Macroscopic dynamics in quadratic nonlinear lattices

Peter D. Miller* and Ole Bang
Australian Photonics Cooperative Research Centre, Optical Sciences Centre, Research School of Physical Sciences and Engineering, The Australian National University, Canberra, Australian Capital Territory 0200, Australia
(Received 26 August 1997)

Fully nonlinear modulation equations are obtained for plane waves in a discrete system with quadratic nonlinearity, in the limit when the modulational scales are long compared to the wavelength and period of the modulated wave. The discrete system we study is a model for second-harmonic generation in nonlinear optical waveguide arrays and also for exciton waves at the interface between two crystals near Fermi resonance. The modulation equations predict their own breakdown by changing type from hyperbolic to elliptic. Modulational stability (hyperbolicity of the modulation equations) is explicitly shown to be implied by linear stability *but not vice versa*. When the plane-wave parameters vary slowly in regions of linear stability, the modulation equations are hyperbolic and accurately describe the macroscopic behavior of the system whose microscopic dynamics is locally given by plane waves. We show how the existence of Riemann invariants allows one to test modulated wave initial data to see whether the modulating wave will avoid all linear instabilities and ultimately resolve into simple disturbances that satisfy the Hopf or inviscid Burgers equation. We apply our general results to several important limiting cases of the microscopic model in question. [S1063-651X(98)11804-4]

PACS number(s): 03.40.Kf, 63.10.+a, 73.50.Fq, 42.82.Et

I. INTRODUCTION

Many nonlinear dispersive equations modeling physical systems have multiparameter families of exact wave-train solutions. Letting the parameters of the wave train be slowly varying functions of space and time typically leads to a dynamical description of the parameters, the so-called *modulation equations*, where nonlinearity dominates over dispersion. Dynamics taking place on the scales of a wavelength and a period are said to be short wave or *microscopic*, while the modulation equations are concerned with the long-wave or *macroscopic* dynamics of the slowly varying parameters. Indeed, the modulation equations asymptotically extend the family of exact wave-train solutions through the introduction of a small parameter ϵ, the ratio between the microscopic and macroscopic scales, which is not present in the underlying microscopic model.

Modulation equations can be obtained formally in several ways, all connected with the method of averaging. Asymptotic expansions using the method of multiple scales can reveal the modulation equations as a solvability condition [1], averaged variational principles can be exploited for Lagrangian systems [2], and local conservation laws can be averaged in an approach that has a somewhat more physical appeal (see, e.g., [3,4] for applications in a nonintegrable discrete setting in closest analogy with what will follow below). When the modulation equations are meaningful, they encode important physical information and can be used to make predictions about macroscopic wave motion and dynamics.

However, when are the formal modulation equations meaningful? One clue comes from the equations themselves. As a set of first-order quasilinear equations (derivatives with respect to space and time enter linearly), the modulation equations can be either hyperbolic or elliptic, perhaps depending on the local values of the wave parameters. If they are elliptic, then the initial value problem of how a wave train with given spatial variation of its parameters evolves in time is ill posed and the modulation equations are not valid as a model of dynamics (although some significance has been attributed to their self-similar solutions [5]). If they are hyperbolic, then the initial value problem can be solved until the time of breaking, when infinite spatial derivatives appear or until the time of change of type when hyperbolic dynamics pushes part of the modulated wave into a region where the equations are locally elliptic, whichever comes first.

However, the fact that the initial value problem can be solved in some time interval does not necessarily mean that the modulation equations are a good model for dynamics in the limit of small-scale ratios ϵ. In fact, the well posedness of the initial value problem for modulated waves is only an indication of stability of those waves to perturbations of nearby wavelength, i.e., to relatively long waves. If other short-wave instabilities are present, then they may obstruct the process of smooth modulation of the waves without being detected by the formal modulation equations themselves. The relevance of this observation has not been explored in the literature, mainly due to a lack of examples of microscopic systems having at the same time the following essential features: a rich multiparameter family of nonlinear traveling-wave solutions, sufficient structure to control slow modulations of the waves in the family, and a set of relatively short-wave instabilities that can exist without long-wave instabilities being present as well.

In this paper we study the dynamics of modulated waves in such a microscopic system where the well posedness of the initial value problem for the modulation equations is not sufficient to guarantee their validity as a model. The microscopic system describes pairs of resonantly interacting nonlinear oscillators arranged on a one-dimensional lattice

*Present address: School of Mathematics, Institute for Advanced Study, Olden Lane, Princeton, NJ 08540.

 © 1998 The American Physical Society

whose sites are labeled with integers n:

$$i\partial_t w_n + w_{n+1} + w_{n-1} + w_n^* v_n = 0,$$

$$i\partial_t v_n + \gamma(v_{n+1} + v_{n-1}) - \alpha v_n + \frac{1}{2} w_n^2 = 0. \tag{1}$$

Here γ and α are real parameters. Without the nearest-neighbor coupling terms, this system is the generic envelope equation for the weakly nonlinear interaction of two (not necessarily mechanical) oscillators in near second-harmonic resonance. To see this, one supposes a Hamiltonian of the form

$$H = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{k_1 q_1^2}{2} + \frac{k_2 q_2^2}{2} + U(q_1, q_2), \tag{2}$$

where the interaction energy U is of third order in the displacements and the near resonance condition is satisfied:

$$\sqrt{\frac{k_2}{m_2}} = 2\sqrt{\frac{k_1}{m_1}} - \epsilon\alpha, \tag{3}$$

with ϵ being a small parameter and α a fixed detuning. For small amplitudes, the two oscillators have displacements of the form

$$q_1 = \epsilon \frac{2\sqrt{k_1}}{U_{112}} \sqrt{2m_2}\, w(\epsilon t) e^{i\Omega t} + \text{c.c.} + O(\epsilon^2),$$

$$q_2 = \epsilon \frac{2\sqrt{k_1}}{U_{112}} \sqrt{m_1}\, v(\epsilon t) e^{2i\Omega t} + \text{c.c.} + O(\epsilon^2), \tag{4}$$

where $\Omega = \sqrt{k_1/m_1}$ is the fundamental frequency, $U_{112} = \partial_{q_1}^2 \partial_{q_2} U(0,0)$, and c.c. denotes the complex conjugate. The envelopes $w(\epsilon t)$ and $v(\epsilon t)$ satisfy the system (1) with the nearest-neighbor coupling terms neglected and with the time derivatives taken with respect to the slow time ϵt. The nonlinear interaction terms are generic for resonant interactions because they dominate whenever U_{112} is nonzero. A weakly coupled lattice of such pairs of resonantly interacting oscillators can be described in the limit $\epsilon \downarrow 0$ with the inclusion of the nearest-neighbor coupling terms, where γ quantifies the relative coupling strengths of the second and first harmonics.

In terms of concrete physical applications, the coupled system of equations (1) is used to model the exciton waves that propagate along the interface between two crystals when the crystals are near Fermi resonance [6,7]. Fermi resonance means exactly that the exciton frequency in one crystal is twice that in the other crystal. In this model, $w_n(t)$ and $v_n(t)$ are interpreted as expectation values of annihilation operators for the bosonic excitations of the two kinds of atoms that are adjacent at the interface at the site n. The same model also describes second-harmonic generation in arrays of optical waveguides, each having core material with a nonzero value of $\chi^{(2)}$, the quadratic nonlinear susceptibility tensor [8,9]. In this case, t is the spatial coordinate of distance along each waveguide, n labels the individual waveguides of the array, and $w_n(t)$ and $v_n(t)$ are the amplitudes of the first- and second-harmonic electric fields in and along each core.

In Sec. II we will describe the family of traveling-wave solutions of the system (1) that will form the basis of our study. The waves will be characterized by their nonlinear dispersion relations, which imply the existence of a band gap in the wave spectrum. By introducing a uniformly small perturbation of fixed relative wave number and linearizing, we will find the linear stability criterion for the waves.

Section III will begin to address the issue of slowly modulated waves, where the parameters of the plane waves, previously held constant, are allowed to be slowly varying functions of space and time. The formal derivation of explicit, fully nonlinear modulation equations will be motivated by numerical simulations of the system (1).

Section IV is the heart of our paper, where we analyze the formal modulation equations. We first do this in order to understand their validity and to underscore our main point: that *short-wave instabilities that can destroy a modulating wave train can be hidden from the modulation equations*. Then we work in regions of linear stability where there is little doubt that the modulation equations are a valid model for modulating wave dynamics and study their implications using the powerful tool of the Riemann invariants.

In Sec. V we explore relevant special cases of the microscopic model (1). First we examine the (dispersive) continuum limit that leads to the equations describing optical solitons in homogeneous $\chi^{(2)}$ media [10] and Fermi resonance interface solitons [6] and show how some of the stability results for plane waves in the continuum model [11,12] can be deduced from the discrete problem. Indeed, the theory of nonlinear plane waves in the continuum model can be reconstructed asymptotically from the neighborhood of the zero solution of Eqs. (1) when $\alpha \approx 2\gamma - 4$. We then briefly consider two discrete limits of Eqs. (1) in which the dynamics simplify and in fact the distinction between the well posedness and validity of the modulation equations becomes blurred: the limit of large detuning α and the limit of high-frequency waves. The former limit reproduces the discrete nonlinear Schrödinger equation, for which the modulation theory of plane waves is simple by comparison [3].

Finally, we conclude in Sec. VI with a brief discussion of unanswered questions.

II. EXACT NONLINEAR PLANE WAVES AND LINEAR STABILITY

The system of equations (1) has a two-parameter family of exact nonlinear plane-wave solutions of the form

$$w_n(t) = We^{i(kn - \omega t)}, \quad v_n(t) = Ve^{2i(kn - \omega t)}, \tag{5}$$

where k, ω, W, and V are real constants satisfying the nonlinear dispersion relations

$$V = -(\omega + 2\cos k),$$

$$W^2 = 2(2\omega - \alpha + 2\gamma \cos 2k)(\omega + 2\cos k). \tag{6}$$

Further solutions can be found using the gauge symmetry of Eqs. (1) for any real constant θ,

$$w_n(t) \mapsto w_n(t) e^{i\theta}, \quad v_n(t) \mapsto v_n(t) e^{2i\theta}, \tag{7}$$

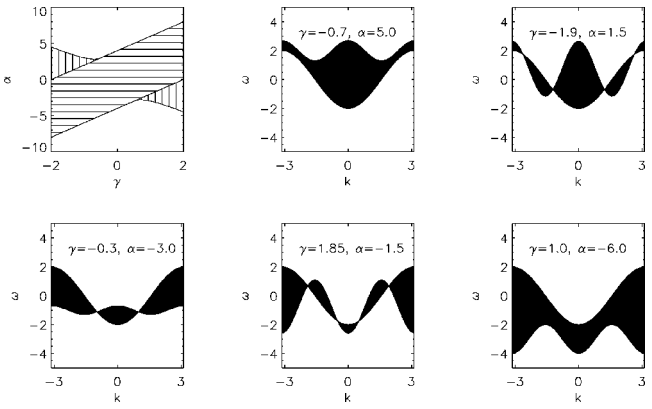

FIG. 1. Band-gap-structure dependence on the parameters γ and α. In the upper left plot, regions of the (γ,α) plane in which the band gap does not twist off at all are blank, regions where there is one twist in the band gap for positive k are shown with horizontal lines, and regions where there are two twists for positive k are shown with vertical lines. The remaining plots show representative band diagrams with the band gap filled in.

which leaves the equations unchanged. We view k and ω as independent parameters, with the real amplitudes W and V determined from Eqs. (6).

In order for W to be real, plane-wave solutions cannot exist everywhere in the (k,ω) plane. Since W^2 is clearly positive for large $|\omega|$, solutions will fail to exist in a certain band gap whose shape is determined by the parameters α and γ. The band gap will pinch off at those values of (k,ω) for which the two factors of W^2 both vanish together. In the range of $0\leqslant k\leqslant\pi$ (the band diagram is even and 2π periodic in k), this can happen once, twice, or not at all. There are two crossings if

$$\gamma<-1/2,\quad 2\gamma+4<\alpha<-1/\gamma-2\gamma \tag{8}$$

or if

$$\gamma>1/2,\quad -1/\gamma-2\gamma<\alpha<2\gamma-4. \tag{9}$$

There is one crossing if $|\alpha-2\gamma|<4$ and there are no crossings otherwise. These three regions of the (γ,α) plane are shown in Fig. 1, along with band diagrams in the (k,ω) plane representative of zero, one, and two crossings. The existence of a gap in the spectrum of nonlinear waves is a feature related to the two-component resonant nature of the system (1). The intriguing shape of the band gap has been studied recently in the context of optics [8].

The stability of the nonlinear plane waves can be investigated by choosing constants V, W, k, and ω that satisfy the dispersion relations (6) and then looking for solutions of Eqs. (1) of the form

$$w_n(t)=[W+\tilde{w}_n(t)]e^{i(kn-\omega t)},$$

$$v_n(t)=[V+\tilde{v}_n(t)]e^{2i(kn-\omega t)}, \tag{10}$$

where the small perturbations have spatial structure of relative wave number β:

$$\tilde{w}_n(t)=[\tilde{w}_{\beta,R}(t)+i\tilde{w}_{\beta,I}(t)]e^{i\beta n},$$

$$\tilde{v}_n(t)=[\tilde{v}_{\beta,R}(t)+i\tilde{v}_{\beta,I}(t)]e^{i\beta n}. \tag{11}$$

The linearized system is then reduced to a fourth-order system for each perturbative wave number β:

$$\partial_t\begin{bmatrix}\tilde{w}_{\beta,R}\\ \tilde{w}_{\beta,I}\\ \tilde{v}_{\beta,R}\\ \tilde{v}_{\beta,I}\end{bmatrix}=\begin{bmatrix}iA & V-C & 0 & -W\\ V+C & iA & W & 0\\ 0 & -W & iB & -D\\ W & 0 & D & iB\end{bmatrix}\begin{bmatrix}\tilde{w}_{\beta,R}\\ \tilde{w}_{\beta,I}\\ \tilde{v}_{\beta,R}\\ \tilde{v}_{\beta,I}\end{bmatrix}, \tag{12}$$

where

$$A=-2\sin k\sin\beta,\quad B=-2\gamma\sin 2k\sin\beta,$$

$$C=\omega+2\cos k\cos\beta,\quad D=2\omega-\alpha+2\gamma\cos 2k\cos\beta. \tag{13}$$

The real parts of the eigenvalues of this matrix are growth rates associated with the mode parametrized by the relative wave number β. The nonlinear plane wave is declared to be *linearly stable* if all four growth rates vanish for all $\beta\in[-\pi,\pi]$. The stability criterion is not easy to express explicitly and analytically as a function of the independent wave parameters k and ω. However, it is easy to compute. In Figs. 4 and 5 we will show the (k,ω) plane for some particular choices of α and γ, indicating the regions of linear instability as calculated from the eigenvalues of the matrix in Eq. (12) with gray shading, but we postpone those figures until we have some information about modulating waves so that we can draw an important conclusion from the comparison.

If the waves are linearly stable, then weakly nonlinear theory can be used to study the small perturbations $\tilde{w}_n(t)$ and $\tilde{v}_n(t)$. In the case of long-wavelength perturbations, Korteweg–de Vries dynamics prevails in the moving frame of each branch of the group velocity of linear waves near $\beta=0$. For some values of the parameters of the underlying plane wave, the dispersion of linear waves at $\beta=0$ may vanish, leading to steepening and shock formation that is regularized by higher-order dispersion. The dynamics of these weakly nonlinear perturbations, which do not have any effect on the background plane wave that supports them, have been recently studied in a one-component discrete system by Konotop and Salerno [13].

III. MODULATED NONLINEAR PLANE WAVES AND MODULATION EQUATIONS

The linear stability analysis described above addresses the question of the behavior of solutions of Eqs. (1) that are *uniformly* close in n to a given plane-wave solution. However, we may also ask whether there exist solutions of Eqs. (1) that are close to a plane-wave solution only locally, in the neighborhood of each fixed n. The plane wave nearest the solution may be different for each neighborhood. The underlying approximate solution of Eqs. (1) is thus generally a

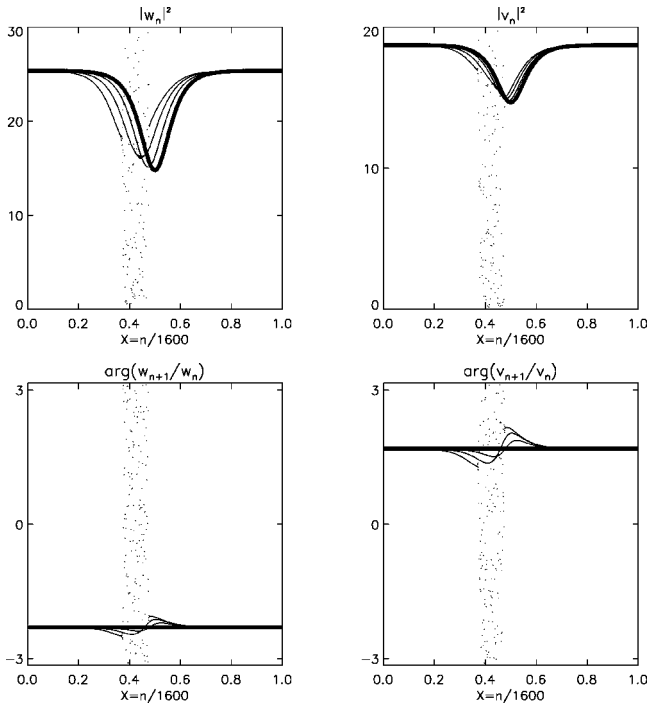

FIG. 2. Snapshots taken 30 time units apart of a numerical simulation of the system (1) with parameter values $\alpha=-3$ and $\gamma=-0.3$ and 1600 points.

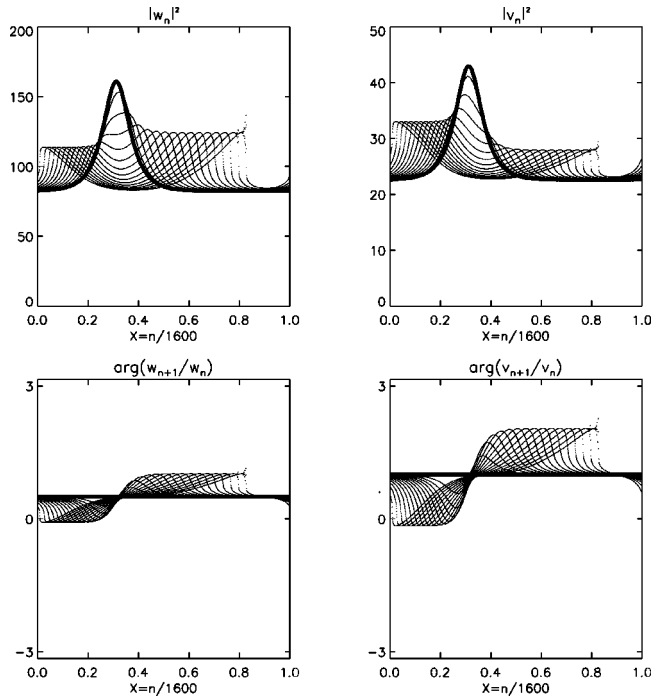

FIG. 3. Snapshots taken 20 time units apart of a numerical simulation of the system (1) with parameter values $\alpha=-3$ and $\gamma=-0.3$ and 1600 points.

slowly modulated plane wave.

To get an idea of what kind of dynamical behavior one might expect of a modulated nonlinear plane wave, we turn to numerical simulation of the microscopic system (1). We prepared a domain of 1600 points with initial conditions that were, near each fixed n, plane waves satisfying the dispersion relations (6). The initial twist $k=\arg(w_{n+1}/w_n)$ of each plane wave was taken to be a constant over the whole domain, while the amplitudes W and V were taken to be slowly varying functions consistent with Eqs. (6) and the choice of k. For convenience, periodic boundary conditions in n were assumed for the whole domain. Snapshots of the quantities $|w_n|^2$, $|v_n|^2$, $\arg(w_{n+1}/w_n)$, and $\arg(v_{n+1}/v_n)$ are shown in Figs. 2 and 3 with the initial conditions highlighted in bold.

It is clear from these simulations that slow variation of the plane-wave parameters in space implies their slow variation in time. In fact, the local period of the microscopic oscillations in Fig. 2 ranges between 2.1 and 2.5 time units compared with 30 time units between snapshots, while in Fig. 3 the local period ranges between 1.3 and 2.1 time units and there are 20 time units between snapshots. It is possible to deduce scaling properties of the slow dynamics with additional numerical experiments; it turns out that pictures that are, to the eye, indistinguishable from these can be made by increasing the number of points sampling the modulated wave by some scaling factor f and then taking snapshots of the numerical simulation at time intervals f times as long. In both simulations, the dynamics of the waves are regular for some time before a local catastrophe appears that breaks the structure of the smoothly modulated plane wave. In Fig. 2 the wave form spontaneously breaks down for apparently no reason, while in Fig. 3 the wave form separates into two parts and steepens until it threatens to become multivalued and something resembling a shock wave appears.

We want to explain these simulations. Slowly modulated plane waves are described by allowing the independent parameters of the family of plane waves, k and ω, to be slowly varying functions of n and t. To proceed, we consider $\epsilon \ll 1$ and introduce equally stretched slow scales $X=\epsilon n$ and $T=\epsilon t$, in accordance with the scaling observations mentioned above. Our immediate goal is to derive evolution equations for the macroscopic quantities $k(X,T)$ and $\omega(X,T)$ from the microscopic dynamics given by Eqs. (1).

We will use the fact that the system of equations (1) implies two local conservation laws

$$\partial_t N_n + F_n - F_{n-1} = 0, \quad \partial_t H_n + G_n - G_{n-1} = 0, \tag{14}$$

with conserved local densities

$$N_n = |w_n|^2 + 2|v_n|^2,$$

$$H_n = \alpha|v_n|^2 - \mathrm{Re}\{2w_n^* w_{n+1} + 2\gamma v_n^* v_{n+1} + w_n^2 v_n^*\} \tag{15}$$

and corresponding fluxes

$$F_n = 2\,\mathrm{Im}\{w_{n+1}w_n^* + 2\gamma v_{n+1}v_n^*\},$$

$$G_n = 2\,\mathrm{Im}\{\alpha\gamma v_n^* v_{n+1} + w_n w_{n+2}^* + \gamma^2 v_n v_{n+2}^* + w_n w_{n+1} v_{n+1}^* + \tfrac{1}{2}\gamma v_n w_{n+1}^{*2}\}. \tag{16}$$

These local forms imply that the sums of N_n and H_n over the periodic domain are global invariants. Evaluating the conserved local densities and their associated fluxes on an exact plane-wave solution with k and ω fixed, one finds that

$$N_n = W^2 + 2V^2,$$

$$F_n = 2(W^2 + 4\gamma V^2 \cos k)\sin k,$$

$$H_n = \alpha V^2 - 2W^2 \cos k - 2\gamma V^2 \cos 2k - W^2 V,$$

$$G_n = -2\gamma^2 V^2 \sin 4k - 2W^2 V \sin k - (2W^2 - 2\alpha\gamma V^2 + \gamma W^2 V)\sin 2k, \tag{17}$$

where the amplitudes W and V are functions of k and ω according to the dispersion relations (6). Note that all of these quantities are *constants* for plane waves and thus in this context are explicit functions of the wave parameters k and ω. Taking k and ω to depend smoothly on X and T and assuming that the dispersion relations continue to hold locally in X and T, we may divide the local conservation laws (14) by ϵ and take the limit $\epsilon \downarrow 0$ to find the macroscopic *modulation equations*

$$\partial_T N + \partial_X F = 0, \quad \partial_T H + \partial_X G = 0. \tag{18}$$

These partial differential equations are themselves local conservation laws, this time in a continuous setting. With an application of the chain rule, these modulation equations can be written in terms of $k(X,T)$ and $\omega(X,T)$ as

$$\partial_T \begin{bmatrix} k \\ \omega \end{bmatrix} + \mathbf{M}(k,\omega)\partial_X \begin{bmatrix} k \\ \omega \end{bmatrix} = \mathbf{0}, \tag{19}$$

where the coefficient matrix is

$$\mathbf{M}(k,\omega) = \begin{bmatrix} 0 & 1 \\ \dfrac{A(k,\omega)}{D(k,\omega)} & \dfrac{B(k,\omega)}{D(k,\omega)} \end{bmatrix} \tag{20}$$

and

$$A(k,\omega) = 2[64\gamma \cos^4 k + 36\gamma\omega \cos^3 k + (8\omega + 4\gamma\omega^2 - 56\gamma - 4\alpha)\cos^2 k + (2\omega^2 - 26\gamma\omega - \alpha\omega)\cos k + 2\alpha + 4\gamma - 4\omega - 2\gamma\omega^2], \tag{21}$$

$$B(k,\omega) = 4[12\gamma \cos^2 k + (4\gamma\omega + 4)\cos k + 4\omega - 2\gamma - \alpha]\sin k, \tag{22}$$

$$D(k,\omega) = 4\gamma \cos^2 k + 8\cos k + 6\omega - 2\gamma - \alpha. \tag{23}$$

The first of the modulation equations confirms *conservation of waves*, proving the existence of a *phase variable* $\theta(X,T)$, a potential from which $k(X,T)$ and $\omega(X,T)$ are derived by $k = \partial_X \theta$ and $\omega = -\partial_T \theta$. Conservation of waves is often imposed in an *ad hoc* manner to provide closure for modulation equations derived from a variational principle [2]. Here we see that it can be derived from an explicit limit of the local conservation laws.

IV. MODULATIONAL STABILITY AND ANALYSIS OF THE MODULATION EQUATIONS

A. Well posedness and the modulational stability criterion

The *characteristic velocities* of the modulation equations are the eigenvalues of the coefficient matrix $\mathbf{M}(k,\omega)$. These eigenvalues are

$$\lambda_{\pm}(k,\omega) = \frac{1}{D(k,\omega)}\{2[12\gamma \cos^2 k + 4(1+\gamma\omega)\cos k + 4\omega - 2\gamma - \alpha]\sin k \pm \sqrt{2(\omega + 2\cos k)p(k,\omega)}\}, \tag{24}$$

where

$$p(k,\omega) = a_2(k)\omega^2 + a_1(k)\omega + a_0(k) \tag{25}$$

and

$$a_2(k) = 24\gamma \cos^2 k + 12\cos k - 12\gamma, \tag{26}$$

$$a_1(k) = -16\gamma^2 \cos^4 k + 144\gamma \cos^3 k + (8 - 4\alpha\gamma + 16\gamma^2)\cos^2 k + (-8\alpha - 88\gamma)\cos k + 8 + 2\alpha\gamma + 4\gamma^2, \tag{27}$$

$$a_0(k) = -16\gamma^2 \cos^5 k + 160\gamma \cos^4 k + (16\gamma^2 - 16\alpha\gamma - 16)\cos^3 k + (-8\alpha - 112\gamma)\cos^2 k + (\alpha^2 + 8\alpha\gamma + 12\gamma^2 + 16)\cos k. \tag{28}$$

If the characteristic velocities are real and distinct for some k and ω, then the modulation equations are locally *hyperbolic*. This means that the initial value problem is locally well posed and is interpreted as implying the *modulational stability* of the plane-wave solution with wave number k and frequency ω. On the other hand, if the velocities have nonzero imaginary parts, then the modulation equations are locally *elliptic*. In this case, the initial value problem is locally ill posed since a typical initial condition for Eq. (19) will excite modes with arbitrarily large growth rates. This is interpreted as implying the *modulational instability* of the plane-wave solution. The modulational stability criterion is thus

$$(\omega + 2\cos k)p(k,\omega) > 0. \tag{29}$$

We wish to compare modulational stability with linear stability. In Figs. 4 and 5 the regions of modulational instability are indicated on the (k,ω) plane with vertical lines, while the regions of linear instability are shaded.

As can be seen in these figures, regions of stability can be isolated from each other both by the band gap and by regions of instability. The way to think of the modulation equations (19) and their relation to the modulational stability criterion is the following. If initial data $k(X)$ and $\omega(X)$ for the modulation equations (19) are parametrically represented in the (k,ω) plane as a curve that lies entirely outside the regions indicated with vertical lines, then the modulation equations describe the dynamics of these functions for slow times T in some finite interval $0 < T < T^*$. The time T^* cannot generally be taken to be infinite because it is possible either for the

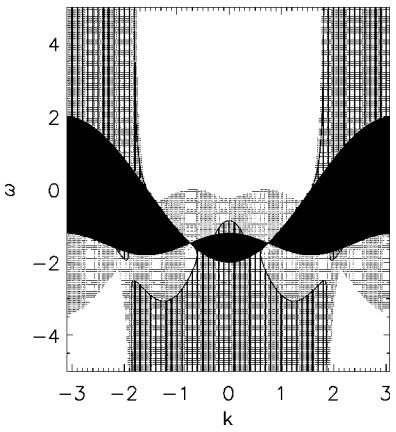

FIG. 4. Regions of instability in the (k,ω) plane for $\alpha=-3$ and $\gamma=-0.3$. The regions where the modulation equations are elliptic are shown with vertical lines and the regions where the plane waves are linearly unstable are shaded.

dynamics of Eq. (19) to drive the system to the boundary of the region of modulational stability at which point the initial value problem for Eq. (19) becomes locally ill posed (elliptic) or for X derivatives of k and ω to become infinite as the solution of Eq. (19) ''tries'' to become multivalued or develop a shock. In the latter case, the modulation equations may continue to have single-valued *weak* solutions (i.e., nonclassical or nonsmooth solutions in the sense of distributions) for times greater than T^*. These weak solutions propagate as shock fronts and require sophisticated numerical methods to resolve their structure [14]. However, the weak solutions *have no meaning* in the microscopic system (1) because they violate the assumption of the smoothness of $k(X)$ and $\omega(X)$ that allowed us to pass to the limit $\epsilon\downarrow 0$ and thus obtain a closed description of the macroscopic dynamics. When infinite derivatives appear in $k(X)$ and $\omega(X)$, one must return to the microscopic system (1) to explain the subsequent dynamics.

These statements concern only the modulation equations (19) and the behavior of their solutions. One hopes, of course, that smooth solutions of a globally well-posed problem in some macroscopic time interval $0<T<T^*$ actually describe the dynamics of a modulating plane wave of the microscopic system (1) in the limit of small ϵ (slow modulation). However, this is not always the case. Figures 4 and 5

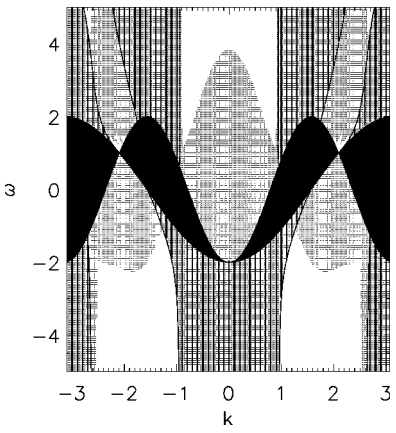

FIG. 5. Same as Fig. 4 but for $\alpha=0$ and $\gamma=2$.

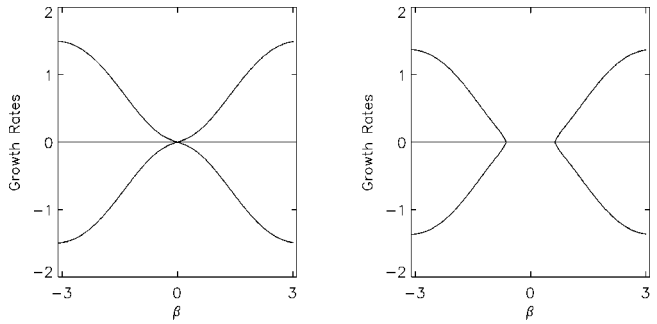

FIG. 6. Linear growth rates for $\alpha=-3$ and $\gamma=-0.3$ as a function of the perturbative wave number β. Left: $k=-1.5$ and $\omega=-3$ in the modulationally unstable region. Right: $k=-1.5$ and $\omega=-2.7$ just across the boundary in the modulationally stable but linearly unstable region.

clearly show that the hyperbolicity of modulation equations implies stability in a more restricted sense than is guaranteed by linear stability. A smooth solution of Eq. (19) will only be a good model for the dynamics of modulated plane waves in Eqs. (1) if the waves additionally avoid all regions of linear instability. In general, if a parametric representation of the modulating wave in the (k,ω) plane is approaching the boundary of the hyperbolic region so that the time of first change of type T^* is on the horizon well before any shocks have the chance to form, *it will first enter a region of linear instability*. This leads to the local breakdown of the wave form at a time T^{**} that is strictly less than the time for first change of type T^* predicted by the modulation equations alone. It is important to observe that the modulation equations (19) are *utterly unaware* of the ''minefield'' of instabilities that can lie between a region of linear stability and a region of modulational instability.

To clarify the difference between modulational stability and linear stability, it is useful to study the four linear growth rates calculated in the course of linear stability analysis as functions of the relative wave number β. By definition, a wave is linearly stable if all growth rates vanish identically as functions of $\beta\in[-\pi,\pi]$. So the interesting comparison is between the growth rates for waves in the two types of unstable regions. In Fig. 6 the linear growth rates are plotted as functions of β for two plane waves near the boundary of modulational stability.

These plots suggest that the hyperbolicity of the modulation equations (19) merely indicates that there exists a neighborhood of relatively long waves near $\beta=0$ that are stable perturbations of the plane wave. This should be contrasted with the more strict linear stability criterion that *all* values of β correspond to stable perturbations. Thus linear stability implies modulational stability, but the converse is not always true. In general, hyperbolicity of modulation equations is equivalent to linear stability *only if all linear instabilities are of Benjamin-Feir type*, that is, if all unstable modes lie in a sideband of the underlying wave train. That the system (1) so clearly demonstrates this fact is one of the main messages of our paper.

Although a proof of this statement in all generality remains to be given, the relationship between linear instability and ellipticity of modulation equations can be demonstrated exactly in this particular problem. If one considers the limit

$\beta \downarrow 0$, the characteristic polynomial with eigenvalue $i\sigma$ of the matrix in Eq. (12) takes the form

$$\sigma^4+[\beta b_3+O(\beta^2)]\sigma^3+[b_2+O(\beta)]\sigma^2+[\beta b_1+O(\beta^2)]\sigma +[\beta^2 b_0+O(\beta^3)]=0, \tag{30}$$

where $b_i=b_i(k,\omega)$ are real coefficients that can be explicitly calculated by direct asymptotic expansion of the characteristic polynomial for small β. Two of the roots are $O(1)$ in the limit and to leading order satisfy

$$\sigma^2+b_2=0. \tag{31}$$

Since b_2 can be shown to be strictly negative for all k and ω, these roots never imply instability. The other two roots are obtained by setting $\sigma=\beta\tilde{\sigma}$, where to leading order

$$b_2\tilde{\sigma}^2+b_1\tilde{\sigma}+b_0=0. \tag{32}$$

These roots will be real and the wave parametrized by the pair (k,ω) will thus be stable to its sideband if the discriminant is positive:

$$b_1^2-4b_2b_0>0. \tag{33}$$

It can be shown that this condition is equivalent to Eq. (29) for all pairs (k,ω) that lie outside the band gap.

The question of the meaning of the modulation equations in the intermediate layer where they are hyperbolic but where short-wave linear instabilities exist is a deep one requiring more investigation. Are the instabilities seen in the numerical experiments artifacts of not working sufficiently close to the limit $\epsilon \downarrow 0$ of vanishing scale ratios (or indeed of numerical noise due to integration and roundoff errors)? That is, in the limit $\epsilon \downarrow 0$, are any short-wave instabilities excited by the modulation itself to destroy the leading order behavior? Or do genuine limiting modulated wave solutions exist that are linearly unstable to short waves?

B. Dynamics of modulated waves

With the caveat that we restrict attention to modulated waves that at $T=0$ lie for all X in a region of linear stability, the modulation equations (19) accurately describe the evolution of the waves for finite times $T<T^{**}$. Let us explore some of the implications of the description of the dynamics offered by the hyperbolic system (19).

1. A brief remark

There are clearly no dispersive terms in the modulation equations (19). These terms have been neglected because (as will be clearly demonstrated below) they play no role in the dynamics until shocks form or until the characteristic velocities $\lambda_{\pm}(k,\omega)$ degenerate and become identical on the boundary of the region of hyperbolicity. In the former case, the weak dispersion necessary to regularize the behavior is supplied by the underlying discreteness of the system and leads (at least at first) to Korteweg–de Vries–like dynamics at the shock front, which sheds solitons. In fact, the numerical experiments of Zabusky and Kruskal that resulted in the coining of the word ''soliton'' [15] clearly demonstrated the phenomenon of solitons escaping from a steepening shock, and in the presence of integrability of the microscopic system (a luxury we do not have available here) the details of how weak dispersion regularizes the shock can be worked out exactly [16,17]. In the latter case of nearly degenerate characteristic velocities, a Galilean transformation can be used to move in the frame of the common characteristic velocity so that the first-order X derivatives disappear altogether and the leading-order balance appears on a slower time scale yielding a nonlinear Schrödinger equation and envelope solitons. Envelope solitons occur, for example, in the limit of small-amplitude waves where $V\approx 0$ and thus $\lambda_+\approx\lambda_-$. However, this case is not always as relevant physically as the finite amplitude case because, as pointed out in the Introduction, the model (1) is typically derived with an assumption of small amplitude (or what is often equivalent, weak nonlinearity) so that $O(1)$ values of the dimensionless amplitudes w_n and v_n already correspond to small physical displacements. Considering even smaller displacements is then a further restriction that removes the dominant nondispersive terms from the problem. In optics, dispersionless terms such as those appearing in the modulation equations (19) characterize group velocity mismatch and the so-called walk-off effect of beams and pulses. Since we keep only these terms, the dynamics that we will obtain can be interpreted as ''nonlinear walk-off'' because the two distinct group velocities $\lambda_{\pm}(k,\omega)$ depend on the local field values. Of course, we do not mean to imply that dispersion can be neglected for *all* solutions of Eqs. (1), only that this is so for those solutions that have the form of fully nonlinear modulated plane waves. For example, isolated localized solutions such as self-trapped stationary excitations and breathers often exist in discrete systems such as (1) and thrive on a dynamical balance between nonlinear effects and dispersion. However, the fact that such excitations are localized in the lattice when their amplitudes are $O(1)$ means that they are not individually well described by the dispersionless modulation equations (19), which are valid as a model for disturbances that are spatially extended. In short, dispersion is important for microscopic dynamics but not for smooth macroscopic dynamics.

2. Validity of the modulation equations

The fact that the modulation equations (19) are the correct model for the macroscopic dynamics of modulated waves is best seen by comparing numerical solutions of Eq. (19) with the data obtained from numerical simulations of the microscopic system (1) shown in Figs. 2 and 3. On the left-hand side of Fig. 7, we show the snapshots of Fig. 2 plotted parametrically in the (k,ω) plane [we take $k=\arg(w_{n+1}/w_n)$ and then determine ω from the dispersion relations using the local amplitudes], while on the right-hand side we show plots of the corresponding snapshots obtained by numerically integrating the modulation equations (19) directly using a stable scheme of centered differences in space and ''leap-frogging'' in time. The first few curves from each simulation are indistinguishable to the eye, illustrating the accuracy of the description offered by the macroscopic system (19). We then see that the dynamics has driven the system locally into a region of linear instability, at which point the actual modulating wave solution of Eqs. (1) breaks up (see Fig. 2). Since

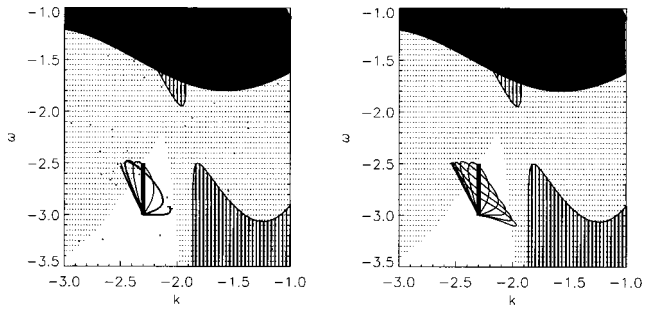

FIG. 7. Left: the snapshots of Fig. 2 plotted parametrically in the (k,ω) plane. Right: corresponding snapshots obtained by direct numerical simulation of the modulation equations. There is excellent agreement until the dynamics move the wave into a region of linear instability. Since these are short-wave instabilities, the modulation equations remain hyperbolic and can be integrated beyond the local destruction of the modulated plane wave in the microscopic system.

this is a region of short-wave instabilities, the modulation equations remain hyperbolic and propagation continues as if nothing were wrong.

Similar parametric plots comparing the numerical data from the microscopic system (1) that was first shown in Fig. 3 to a numerical simulation of the macroscopic equations (19) are shown in Fig. 8. Here again one sees that the modulation equations are an excellent model for the dynamics. We also see that the shock wave that forms on the right-hand side of Fig. 3 is different from the breakdown that occurred in the previous simulations, as it is not related to any instabilities; the evolution remains outside of all unstable regions. Other important features to observe are the fact that the curve seems to trace out a four-sided caustic as it evolves and that the curve rapidly contracts to the two sides of this figure that are joined at the vertex corresponding to the background plane wave in Fig. 3 on which the disturbances are propagating. More careful observations reveal that each of the two waves that seem to have separated in Fig. 3 is confined to just one of the sides of this figure.

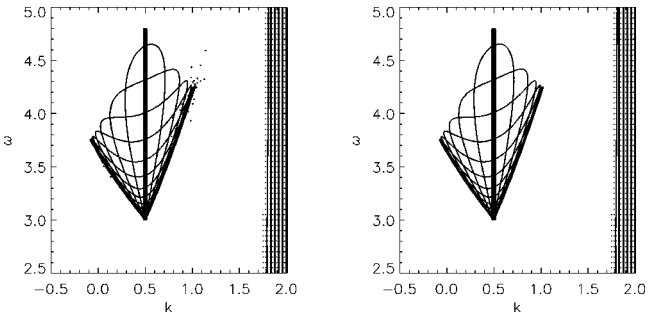

FIG. 8. Left: snapshots of Fig. 3 plotted parametrically in the (k,ω) plane. Right: Corresponding snapshots obtained by direct numerical simulation of the modulation equations. The waves ultimately lie along two sides of a distorted rectangle joined at the point of the background wave. These two sides correspond to independently propagating simple waves and the distorted rectangle is the image of a genuine rectangle in the plane of the Riemann invariants. The scattered points in the upper right corner of the left figure are the oscillations near the shock front; they do not appear in the right figure because the modulation equations cannot be integrated sensibly beyond the shock time.

3. Existence of Riemann invariants

In explaining these observations, the most important tool for interpreting the modulation equations (19) is their representation in terms of *Riemann invariants*. Riemann invariants for a hyperbolic quasilinear system are new dependent variables for which the matrix of coefficients of X derivatives is the diagonal matrix of characteristic velocities. In the case of the equations (19), these are variables $r_\pm(k,\omega)$ in which the modulation equations take the form

$$\partial_T r_\pm + c_\pm(r_+,r_-)\partial_X r_\pm = 0. \tag{34}$$

Thus the equations are coupled only through the characteristic velocities $c_\pm(r_+,r_-)=\lambda_\pm(k,\omega)$, now expressed in terms of $r_\pm$ rather than k and ω. It is also easy to visualize the dynamics implied by these equations: Each point on the graph of $r_\pm(X,T)$ as a function of X is moving to the right with speed $c_\pm(r_+(X,T),r_-(X,T))$. Finding the Riemann invariants amounts to solving the Pfaffian differential equation

$$\mathbf{D}\begin{bmatrix} dr_+ \\ dr_- \end{bmatrix} = \begin{bmatrix} s_1^+ & s_1^- \\ s_2^+ & s_2^- \end{bmatrix}^{-1}\begin{bmatrix} dk \\ d\omega \end{bmatrix}, \tag{35}$$

where $s_j^\pm(k,\omega)$ are the components of linearly independent eigenvectors of the coefficient matrix $\mathbf{M}(k,\omega)$ belonging to the distinct eigenvalues $\lambda_\pm(k,\omega)$ and $\mathbf{D}$ is an arbitrary invertible diagonal matrix. Rewritten in terms of the Jacobian matrix of the transformation, this becomes

$$\frac{\partial(r_+,r_-)}{\partial(k,\omega)}\mathbf{S} = \mathbf{D}^{-1}, \tag{36}$$

where $\mathbf{S}$ is the eigenvector matrix. Since the only constraint is that the left-hand side be diagonal, this amounts to N^2-N equations in N unknowns for an $N\times N$ hyperbolic system. The Riemann invariants are thus generally overdetermined, although there are special cases where there is enough structure for solutions to exist in the $N\times N$ case nonetheless; see, e.g., [18]. However, for two-component systems such as (19), there are exactly as many equations as unknowns and thus there exist solutions [14], although they may be very difficult to write down explicitly. Using the fact that $M_{11}(k,\omega)=0$ and $M_{12}(k,\omega)=1$, it is easy to see that the Riemann invariants satisfy

$$\partial_k r_\pm(k,\omega) + \lambda_\mp(k,\omega)\partial_\omega r_\pm(k,\omega) = 0. \tag{37}$$

It follows that $r_\pm(k,\omega)$ is constant along curves $\omega=\omega(k)$ in the (k,ω) plane where

$$\omega'(k) = \lambda_\mp(k,\omega(k)). \tag{38}$$

Although at this time we do not have explicit expressions for $r_\pm(k,\omega)$, the fact that Riemann invariants exist for this problem allows us to make several key observations about the behavior of modulated waves in the microscopic system (1).

4. Riemann signatures, signature boxes, and instabilities

We recall some useful terminology introduced in [3]. In terms of the Riemann invariants, given fields $k(X)$ and $\omega(X)$ can be represented parametrically in the (r_+,r_-) plane by a

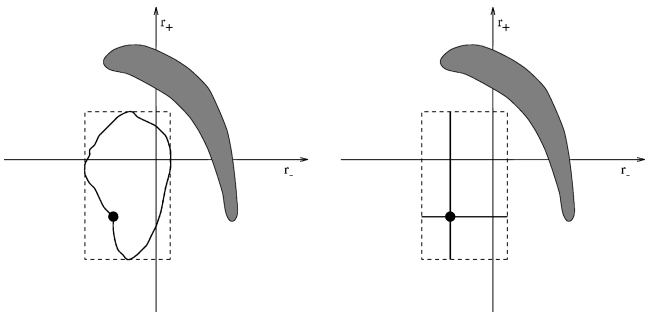

FIG. 9. Left: schematic illustration of the Riemann signature of initial data $r_{\pm}(X)$ homoclinic to the point $(r_+^{\infty},r_-^{\infty})$ as $|X|\to\infty$. The homoclinic point is shown with a large dot. The time evolution of these data will avoid the shaded linearly unstable regions because its dashed signature box does, and the signature cannot leave the box. Right: signature of the data in the limit of large time T if the criterion for resolution into simple waves is satisfied. Although the signature becomes fixed, nontrivial hyperbolic dynamics are taking place on the two line segments $r_{\pm}=r_{\pm}^{\infty}$ in the signature box.

graph that we call the *Riemann signature* of the data. Each Riemann signature is contained in a unique *signature box*, which is the smallest rectangle with edges having constant values of either r_+ or r_- that contains the signature. An easy consequence of writing the hyperbolic system of modulation equations in Riemann invariant form is that, while the Riemann signature of the fields evolves in time, the signature box remains the same.

The fact that the Riemann signature can never leave the box that initially contained it allows us to see that there exist some modulated waves that never enter the region of linear instability until shocks form in the macroscopic fields. In general, initial data $k(X)$ and $\omega(X)$ contained completely in the linearly stable region might evolve under the modulation equations into a region of instability, as in Fig. 7. However, observe that if not just the Riemann signature but also the signature box of the data lies completely in the region of linear stability, then this situation is avoided for as long as the solutions of the modulation equations remain smooth (see Fig. 9, left). This latter case is exactly what is going on in Fig. 8. In fact, the four-sided figure that is being traced out by the dynamics of the curve is nothing but the image of the rectilinear signature box under the mapping $(r_+,r_-)\mapsto(k,\omega)$. If we had formulas for the Riemann invariants $r_{\pm}(k,\omega)$, we could use these observations to establish sufficient conditions on initial modulated wave data $k(X)$ and $\omega(X)$ to avoid spontaneous linear instabilities of the kind shown in Figs. 2 and 7.

5. Simple wave solutions and asymptotic resolution into simple waves

Another easy consequence of writing the modulation equations in Riemann invariant form (34) is that the system admits the reduction of taking either $r_+(X,T)$ or $r_-(X,T)$ to be constant. For example, one can set $r_-(X,T)=K$ so that the system reduces to the standard form of the Hopf or inviscid Burgers equation

$$\partial_T u+u\partial_X u=0 \tag{39}$$

in the single-dependent variable $u=c_+(r_+,K)$, which is easily solved by the method of characteristics [2]. Such special solutions of the modulation equations, where only one of the Riemann invariants is nonconstant, are called *simple waves*. The simple waves are the nonlinear analogs of the right- and left-going components in d'Alembert's solution $u(X,T)=f(X-cT)+g(X+cT)$ of the linear wave equation $\partial_T^2 u-c^2\partial_X^2 u=0$.

Of course, one of the most striking properties of d'Alembert's solution is that every localized initial disturbance eventually resolves itself into isolated right- and left-going waves. This is not always true in the nonlinear system (34), but let us see what we can conclude in this case nonetheless. Assume for the moment that we have expressions for $r_{\pm}(k,\omega)$ and that we have found initial data $r_{\pm}(X)$, such as that from Figs. 3 and 8, for which we do not have to worry about instabilities. Assume also that these data represent a localized disturbance on a background wave, so that as $|X|\to\infty$, the Riemann invariants take on constant values: $r_{\pm}(X)\to r_{\pm}^{\infty}$. The Riemann signature of these kinds of initial data is shown schematically on the left-hand side in Fig. 9. Now, without loss of generality we can assume that $c_+(r_+,r_-)>c_-(r_+,r_-)$ at each point in the strictly hyperbolic region of the (r_+,r_-) plane. According to Eq. (34), points on the graph of $r_+(X)$ will move to the right with a speed bounded below by

$$c_+^{\inf}\doteq \inf_{-\infty<X<\infty,T>0} c_+(r_+(X,T),r_-(X,T)), \tag{40}$$

while points on the graph of $r_-(X)$ will move to the right with a speed bounded above by

$$c_-^{\sup}\doteq \sup_{-\infty<X<\infty,T>0} c_-(r_+(X,T),r_-(X,T)). \tag{41}$$

This means that if $c_+^{\inf}$ is strictly greater than $c_-^{\sup}$, then the region in which the field $r_+(X)$ differs significantly from r_+^{∞} will eventually overtake the region in which the field $r_-(X)$ differs significantly from r_-^{∞} and the coupled problem will have resolved itself into two uncoupled simple wave problems. We would like to be able to characterize the initial data for which this simplification will ultimately occur. Unfortunately, the extreme speeds defined by Eqs. (40) and (41) cannot be evaluated directly without integrating the coupled system (34). However, it is easy to bound these quantities because we know that the signature box contains the dynamics for all X and T. If B is the region bounded by the signature box in the (r_+,r_-) plane (B is determined from the initial data alone), then we have

$$d_+^{\inf}\doteq \inf_B c_+(r_+,r_-)\leqslant c_+^{\inf} \tag{42}$$

and

$$d_-^{\sup}\doteq \sup_B c_-(r_+,r_-)\geqslant c_-^{\sup}. \tag{43}$$

These quantities are easy to evaluate given the Riemann invariants and one now can say that if the initial data $r_{\pm}(X)$ has the signature box B, then the condition

$$d_+^{\inf} > d_-^{\sup} \tag{44}$$

is a sufficient condition for the dynamics to naturally separate into two simple waves. The signature will ultimately contract onto the two line segments $r_\pm = r_\pm^\infty$ in the signature box, as shown on the right-hand side in Fig. 9. The dynamics on each of these segments is given by independent Hopf equations (39) for the variables $u_+(X,T) = c_+(r_+(X,T), r_-^\infty)$ and $u_-(X,T) = c_-(r_+^\infty, r_-(X,T))$. The simulation shown in Figs. 3 and 8 is one in which an initial condition is resolved into simple waves. In this case, the homoclinic point (r_+^∞, r_-^∞) lies at one of the vertices of the signature box B. For this simulation, the condition (44) can be shown to be satisfied and the dramatic effect of the separation of the two simple waves can be seen in Fig. 3. Indeed, the phenomenon of resolution of an initial condition into simple waves appears to be quite common in the particular set of modulation equations (19).

The arguments above presume that shocks do not form before the separation into simple waves is complete. If $r_+(X)$ is initially different from r_+^∞ only for $X > X_+$ and $r_-(X)$ is initially different from r_-^∞ only for $X < X_-$, then the separation time is

$$T_{\rm sep} = \frac{X_+ - X_-}{c_+^{\inf} - c_-^{\sup}} \leqslant \frac{X_+ - X_-}{d_+^{\inf} - d_-^{\sup}}. \tag{45}$$

If shocks do not form before $T = T_{\rm sep}$, then the time for shock formation can be calculated exactly [2] from the independent equations for the two simple waves.

V. IMPORTANT LIMITING CASES

With its two free ''material'' parameters α and γ, the discrete quadratic chain (1) is a very rich system that includes other known models as limiting cases. Also, for fixed values of the material parameters α and γ, the plane-wave parameters k and ω can be scaled so that the dynamics of modulated plane waves simplifies considerably.

A. The continuum limit

Introduce a small parameter $\epsilon \ll 1$ and slow scales $X = \epsilon n$ and $\tau = \epsilon^2 t$ and make the following transformations of parameters in Eqs. (1):

$$\alpha = 2\tilde{\gamma} - 4 + \epsilon^2 \tilde{\alpha}, \quad \gamma = \tilde{\gamma}. \tag{46}$$

Then set

$$w_n(t) = \epsilon^2 \tilde{w}(X,\tau) e^{2i\tau/\epsilon^2}, \quad v_n(t) = \epsilon^2 \tilde{v}(X,\tau) e^{4i\tau/\epsilon^2} \tag{47}$$

and assume that $\tilde{w}$ and $\tilde{v}$ are smooth functions of X and τ to obtain the system

$$i\partial_\tau \tilde{w} + \partial_X^2 \tilde{w} + \tilde{w}^* \tilde{v} = 0,$$

$$i\partial_\tau \tilde{v} + \tilde{\gamma}\partial_X^2 \tilde{v} - \tilde{\alpha}\tilde{v} + \frac{1}{2}\tilde{w}^2 = 0. \tag{48}$$

This is a new family of systems, with material parameters $\tilde{\alpha}$ and $\tilde{\gamma}$, that includes the model for two-wave solitons due to cascading in quadratic nonlinear optical media or so-called $\chi^{(2)}$ materials [10]. For a general review of cascading in optics see [19]. Note that as mentioned above, the continuum limit leading from the discrete system (1) to (48) is not always very relevant *if the discrete system already represents the dominant weakly nonlinear physics*. As we will now see, the limit is, however, generally useful as a mathematical device in the study of the continuum system (48) when it arises on its own, more or less from first principles.

Nonlinear plane waves of the discrete system (1) go over to nonlinear plane waves of the continuum system (48) in the limit $\epsilon \downarrow 0$. The latter waves have the form

$$\tilde{w}(X,\tau) = \tilde{W} e^{i(\tilde{k}X - \tilde{\omega}\tau)}, \quad \tilde{v}(X,\tau) = \tilde{V} e^{2i(\tilde{k}X - \tilde{\omega}\tau)}. \tag{49}$$

The relations between the continuum and discrete plane-wave parameters are

$$W = \epsilon^2 \tilde{W}, \quad V = \epsilon^2 \tilde{V},$$

$$k = \epsilon \tilde{k}, \quad \omega = -2 + \epsilon^2 \tilde{\omega} \tag{50}$$

and the corresponding nonlinear dispersion relations for Eqs. (48) are obtained by substituting these into the discrete dispersion relations (6) and taking the limit $\epsilon \downarrow 0$.

These facts indicate that the linear stability results for the continuum system (48) can be obtained from the family of discrete systems (1) satisfying $\alpha \approx 2\gamma - 4$ by examining the stability of the waves in the vicinity of $\omega = -2$ and $k = 0$. The (k,ω) plane of one such system is shown in Fig. 5. By considering a perturbative wave number $\beta = \epsilon\tilde{\beta}$ and then using the ϵ dependence to ''unfold'' the degenerate eigenvalues (for $\epsilon = 0$) of the matrix of the linearized problem for the discrete system, the stability properties of *all waves* in the continuum limit system (48) are determined from the stability of an arbitrarily small neighborhood of a *single wave* of the discrete system (1). Equivalently, by direct linearization about plane-wave solutions of the continuum problem (48), it has been shown [11] that all plane waves with $\tilde{k} = 0$ are unstable and if $\tilde{\gamma} = 1/2$ then the system is Galilean invariant, so that instability for $\tilde{k} = 0$ implies instability for all $\tilde{k}$. Experimental results [12] suggest that all plane waves may be unstable for all values of $\tilde{\gamma}$.

Although the boundaries between regions of stability and instability may disappear in the continuum limit, with all waves becoming unstable, the distinction between modulational instability and linear stability does not disappear. For example, taking $\tilde{\alpha} = 0$, $\tilde{\gamma} = 2$, and $\tilde{k} = 0$, one can see that for $\tilde{\omega}$ negative (below the band gap) there are always instabilities in a sideband of $\tilde{\beta} = 0$, while for $\tilde{\omega}$ positive (above the band gap) the longest unstable mode appears for finite nonzero $\tilde{\beta}$. We suspect that this is true more generally and that below the band gap the linear instabilities of plane waves in the continuum system are of Benjamin-Feir type, while above the band gap they are not. The continuum version of the modulation equations (19) are hyperbolic above the band gap, but are elliptic below the band gap. The coexistence of

both stable and unstable waves in the system (1) is a direct consequence of its discreteness, but clearly the existence of short-wave instabilities without sideband instabilities is not a feature of discrete systems alone.

B. The discrete NLS limit and the high-frequency limit

Another system that can be obtained from Eqs. (1) by a passage to a limit in the material parameters is the discrete nonlinear Schrödinger equation. Make the scaling

$$w_n(t)=\sqrt{|\alpha|}\widetilde{w}_n(t) \tag{51}$$

and consider the limit of large detuning $|\alpha|\uparrow\infty$, holding v_n and $\widetilde{w}_n$ fixed. This leads to the system

$$i\partial_t\widetilde{w}_n+\widetilde{w}_{n+1}+\widetilde{w}_{n-1}+\frac{\eta}{2}|\widetilde{w}_n|^2\widetilde{w}_n=0, \tag{52}$$

where $\eta=\mathrm{sgn}(\alpha)$. The weak second harmonic is slaved to the strong fundamental by $v_n=\eta\widetilde{w}_n^2/2$. This model is a discrete version of the nonlinear Schrödinger equation. Its stationary solutions and their linear stability properties are known [20] and the modulational behavior of its plane-wave solutions was studied in [3].

Let us deduce the linear growth rates for plane waves in the system (52) from the characteristic polynomial of the matrix of the linearized problem (12) by taking the limit of large α and keeping all other parameters fixed. Calling the eigenvalue $i\sigma$, we find in the limit that two of the roots are given by $\sigma=\pm\alpha+O(1)$. The remaining two roots σ are $O(1)$ and are real if the discriminant is positive:

$$[(3-4\cos\beta+\cos^2\beta)\cos k+\omega(1-\cos\beta)]\cos k>0. \tag{53}$$

The large $|\alpha|$ behavior of the dispersion relations (6) gives $\omega=-2\cos k-\eta\widetilde{W}^2/2$, where $\widetilde{W}^2=|\widetilde{w}_n|^2$, so that $\alpha(\omega+2\cos k)$ is negative. From this it follows that if $\alpha\cos k<0$ the discriminant is positive for all β, while for $\alpha\cos k>0$ the discriminant is strictly negative for all β in the left and right sidebands of $\beta=0$ defined by

$$1-\frac{\eta\widetilde{W}^2}{2\cos k}<\cos\beta<1. \tag{54}$$

Therefore, if linear instabilities exist at all, then they certainly exist for a sideband of $\beta=0$. This asymptotic analysis verifies what was shown in [3], that in the discrete nonlinear Schrödinger equation linear stability in general is equivalent to linear stability to perturbations of relative wave number $\beta\approx0$ and thus to modulational stability as determined from the hyperbolicity of a pair of modulation equations. The discreteness of a system does not imply a difference between the two kinds of stability.

We also note that the difference between instability of the neighborhood $\beta\approx0$ and instability in general also vanishes from the system (1) for fixed arbitrary values of the material parameters α and γ in the limit of high frequency ω. Taking the limit $|\omega|\uparrow\infty$, the modulational stability criterion (29) becomes simply

$$\mathrm{sgn}(\omega)[2\gamma\cos^2 k+\cos k-\gamma]>0. \tag{55}$$

On the other hand, two of the roots $i\sigma$ of the characteristic polynomial of the matrix in the linearized system (12) are of the form $\sigma=\pm\omega\sqrt{12}+O(1)$, while the other two are of the form $\sigma=\widetilde{\sigma}\sqrt{|\omega|}$, where to leading order

$$\widetilde{\sigma}^2=2\,\mathrm{sgn}(\omega)[1-\cos\beta][2\gamma\cos^2 k+\cos k-\gamma]. \tag{56}$$

Since $1-\cos\beta$ is non-negative for all β, the reality of these roots is equivalent to Eq. (55).

VI. CONCLUSION

The quadratic nonlinear two-component lattice model (1) has a family of plane-wave solutions, some of which are linearly stable and some of which are not. For linearly stable waves, spatiotemporal modulations of the wave parameters k and ω that are slowly varying but have finite amplitude satisfy a dispersionless nonlinear hyperbolic system of macroscopic modulation equations (19). The modulation equations can be hyperbolic even in regions of the (k,ω) plane where the plane waves are linearly unstable. In this case, the linear instabilities are not Benjamin-Feir instabilities. The wave is stable to a sideband of relatively long waves and the unstable perturbation with the longest wavelength has a strictly nonzero relative wave number β. To our knowledge, this is the first concrete example of a nonlinear system with modulated waves that are susceptible to such instabilities.

Of course, there are other models with similar short-wave instabilities. For instance, all plane waves in the continuum equations (48) that lie above the band gap are unstable to some modes, but the sideband neighborhood of $\beta=0$ consists only of stable modes. By contrast, the waves below the band gap are also unstable, but always suffer from the long-wave Benjamin-Feir instability. This means that having the longest wavelength unstable perturbation with nonzero wave number (so that modulation equations are hyperbolic in spite of linear instability) is not a phenomenon restricted to discrete systems alone. Nor is the phenomenon implied by discreteness. The discrete nonlinear Schrödinger equation (52) is a system that is contained in the general model (1) in the limit of large detuning α and for which hyperbolicity of modulation equations always goes hand in hand with linear stability.

The phenomenon might be related to multicomponent systems, as this is the obvious common feature of the discrete system (1) and the continuum system (48) that is not shared by the discrete nonlinear Schrödinger equation (52) or other familiar models where Benjamin-Feir instabilities are the rule. Having contributed this observation, we view the classification of wave equations for which all linear instabilities of a wave are restricted to its sideband as an open problem.

Another question raised by our analysis of the plane waves in the system (1) is that of the meaning of the modulation equations in the twilight region where, in the presence of linear instabilities, they are nonetheless hyperbolic. The most challenging form of this question asks whether there is a dispersionless macroscopic limit in such a region at all.

Similarly, the problem of establishing whether there exists a macroscopic description of plane waves in the focusing nonlinear Schrödinger equation (where the modulation equations are always elliptic) remains open despite the integrability of that system (but see [21] for recent progress). There are fewer tools available to study the system (1) since it is not integrable, but the problem is more interesting because, as we have seen in this paper, there is good evidence that the macroscopic limit may not exist even in some regions where the modulation equations are hyperbolic.

When instabilities of all types can be avoided at least initially, the modulation equations (19) are evidently a good model for the time evolution of a smoothly modulated wave train over some finite time interval. The length of this interval is limited only by the possibility that the wave might encounter instabilities during modulation or that the wave might break. Because we know that Riemann invariants exist for the modulation equations, we know that we can in principle identify spatial modulations that will never encounter any instabilities as they modulate. In some cases, which can also be characterized in terms of the Riemann invariants, the modulated wave will separate into two isolated pulses, the evolution of each of which can be found from the Hopf equation for simple waves by the method of characteristics. These simple wave solutions will be valid until they form infinite derivatives, at which point dispersion introduced by the discreteness of the microscopic system must be taken into account. Obviously, the predictive capacity of the theory of the hyperbolic system (19) would be strengthened if we had explicit expressions for the Riemann invariants $r_{\pm}(k,\omega)$. Analysis of the ordinary differential equations (38) for the contours of constant $r_{\pm}(k,\omega)$ leading to solutions of the corresponding partial differential equations (37) is clearly an avenue for future work.

When discreteness makes itself known at the front of a forming shock wave, one expects microscopic oscillations with wave numbers possibly close to the edge of the Brillouin zone to appear in the formerly smooth functions $k(X)$ and $\omega(X)$. If these oscillations are regular, in the sense that they have envelopes and mean values that appear to be smooth functions of X (this is at first a matter of interpreting numerical experiments), then it is possible that moments constructed from local averages such as $\langle k^p \rangle$ and $\langle \omega^p \rangle$ could be found to satisfy a larger quasilinear system of equations in X and T and the analysis could thus be continued beyond the shock formation time. As has been shown several times [16,17], this is indeed the case if the microscopic system is integrable. To our knowledge, there is not yet an example of a nonintegrable microscopic system that has enough local conservation laws to allow wave modulation to be continued through first shock formation in some weak sense.

ACKNOWLEDGMENTS

We would like to thank Gennady El', David Levermore, and Alan Newell for useful discussions.

[1] J. C. Luke, Proc. R. Soc. London, Ser. A **292**, 403 (1966).
[2] G. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[3] M. H. Hays, C. D. Levermore, and P. D. Miller, Physica D **79**, 1 (1994); **85**, 304(E) (1995).
[4] C. D. Levermore and J.-G. Liu, Physica D **99**, 191 (1997).
[5] G. A. El', A. V. Gurevich, V. V. Khodorovskiĭ, and A. L. Krylov, Phys. Lett. A **177**, 357 (1993); A. M. Kamchatnov, Phys. Rep. **286**, 199 (1997).
[6] V. M. Agranovich and A. M. Kamchatnov, JETP Lett. **59**, 424 (1994); V. M. Agranovich, S. A. Darmanyan, O. A. Dubovsky, A. M. Kamchatnov, Th. Neidlinger, and P. Reineker, Phys. Rev. B **53**, 15 451 (1996).
[7] O. A. Dubovskiĭ and A. V. Orlov, Fiz. Tverd. Tela **38**, 1221 (1996) [Phys. Solid State **38**, 675 (1996)]; **38**, 1931 (1996) **[38**, 1067 (1996)].
[8] T. Peschel, U. Peschel, and F. Lederer, Phys. Rev. E **57**, 1127 (1998).
[9] O. Bang, P. L. Christiansen, and C. Balslev Clausen, Phys. Rev. E **56**, 7257 (1997).
[10] Yu. N. Karamzin and A. P. Sukhorukov, JETP Lett. **20**, 339 (1974); C. R. Menyuk, R. Schiek, and L. Torner, J. Opt. Soc. Am. B **11**, 2434 (1994).
[11] A. V. Buryak and Y. S. Kivshar, Phys. Lett. A **197**, 407 (1995); H. He, P. D. Drummond, and B. A. Malomed, Opt. Commun. **123**, 394 (1996).
[12] R. A. Fuerst, D. M. Baboiu, B. Lawrence, W. E. Torruellas, G. I. Stegeman, S. Trillo, and S. Wabnitz, Phys. Rev. Lett. **78**, 2756 (1997).
[13] V. V. Konotop and M. Salerno, Phys. Rev. E **56**, 3611 (1997).
[14] P. D. Lax, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves* (SIAM, Philadelphia, 1973).
[15] N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965).
[16] P. D. Lax and C. D. Levermore, Commun. Pure Appl. Math. **36**, 253 (1983); **36**, 571 (1983); **36**, 809 (1983).
[17] A. M. Bloch and Y. Kodama, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **52**, 909 (1992).
[18] H. Flaschka, M. G. Forest, and D. W. McLaughlin, Commun. Pure Appl. Math. **33**, 739 (1980).
[19] G. I. Stegeman, D. J. Hagan, and L. Torner, Opt. Quantum Electron. **28**, 1691 (1996).
[20] J. Carr and J. C. Eilbeck, Phys. Lett. **109A**, 201 (1985); B. M. Herbst, A. R. Mitchell, and J. A. C. Weideman, J. Comput. Phys. **60**, 263 (1985); J. A. C. Weidemann and B. M. Herbst, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **23**, 485 (1986).
[21] J. C. Bronski, Physica D **97**, 376 (1996); J. C. Bronski and D. W. McLaughlin, in *Singular Limits of Dispersive Waves*, edited by N. M. Ercolani *et al.* (Plenum, New York, 1994), p. 21; Shan Jin, C. D. Levermore, and D. W. McLaughlin, in *Singular Limits of Dispersive Waves* (*ibid.*), p. 235.