, \qquad \alpha := a\gamma,$$

and Δt denotes the fixed time step.

The dispersion relation pertaining to the linear equation

$$U_j^{n+1} - U_j^{n-1} + \alpha(U_{j+1}^n - U_{j-1}^n) = 0 \tag{44}$$

is obtained by assuming a solution of the form

$$U_j^n = e^{i(kx_j - \omega t_n)}$$

and is given by

$$\sin(\omega_k \Delta t) = \alpha \sin(kh). \tag{45}$$

This gives two values of ω_k for each k,

$$\omega_k^1 = \frac{1}{\Delta t} \arcsin(\alpha \sin(kh)) \tag{46a}$$

$$\omega_k^2 = \frac{1}{\Delta t}(\pi - \Delta t\, \omega_k^1). \tag{46b}$$

The main, qualitative, difference between (45) and (5) lies in (46b). Waves travelling at speeds according to this value of ω_k are of purely numerical origin and are known as parasitic waves. It will be shown that these waves are the primary, but not only, destabilising agent in (43).

8. *Grid Resonances*

We proceed as in Part I and assume a solution for (13) of the form

$$\tilde{U}_j^n = \varepsilon \left(\sum_{l=1}^{2} e^{i(kx_j - \omega_k^l t_n)} + \text{c.c.} \right) + O(\varepsilon^2). \tag{47}$$

In order to keep our notation as simple as possible we prefer not to assign different weights to the two waves $l=1$ and $l=2$ in (47), although the relative weights depend on the way the starting values, U_j^1, are provided in (43).

Assuming a perturbed solution of (43) of the form

$$U_j^n = \tilde{U}_j^n + \phi_j^n, \tag{48}$$

where $\phi_j^n \ll \tilde{U}_j^n$ and

$$\phi_j^n = \xi \sum_{m=0}^{(1/2)N} (\alpha_m^n e^{imx_j} + \text{c.c.}), \qquad \xi \ll \varepsilon, \tag{49}$$

we may substitute (48) into (43) to obtain to first order in ϕ_j^n,

$$\begin{aligned} &\phi_j^{n+1} - \phi_j^{n-1} + \alpha(\phi_{j+1}^n - \phi_{j-1}^n) + \gamma\theta(\tilde{U}_{j+1}^n \phi_{j+1}^n - \tilde{U}_{j-1}^n \phi_{j-1}^n) \\ &\quad + \gamma(1-\theta)[(\tilde{U}_{j+1}^n - \tilde{U}_{j-1}^n)\phi_j^n + \tilde{U}_j^n(\phi_{j+1}^n - \phi_{j-1}^n)] = 0. \end{aligned} \tag{50}$$

If we substitute (47) and (49) into (50) and collect coefficients of e^{isx_j}, we obtain

$$\begin{aligned} &(\alpha_s^{n+1} - \alpha_s^{n-1})/(2\Delta t) + i w_s \alpha_s^n \\ &\quad + i\varepsilon C_s \left(\sum_{l=1}^{2} e^{-i\omega_k^l t_n} \right) (\alpha_{s-k}^n + \alpha_{k-s}^{n*}) \\ &\quad + i\varepsilon D_s \left(\sum_{l=1}^{2} e^{i\omega_k^l t_n} \right) (\alpha_{s+k}^n + \alpha_{N-(k+s)}^{n*}) = 0, \end{aligned} \tag{51}$$

where C_s and D_s are given by (17) and

$$w_s = (a/h)\sin(sh). \tag{52}$$

As before, we choose $\alpha_q(\alpha_q^*)$, where $0 \leqslant q \leqslant \frac{1}{2}N$.

In order to develop our intuition for the qualitative effect of the nonlinear interaction, we again consider the $O(\varepsilon)$ terms in (51) to be source terms for the homogeneous equation.

$$(\alpha_s^{n+1} - \alpha_s^{n-1})/(2\,\Delta t) + i w_s \alpha_s^n = 0. \tag{53}$$

The solution of (53) is of the form

$$\alpha_s^n \sim e^{-i\theta_s t_n}, \tag{54}$$

where

$$\sin(\Delta t\,\theta_s) = \Delta t\, w_s. \tag{55}$$

It is clear that (55) allows two values for θ_s, namely

$$\theta_s^l = \omega_s^l, \qquad l = 1, 2,$$

where ω_s^l is given by (46).

Substituting (54) into the "source terms" in (51), we find that secular terms appear and α_s will grow linearly in time if

$$\omega_s^m = \omega_k^l + \omega_{s-k}^p \tag{56a}$$

or

$$\omega_s^m = -\omega_k^l + \omega_{s+k}^p, \qquad m, l, p = 1, 2. \tag{56b}$$

The resonance condition (56a) is satisfied if

$$s = k; \qquad m = l = 1, 2; \qquad p = 1 \tag{57a}$$

and

$$s = \tfrac{1}{2}N; \qquad m = 2; \qquad l \neq p; \qquad \text{and} \qquad l, p = 1, 2. \tag{57b}$$

Condition (56b) is satisfied if

$$s = \tfrac{1}{2}N - k; \qquad m = 2; \qquad l = 1,\ p = 2. \tag{57c}$$

It is already clear from conditions (57b) and (57c) that the two modes, $\frac{1}{2}N$ and $\frac{1}{2}N - k$, should become unstable due to the numerical errors given by (49). We also

observe that these conditions can only be satisfied if both modes, $l=1, 2$, are present. This prediction will be tested numerically in Section 10. First we investigate the stability of the modes in the vicinity of the resonance conditions (57).

9. *Formal Multiple Scales Analysis*

Having extracted the essential information in the form of (56) and (57) from the fully discrete dispersion relation (55), we now proceed to a formal multiple scales analysis. For this purpose we define the leapfrog operator at a time t as

$$\partial_t U(t) = [U(t+\Delta t) - U(t-\Delta t)]/(2\,\Delta t).$$

Furthermore, we expand

$$\alpha_s(t) = \sum_j \varepsilon^j \alpha_s^j(T_0, T_1, \ldots), \tag{58a}$$

where the slow variables T_j are given by

$$T_j = \varepsilon^j t, \qquad j = 0, 1, \ldots . \tag{58b}$$

Schoombie [9] constructed discrete linear operators ∂_{T_j} satisfying

$$\partial_t = \sum_{j=0}^{\infty} \varepsilon^j \partial_{T_j}. \tag{59}$$

In this paper we do not make use of the actual form of the discrete operators ∂_{T_j} appearing in (59). Instead, the qualitative features of the time discretization as reflected in (56) and (57) are emphasized which allows us to treat the discrete operators similar to their continuous counterparts. The quantitative features where the structure of the operators appearing in (59) are taken into account will be discussed elsewhere.

9.1. $s = k \pm r$

We first investigate resonances in the vicinity of k and put

$$s = k \pm r$$

where r is small. The resonance conditions (56) become

$$\omega_{k+r}^l = \omega_k^l + \omega_r^1 + \varepsilon\delta_+ \tag{60a}$$

$$\omega_{k-r}^l = \omega_k^l - \omega_r^1 - \varepsilon\delta_-, \qquad l = 1, 2. \tag{60b}$$

It also follows from (52) and (46) that

$$w_{k\pm r}=\frac{1}{\Delta t}\left[\sin(\omega_k^l\pm\omega_r^1)\,\Delta t\pm\varepsilon\delta_\pm\cos(\omega_k^l\pm\omega_r^1)\,\Delta t\right] \tag{61a}$$

$$w_r=\frac{1}{\Delta t}\left[\sin(\tfrac{1}{2}(\omega_{k+r}^l-\omega_{k-r}^l)\,\Delta t)-\tfrac{1}{2}\varepsilon(\delta_+ +\delta_-)\cos(\tfrac{1}{2}(\omega_{k+r}^l-\omega_{k-r}^l)\,\Delta t)\right]. \tag{61b}$$

If we now substitute (58), (49), and (61) into (50) and proceed as in Part I, it follows that the secular terms at $O(\varepsilon^2)$ are removed by the following equivalent of (27),

$$\frac{dA_{k+r}}{dT_1}+iM_{k+r}^l\delta_+A_{k+r}+iC_{k+r}A_r=0 \tag{62a}$$

$$\frac{dA_{k-r}}{dT_1}-iM_{k-r}^l\delta_-A_{k-r}+iC_{k-r}A_r^*=0 \tag{62b}$$

$$\frac{dA_r}{dT_1}-\tfrac{1}{2}iM_r^l(\delta_+ +\delta_-)\,A_r+iC_rA_{k-r}^*+iD_rA_{k+r}=0, \tag{62c}$$

where

$$M_{k\pm r}^l=\cos((\omega_k^l\pm\omega_r^1)\,\Delta t), \tag{63a}$$

$$M_r^l=\cos(\tfrac{1}{2}(\omega_{k+r}^l-\omega_{k-r}^l)\,\Delta t). \tag{63b}$$

C_j and D_j are given by (17).

In order to solve (62) we make a few simplifying assumptions. Numerical calculations show it to be reasonable to take, for $rh\ll 1$,

$$\delta_+=\delta_-=\delta,\qquad M_{k+r}^l=M_{k-r}^l=M^l,\qquad M_r^l=1.$$

If we now assume that $A_{k\pm r}$ is of the form

$$A_{k\pm r}\sim e^{i\lambda_\pm T_1},$$

it follows that

$$\lambda_\pm^4\mp 2\delta(1-M^l)\,\lambda_\pm^3-\{\Delta+2M^l\delta^2-(1-M^l)^2\,\delta^2\}\,\lambda_\pm^2$$
$$\mp(M^l-1)\,\delta(\Delta+2M^l\delta^2)\,\lambda_\pm+M^l\delta^2(M^l\delta^2+\Delta)=0, \tag{64}$$

where Δ is given by (29). This equation is readily solved to give

$$\lambda_+=\begin{cases}\delta\\ -M^l\delta\\ \tfrac{1}{2}[\delta(1-M^l)\pm\{\delta^2(1-M^l)^2+4(\Delta+M^l\delta^2)\}^{1/2}].\end{cases} \tag{65}$$

The expression for λ_- is obtained by substituting δ with $-\delta$. It is clear that we will have an instability on the T_1 time scale if

$$\delta^2(1+M^l)^2+4\Delta<0. \tag{66}$$

This condition reduces to (30) if $M^l=1$. A closer look at (63a) reveals that

$$0\leqslant M^1\leqslant 1, \tag{67a}$$

$$M^2=-M^1 \tag{67b}$$

For small values of α, roughly $\alpha\leqslant 0.5$, $M^1=1$; for larger values of α there is a significant decrease in the value of M^1 for most values of k. It is clear from (66) and (60) that values of M^l less than 1 have a destabilizing effect, since it implies that instability will occur for bigger values of δ, hence for smaller values of ε.

If, in Part I, we needed

$$\varepsilon>\varepsilon_I$$

for instability, we now need (approximately)

$$\varepsilon>\tfrac{1}{2}(1+M^1)\,\varepsilon_I. \tag{68}$$

However, a more severe instability occurs in the case of the parasitic wave which corresponds to the choice M^2 in (66). In this case the bound on ε becomes

$$\varepsilon>\tfrac{1}{2}(1-M^1)\,\varepsilon_I. \tag{69}$$

Since M^1 may approach 1, (69) provides a bound on the amplitude which may be well inside the limit imposed by the nonlinear instability condition.

9.2. $s=\frac{1}{2}N-r$

For values of s satisfying

$$s=\tfrac{1}{2}N-r,$$

where r is small, let

$$\omega^2_{(1/2)N-r}=\omega^l_k+\omega^p_{(1/2)N-k-r}-\varepsilon\delta, \tag{70}$$

where from (57b), $l\neq p$ and $l, p=1, 2$. Proceeding in a similar way as before, we obtain

$$\frac{dA_{(1/2)N-r}}{dT_1}-i\delta M_+A_{(1/2)N-r}+iC_{(1/2)N-r}A_{(1/2)N-(k+r)}=0 \tag{71a}$$

$$\frac{dA_{(1/2)N-(k+r)}}{dT_1}+i\delta M_-A_{(1/2)N-(k+r)}+iD_{(1/2)N-(k+r)}A_{(1/2)N-r}=0, \tag{71b}$$

where

$$M_+ = \cos((\omega_k^l + \omega^p_{(1/2)N-(k+r)})\,\Delta t)$$
$$M_- = \cos((\omega_k^l - \omega_r^2)\,\Delta t).$$

Assuming

$$A_{(1/2)N-r} \sim e^{i\lambda T_1},$$

and similarly for $A_{(1/2)N-(k+r)}$, it follows from (71) that the eigenvalues λ for both $A_{(1/2)N-r}$ and $A_{(1/2)N-(k+r)}$ satisfy

$$\lambda = (M_+ - M_-)\,\delta \pm \{(M_+ + M_-)^2\,\delta^2 + 4C_{(1/2)N-r}D_{(1/2)N-(k+r)}\}^{1/2}. \tag{72}$$

Clearly we have instability on the T_1 time scale if

$$(M_+ + M_-)^2\,\delta^2 + 4C_{(1/2)N-r}D_{(1/2)N-(k+r)} < 0, \tag{73}$$

since $C_s \geqslant 0$; instability will only occur if

$$D_{(1/2)N-(k+r)} < 0. \tag{74}$$

Thus we are again provided with a lower bound on ε. In Table II, this bound was calculated using $N = 120$, $\gamma = 1$, $\alpha = 0.5$, and $r = 1$. We also used $l = 1$ for $k < \frac{1}{4}N$ and $l = 2$ for $k > \frac{1}{4}N$ in (70), since these values correspond to the strongest resonance.

The final resonance condition (57c) provides us with a system similar to (71), but involving $A_{(1/2)N-r}$ and $A_{(1/2)N-k+r}$. This shows that the instabilities in $A_{(1/2)N-r}$, $A_{(1/2)N-k-r}$, and $A_{(1/2)N-k+r}$ are connected. This will be verified numerically in the next section.

10. *Numerical Results*

The numerical results given by Briggs *et al.* [1] and Sloan and Mitchell [10] illustrate many of our theoretical results. For instance, Sloan and Mitchell give

TABLE II

Minimum Values of ε Required for Instability

k \ θ	0	0.5
15	6.75×10^{-3}	—
45	1.70×10^{-4}	2.3×10^{-3}

Note. There is no entry for $k = 15$ and $\theta = 0.5$, since $D_{(1/2)N-(k+r)} > 0$ for these values.

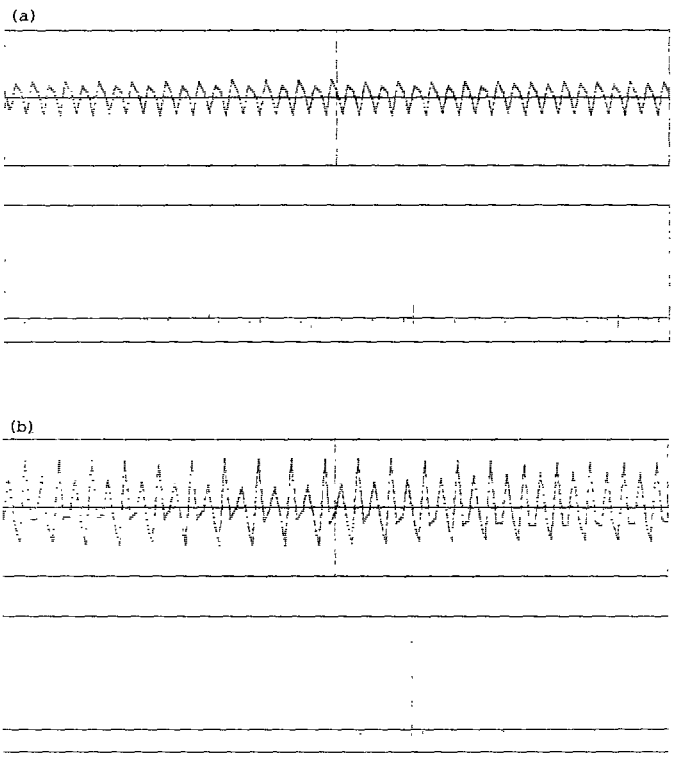

FIG. 1. Solution and its Fourier transform after (a) 325, (b) 350.

numerical and theoretical evidence that the instability is related to the leap-frog time discretization. They also show that the instabilities are absent if there is insufficient energy in the fundamental mode. Briggs *et al.* give extensive numerical evidence of the destabilization of the modes in the vicinity of the fundamental mode which formed the basis of the theoretical investigations of Sloan and Mitchell.

Hence, we concentrate on those results of our theoretical investigations which, in our opinion, have not yet been fully appreciated. This will both serve as a verification of our theoretical results and also improve our understanding of the mechanism causing the instabilities. All the numerical experiments were done on a Sperry Micro IT utilizing a 80287 coprocessor.

The resonance conditions (57b) and (57c) predict a growth in modes $\frac{1}{2}N$ and $\frac{1}{2}N-k$. In order to test this prediction, we solve (43) numerically with $N=120$, $\gamma=1$, $\alpha=0.9$, and $\theta=0$. The initial condition is given by

$$U_j^0 = U_j^1 = \tfrac{1}{2}\sigma(1+i)\, e^{ikx_j} + \text{c.c.} \tag{75}$$

with $k=\frac{1}{3}N$ and $\sigma=0.05$. Note that in these results no disturbances were added to

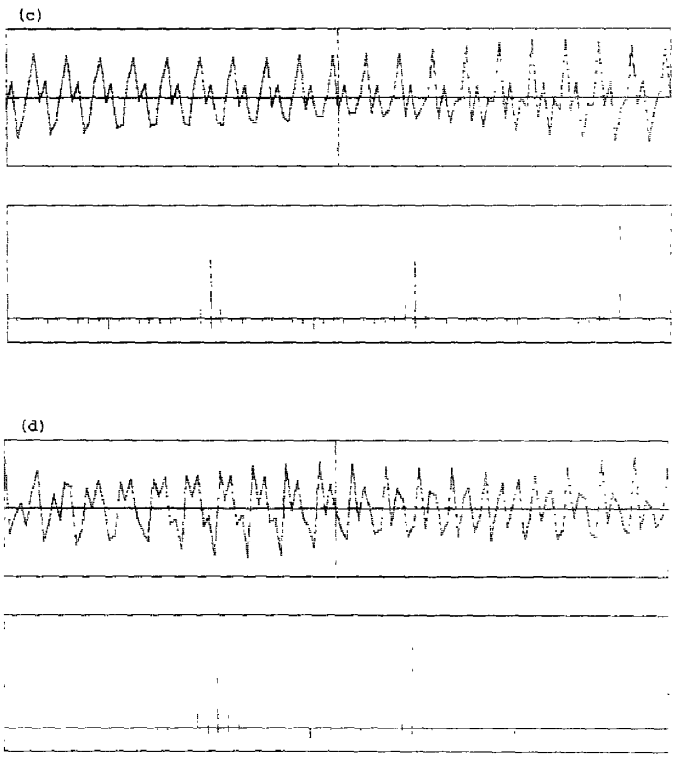

FIG. 1 (*continued*). (c) 375, (d) 400 time steps using the initial condition (75).

stimulate any additional modes. The results after 325, 350, 375, and 400 time steps are shown in Fig. 1. Note that in all our experiments we only show half of the Fourier modes, the other half being symmetrical. After 325 time steps the modes $\frac{1}{2}N$ and $\frac{1}{2}N-k$ appear, as predicted by (57b) and (57c). After 350 time steps more modes appear in the vicinity of the fundamental mode and the modes $\frac{1}{2}N$ and $\frac{1}{2}N-k$, as predicted by the multiple scales analysis of Section 9. From (62) we also expect the low frequency modes to appear. Although these modes do not show up in Fig. 1, they were observed in many of our experiments, cf. for instance, Fig. 2.

We observed in Section 7 that the main qualitative difference between the semi-discrete equations (3) and the leapfrog system (43) is the parasitic wave allowed by the latter. Also in Section 9, e.g., (69), we contributed the cause of instability to the parasitic wave. Accordingly, we conducted numerical experiments with the following initial conditions

$$U_j^0 = A_0 e^{ikx_j} + \text{c.c.} \tag{76a}$$

$$U_j^1 = r_1 A_0 e^{ikx_j} + \text{c.c.} \tag{76b}$$

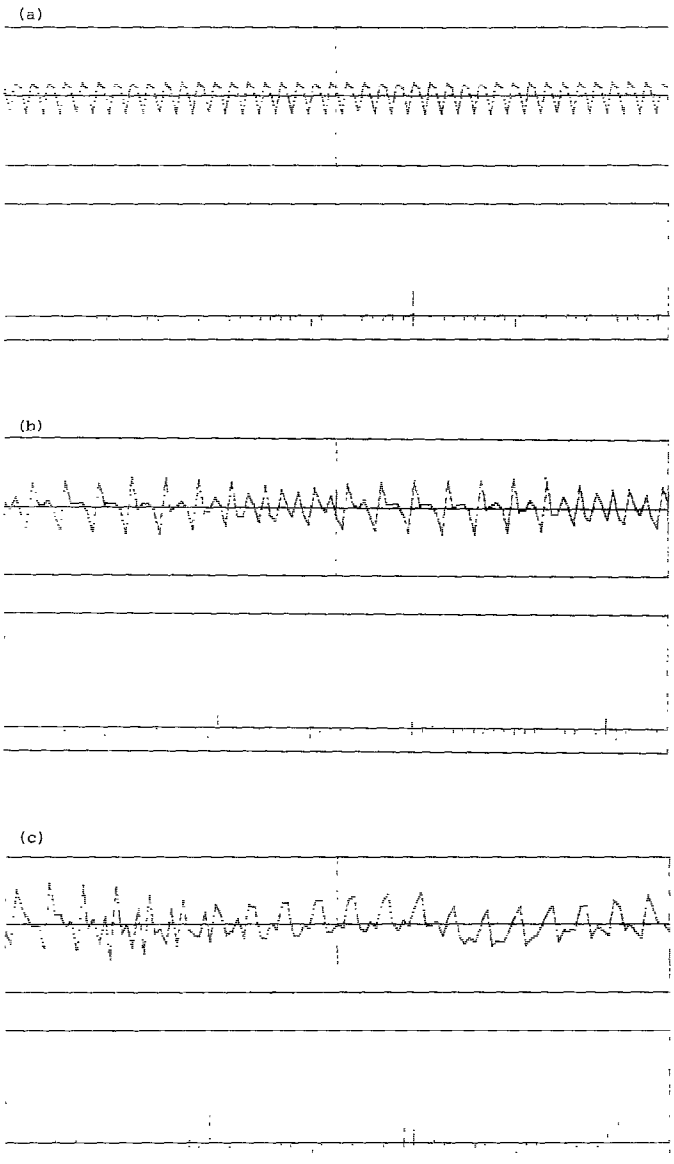

FIG. 2. Solution and its Fourier transform after (a) 400, (b) 600, (c) 1000 time steps using the initial condition (76).

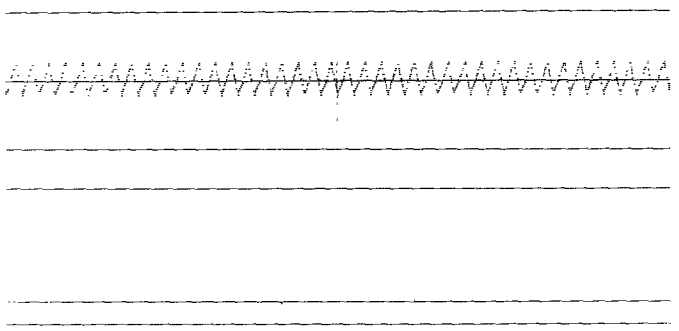

FIG. 3 Solution and its Fourier transform after 1000 time steps using the initial condition (77).

and

$$U_j^0 = A_0 e^{ikx_j} + \text{c.c.} \tag{77a}$$

$$U_j^1 = r_2 A_0 e^{ikx_j} + \text{c.c.}, \tag{77b}$$

where

$$A_0 := \tfrac{1}{2}\sigma(1+i)$$

$$r_1 := -i\alpha \sin(kh) + (1-\alpha^2 \sin^2(kh))^{1/2} \tag{78a}$$

$$r_2 := -i\alpha \sin(kh) - (1-\alpha^2 \sin^2(kh))^{1/2}. \tag{78b}$$

These initial conditions were obtained by applying the leapfrog scheme to the linearization of (7), i.e., they were obtained from

$$A_{n+1} - A_{n-1} + 2i\alpha \sin(kh)\, A_n = 0. \tag{79}$$

It is easily seen that

$$A_1 = r_1 A_0$$

TABLE III

Minimum Values of ε Required for Instability for Different Values of r ($\theta = 0$, $N = 120$, $a = 0.5$)

r	1	2	3	4	5	6	7	8	9	10	11
$\varepsilon \times 10^{-4}$ for $k = 50$	0.31	0.42	0.37	0.22	0.00	0.22	0.40	0.47	—	—	—
$\varepsilon \times 10^{-3}$ for $k = 40$	0.28	0.46	0.57	0.61	0.59	0.53	0.43	0.31	0.16	0.00	0.33

eliminates the parasitic contribution from (79) and

$$A_1 = r_2 A_0$$

eliminates the physical contribution from (79). Thus, (76) admits only a small contribution from the parasitic wave and (77) only a small contribution from the physical wave.

Figure 2 shows the results obtained from (76), the values of all the parameters are as in Fig. 1. After 400 time steps there is no indication of any instability, as compared to the significant instability observed at the same time in Fig. 1, where both waves were present. From 600 to 1000 time steps the instability has developed in much the same fashion as in Fig. 1. In this case, however, we do note the appearance of low wave numbers. The instability need not surprise us; the choice (76) does not remove all of the parasitic wave from the nonlinear problem.

Figure 3 shows the result obtained from (77) after 1000 time steps. Perhaps surprisingly, there is no evidence of any instability. However, the meaning of (57b) and (57c) is now clear. We need both waves for these resonance conditions to be satisfied. The instability appears as a result of the resonance caused by an interaction of the two waves.

Finally we observe that our analysis may also be able to account for the observation of Sloan and Mitchell [2] that the most unstable side-mode is not necessarily the one closest to the fundamental mode. In our terminology this means that the modes with $r = 1$ are not always the most unstable ones. Some indication of this is obtained from (73). In Table III the bounds on the amplitude ε required for instability for various values of r as calculated from (70) and (73), are shown.

11. *Conclusions*

We have demonstrated how the interaction between the physical wave and the parasitic wave, arising from a leapfrog time discretization, may be responsible for the instabilities described by Briggs *et al.* [1]. Although we believe this to shed more light on the structure of the mechanism responsible for the instability, some questions remain unanswered. For instance, our analyses are not sharp enough to predict the time of the onset of the instability. As a result the significant difference between the results obtained from (76) and (77) remains puzzling.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the financial support of the C.S.I.R., Pretoria, for this research. We are also grateful to Dr. Dave Sloan for his encouragement and many useful discussions as well as the reviewers for their comments.

Note added in proof. We are now in a position to improve on the approximations made in the multiple scales analyses. Hopefully this will remove some of the remaining puzzles.

REFERENCES

1. W. L. BRIGGS, A. C. NEWELL, AND T. SARIE, *J. Comput. Phys.* **50**, 83 (1983).
2. B. FORNBERG, *Math. Comput.* **27**, 45 (1973).
3. A. JEFFREY AND T. KAWAHARA, *Asymptotic Methods in Nonlinear Wave Theory* (Pitman, Boston, 1982).
4. D. W. JORDAN AND P. SMITH, *Nonlinear Ordinary Differential Equations* (Oxford Press, Clarendon, Oxford, 1977).
5. D. W. MOORE, *IMA J. Appl. Anal.* **31**, 1 (1983).
6. A. C. NEWELL, *SIAM J. Appl. Math.* **33**, 133 (1977).
7. R. D. RICHTMYER AND K. W. MORTON, *Difference Methods for Initial Value Problems* (Interscience, New York, 1967).
8. J. M. SANZ-SERNA, *SIAM J. Sci. Statist. Comput.* **6**, 923 (1985).
9. S. W. SCHOOMBIE, Department of Applied Mathematics, University of the Orange Free State, Bloemfontein, private communication (1986).
10. D. M. SLOAN AND A. R. MITCHELL, *J. Comput. Phys.* **67**, 372 (1986).
11. F. VADILLO AND J. M. SANZ-SERNA, *J. Comput. Phys.* **66**, 225 (1986).
12. G. B. WHITHAM, *Linear and Nonlinear Waves* (Wiley, New York, 1974).