ELSEVIER

Physica D 92 (1996) 1–27

PHYSICA D

Modulation of multiphase waves in the presence of resonance

Peter D. Miller [a,1], Nicholas M. Ercolani [b,2], C. David Levermore [b,3]

[a] *Optical Sciences Centre, Australian National University, Canberra ACT 0200, Australia*

[b] *Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA*

Received 13 December 1994; revised 10 July 1995; accepted 19 October 1995

Communicated by H. Flaschka

Abstract

The phenomenon of spatio-temporal phase modulation made possible by resonance is investigated in detail through the analysis of an example problem. A simple family of exact solutions to the Ablowitz-Ladik equations is found to be modulationally stable in some regimes. This family of solutions is determined by fixing antiperiod 2 boundary conditions, which determines two wavenumbers. Within the family of solutions, the frequencies do not depend on amplitude; this feature ensures that the antiperiod 2 boundary conditions will be enforced under modulation. The family of solutions is described by four parameters, two being actions that foliate the phase space, and two being macroscopically observable functions of the phase constants. The modulation of the actions is described by a closed hyperbolic system of first order equations, which is consistent with the full set of four genus 1 modulation equations. The modulation of the phase information, easily observed due to the presence of two resonances, is described by two more equations that are driven by the actions. The results are confirmed by numerical experiments.

1. Introduction

In the study of linear dispersive waves, one often asks about the behavior of waves in a slowly varying medium. For example, in optical problems with a slowly varying index of refraction, one uses the ideas of geometrical optics to describe the slow variations in the wavefront geometry and amplitude of a rapidly oscillating electromagnetic field. In extending these ideas to nonlinear problems, two new effects appear. First, the nonlinearity may introduce slow variation in the waves even in a homogeneous medium (as in the Kerr media of nonlinear optics, where the refractive index depends on the local field intensity). Second, the fundamental wave modes may depend periodically on more than one independent phase variable of the form $kx - \omega t$. In *integrable* nonlinear dispersive wave systems, the multiphase solutions constructed in terms of the algebraic

[1] Program in Applied Mathematics, University of Arizona, Tucson, AZ 85721. E-mail: pdm124@rsphysse.anu.edu.au.

[2] E-mail: ercolani@math.arizona.edu.

[3] E-mail: lvrmr@math.arizona.edu.

SSDI 0167-2789(95)00281-2

geometry of Riemann surfaces [4] are the fundamental oscillatory modes that are the analogs of the Fourier modes of linear dispersive wave theory. Like their counterparts in the linear theory, these modes may be spatially and temporally modulated. Describing the modulations requires a nonlinear generalization of the averaging methods of geometrical optics.

The extension of these averaging methods to multiphase solutions of integrable nonlinear wave problems was accomplished by Flaschka, Forest, and McLaughlin [9] using the machinery of the inverse spectral transform for the Korteweg-deVries equation. Their formulation has turned out to be quite general for nonlinear wave problems with inverse spectral transforms (see [10] for the cubic nonlinear Schrödinger equation [3], for the Toda lattice, and [21] for the Ablowitz-Ladik equations) and makes accurate predictions about the modulational stability of multiphase solutions.

The modulation theory obtained from the prescription of multiphase averaging requires that certain conditions on the modulating variables are satisfied in order to avoid resonances. Being nondegeneracy conditions, they are satisfied for "most" values of the modulating variables, so that the regions of phase space corresponding to resonances are not of maximal dimension. If the modulating wavetrain manages to avoid these resonance regions, the modulation equations derived from multiphase averaging give the correct description of the macroscopic dynamics. On the other hand, it is not clear what differences to expect in the theory when the modulating wavetrain comes close to a resonance.

One expects some analogy with the idea of resonance as it occurs in the theory of averaging of perturbations of completely integrable finite-dimensional systems of ordinary differential equations. In that context, the hallmark of resonance is the rational dependence of frequencies. In the context of spatio-temporal modulations in infinite-dimensional systems however, it is not even clear how to generalize the simple idea of rational dependence – now there are wavenumbers associated with the oscillations as well as frequencies. Also associated with resonance in the finite-dimensional case is the slow evolution of phase information. This phase modulation surely also occurs in the infinite-dimensional case, and it is of some interest to understand the structure of the equations that should describe the phase modulation near a resonance.

Our aim in this paper is to begin to address the problems of modeling self-consistent spatio-temporal modulations in the presence of resonance by closely examining a simple example that occurs in an integrable system whose modulational description away from resonance is understood. We will be considering certain modulated wavetrains in the Ablowitz-Ladik equations

$$\begin{aligned} -i\partial_t Q_n - Q_{n+1} + 2Q_n - Q_{n-1} + Q_n R_n\left[Q_{n+1} + Q_{n-1}\right] &= 0\,, \\ -i\partial_t R_n + R_{n+1} - 2R_n + R_{n-1} - R_n Q_n\left[R_{n+1} + R_{n-1}\right] &= 0\,, \end{aligned} \tag{1}$$

where n is an integer index and t is a continuous time variable. These equations are known to be completely integrable by an inverse spectral transform [1]. Under the constraint

$$R_n(t) = \kappa\overline{Q_n(t)}\,, \qquad \kappa = \pm 1\,, \tag{2}$$

for real t, one obtains from (1) an integrable discretization of the cubic nonlinear Schrödinger equation:

$$-i\partial_t Q_n - \left(Q_{n+1} - 2Q_n + Q_{n-1}\right) + \kappa|Q_n|^2\left(Q_{n+1} + Q_{n-1}\right) = 0\,, \tag{3}$$

where $\kappa = 1$ (-1) corresponds to the defocusing (focusing) case.

[4] See Krichever [15] and Dubrovin [6] for reviews of the general methods; the methods are applied to the Ablowitz-Ladik equations in [21], but we will give a self-contained version of the details we need below.

We will begin in Section 2 by introducing a class of exact wavetrain solutions that are indexed by four complex parameters. In some regimes of parameter space, the waves will be linearly stable, and in these regimes we will find four well-posed modulation equations for the self-consistent evolution of these parameters by a straightforward multiple scaling procedure applied directly to (1). In Section 3 we will compare these modulation equations to the four modulation equations obtained by an algebro-geometric version of averaging in local conservation laws, a procedure that is formally applicable to integrable systems. We will find that the two different approaches yield different modulation equations. In particular, the modulation equations described in Section 3 will need to be constrained in order to be consistent with those derived in Section 2, and then this constrained system of equations is incomplete, there being two additional equations among those derived in Section 2 that model the modulation of phase information made visible by resonances that are preserved as the waves modulate. The modulation equations derived in Section 2 will thus be seen to model the slow evolution of phase information as well as expected wave-action variables. Specifically, we will show that

- the solution class we are considering is parametrized by the class of genus 1 elliptic curves satisfying two constraints among the four branch points,
- these constraints are preserved under the modulation equations given in Section 3, which after reducing out the constraints become two of the four modulation equations derived in Section 2, and
- these constraints imply two resonances, with two corresponding slow phase variables the evolution of which is modeled by the remaining two modulation equations derived in Section 2.

Section 4 generalizes these ideas, extending them to more complicated resonances in $1+1$ dimensional wave systems. Finally, we conclude with a discussion in which we will try to connect our work to work done by others on the topics of resonances and constraints preserved by modulation equations.

2. Antiperiod 2 solutions of the Ablowitz-Ladik equations

The system of ordinary differential equations (1) has a class of exact solutions satisfying spatially antiperiod 2 boundary conditions,

$$Q_{n+2}(t) = -Q_n(t)\,, \quad R_{n+2}(t) = -R_n(t)\,. \tag{4}$$

These exact solutions are very simple by virtue of the structure of the nonlinear terms. In fact, for such boundary conditions, the nonlinearity and dispersion disappear altogether, and the equations reduce to

$$-i\partial_t Q_n + 2Q_n = 0\,, \quad -i\partial_t R_n - 2R_n = 0\,, \tag{5}$$

which are trivially integrated, giving the solution

$$Q_n(t) = i^n \exp(-2it) \begin{cases} Q_o\,, & n \text{ odd,} \\ Q_e\,, & n \text{ even,} \end{cases} \tag{6}$$

and

$$R_n(t) = (-i)^n \exp(2it) \begin{cases} R_o\,, & n \text{ odd,} \\ R_e\,, & n \text{ even.} \end{cases} \tag{7}$$

The class of all such solutions of (1) is parametrized by the independent complex parameters Q_o, Q_e, R_o, and R_e. In the focusing ($\kappa = -1$) and defocusing ($\kappa = 1$) cases, we should choose $R_{e,o} = \kappa\overline{Q_{e,o}}$. In these cases, the solutions will be parametrized by the amplitudes and phases of Q_e and Q_o.

Of course in practical applications, a family of exact solutions to some system of equations is only useful if the solutions are stable to small perturbations. In particular, stability to small perturbations will imply modulational stability, which we aim to study in this paper. It is easy to assess the stability of the antiperiod 2 solutions given above in the focusing and defocusing cases using linear stability analysis. Let

$$Q_n(t) = i^n \exp(-2it) \begin{cases} Q_e + \epsilon_n(t)\,, & \text{for } n \text{ even}\,, \\ Q_o + \delta_n(t)\,, & \text{for } n \text{ odd}\,, \end{cases} \tag{8}$$

and suppose that the small perturbation has spatial structure described by a wavenumber $\tilde{k}$, so that

$$\mathrm{Re}(\epsilon_n(t)) = \epsilon_R(t)\exp(in\tilde{k}) + \overline{\epsilon_R(t)}\exp(-in\tilde{k})\,, \tag{9}$$

$$\mathrm{Im}(\epsilon_n(t)) = \epsilon_I(t)\exp(in\tilde{k}) + \overline{\epsilon_I(t)}\exp(-in\tilde{k})\,, \tag{10}$$

$$\mathrm{Re}(\delta_n(t)) = \delta_R(t)\exp(in\tilde{k}) + \overline{\delta_R(t)}\exp(-in\tilde{k})\,, \tag{11}$$

$$\mathrm{Im}(\delta_n(t)) = \delta_I(t)\exp(in\tilde{k}) + \overline{\delta_I(t)}\exp(-in\tilde{k})\,. \tag{12}$$

Then, defining

$$\boldsymbol{x} = (\epsilon_R(t), \epsilon_I(t), \delta_R(t), \delta_I(t))^T\,, \tag{13}$$

the small perturbation $\boldsymbol{x}$ solves the equation

$$\partial_t \boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}\,, \tag{14}$$

with the matrix $\boldsymbol{A}$ given by

$$\boldsymbol{A} = -2i\sin\tilde{k}\begin{bmatrix} 0 & 0 & 1-\kappa|Q_e|^2 & 0 \\ 0 & 0 & 0 & 1-\kappa|Q_e|^2 \\ 1-\kappa|Q_o|^2 & 0 & 0 & 0 \\ 0 & 1-\kappa|Q_o|^2 & 0 & 0 \end{bmatrix}. \tag{15}$$

To assess the stability, we only need to determine conditions sufficient to guarantee that the eigenvalues of $\boldsymbol{A}$ lie on the imaginary axis for all $\tilde{k}$, and that there are four linearly independent eigenvectors. The eigenvalues are

$$\lambda = \pm 2i\sin\tilde{k}\sqrt{(1-\kappa|Q_e|^2)(1-\kappa|Q_o|^2)}\,, \tag{16}$$

and both eigenvalues are double (all four coincide only if $\lambda = 0$). When the eigenvalues are imaginary for all perturbative wavenumbers $\tilde{k}$, and when there are four linearly independent eigenvectors, the antiperiod 2 oscillations will be neutrally stable. The only time the eigenvectors degenerate is in the defocusing case, if $|Q_{e,o}|^2 = 1$, and the dependence of the phase of the eigenvalues on the parameters is clear from (16). Thus, linear stability analysis makes the following predictions:

- In the focusing case, where $R_n(t) = -\overline{Q_n(t)}$ for real t, the oscillations are always (neutrally) stable.
- In the defocusing case, where $R_n(t) = \overline{Q_n(t)}$ for real t, the oscillations are (neutrally) stable whenever $|Q_o|$ and $|Q_e|$ are both less than or both greater than 1.

At face value, linear (neutral) stability means that a slightly perturbed initial condition will evolve in time so as to remain close to the exact solution parametrized by the constant values of Q_e, Q_o, R_e, and R_o for a long time. However, as first observed by Whitham [25] and verified again and again in applications, the conditions sufficient for linear stability agree with the conditions sufficient for *modulational* stability, as determined from the hyperbolicity of a set of modulation equations [5]. That is, taking the initial data to correspond to an antiperiod 2 wavetrain with slowly varying values of Q_e, Q_o, R_e, and R_o along the lattice (say described by functions of a slow continuous spatial variable $X = hn$ for a small spacing h), one expects that the field $Q_n(t)$ will evolve in such a way that on the time scale $T = ht$, the four parameters will continue to be functions of X. The evolution may not be trivial, but the scaling will be preserved.

The modulational stability of antiperiod 2 solutions to the Ablowitz-Ladik equations can be corroborated by numerical experiments. We want to integrate the ordinary differential equations (1) subject to an initial condition that has the spatial structure of an antiperiod 2 wavetrain, but with slowly varying values of the parameters Q_e, Q_o, R_e, and R_o. As the solution evolves in time under (1), we would like to extract the values of the four parameters, using the assumption that the local microstructure is indeed antiperiod 2. Ideally, the parameter values should be calculated from local measurements only, that is, measurements that are carried out over a few lattice points and over a short interval of time t. In the jargon of statistical mechanics [23], such measurements correspond to the values of *intensive variables* which are independent of the size of the system, as opposed to *extensive variables* whose values depend on nonlocal information such as the total number of lattice points.

A practical algorithm for calculating the parameters Q_e, Q_o, R_e, and R_o in the vicinity of a given lattice point n_0 and a given time t_0 might be simply to use the field values $Q_{n_0}(t_0)$, $Q_{n_0+1}(t_0)$, $R_{n_0}(t_0)$, $R_{n_0+1}(t_0)$ in the formulas (6) and (7), in which n is replaced by $n - n_0$ and t is replaced by $t - t_0$. The trouble with this algorithm is that it is very sensitive to the choice of the point (n_0, t_0) due to the highly oscillatory nature of the solution. In particular, the overall factor of $\exp(-2it)$ (respectively $\exp(2it)$) that multiplies $Q_n(t)$ (respectively $R_n(t)$) makes a precise measurement of the phase of the four parameters impossible (although slow variations in the phase can be observed), and without a way to tie the observation point n_0 to a fixed origin, the pair (Q_e, R_e) cannot be distinguished from the pair (Q_o, R_o).

In order to avoid these problems in directly measuring Q_e, Q_o, R_e, and R_o and yet still characterize the macroscopic properties of a slowly varying antiperiod 2 wavetrain, we define the four symmetrical quantities:

$$u_n(t) = \log\big(1 - Q_n(t)R_n(t)\big) + \log\big(1 - Q_{n+1}(t)R_{n+1}(t)\big), \tag{17}$$

$$v_n(t) = \frac{1}{2i}\big(R_{n+1}(t)Q_n(t) - Q_{n+1}(t)R_n(t)\big), \tag{18}$$

$$w_n(t) = Q_n(t)R_n(t) + Q_{n+1}(t)R_{n+1}(t), \tag{19}$$

$$y_n(t) = \frac{1}{h}\left[\frac{1}{8i}\log\left(\frac{Q_{n+2}(t)Q_{n+1}(t)^2Q_n(t)}{R_{n+2}(t)R_{n+1}(t)^2R_n(t)}\right) - \frac{1}{8i}\log\left(\frac{Q_{n+1}(t)Q_n(t)^2Q_{n-1}(t)}{R_{n+1}(t)R_n(t)^2R_{n-1}(t)}\right)\right], \tag{20}$$

where h is the small lattice spacing. In the focusing and defocusing cases, $v_n(t)$, $w_n(t)$, and $y_n(t)$ are real valued, and $u_n(t)$ is real valued modulo πi (so that the essential information is contained in the real part). These quantities represent local measurements, and hence intensive variables, since they can be calculated from any four consecutive lattice points at any fixed time. Note that in the case of an *exact* antiperiod 2 solution, these quantities have no variation in n or t. In terms of our four parameters, $u_n(t)$, $v_n(t)$, and $w_n(t)$ become

$$u = \log\big(1 - Q_eR_e\big) + \log\big(1 - Q_oR_o\big), \tag{21}$$

$$v = -\tfrac{1}{2}\big(Q_eR_o + R_eQ_o\big), \tag{22}$$

[5] Direct linear stability analysis is difficult, except for the analysis of the simplest solutions, where the coefficients of the linearization are constant (or of fixed discrete period on a lattice). In these simple cases, the equivalence between linear stability and hyperbolicity of modulation equations can be verified, as we will show below, and as done in [19] and [12], for example. Similar correspondences may be established in a diffusive, rather than dispersive setting [22].

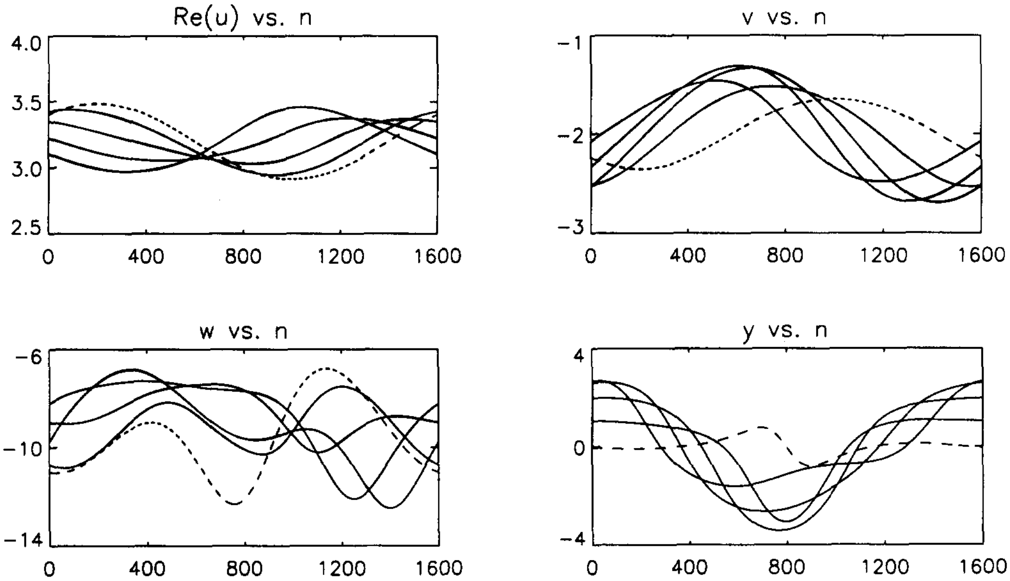

Fig. 1. A numerical simulation of the focusing Ablowitz-Ladik equations. The initial conditions locally satisfy antiperiod 2 boundary conditions, with parameters u, v, w, and y given as functions of X, a slow spatial variable. The initial fields $u(X,0)$, $v(X,0)$, $w(X,0)$, and $y(X,0)$ are shown with dashed curves, and several snapshots of $u(X,T_k)$, $v(X,T_k)$, $w(X,T_k)$, and $y(X,T_k)$ with $T_{k+1} - T_k = \mathcal{O}(1)$ (separated by a distance $\mathcal{O}(h^{-1})$ on the t scale) are superimposed.

$$w = Q_e R_e + Q_o R_o \,. \tag{23}$$

In the focusing and defocusing cases, there may be sharp inequalities satisfied by these quantities, in addition to the reality conditions. For example, the sharp inequalities $u \geq 0$ and $v^2 \leq \exp(u) - 1$ hold in the focusing case[6]. As a result of the spatial differencing for nonzero h, the quantity $y_n(t)$ vanishes identically for an exact antiperiod 2 solution. It is necessary to use the difference quotient because the spatial antiderivative of y,

$$Y_n(t) = \frac{1}{8i} \log\left(\frac{Q_{n+1}(t) Q_n(t)^2 Q_{n-1}(t)}{R_{n+1}(t) R_n(t)^2 R_{n-1}(t)} \right) , \tag{24}$$

grows linearly in time, with a rate that does not depend on n; thus $Y_n(t)$ itself cannot be accurately measured, although it is slowly varying in space. The spatial differencing eliminates the dependence on t, and when scaled by $1/h$ the difference converges as $h \downarrow 0$ to a smooth function of X and T, namely

$$\lim_{h \downarrow 0} y_n(t) = y(X,T) = \partial_X \frac{1}{4i} \log\left(\frac{Q_e(X,T) Q_o(X,T)}{R_e(X,T) R_o(X,T)} \right) . \tag{25}$$

An example evolution, corresponding to the focusing case, is shown in Fig. 1. Data corresponding to 1600 lattice points worth of spatially modulated antiperiod 2 microstructure were used to initialize the focusing Ablowitz-Ladik equations. The microscopic system was integrated for a period of time long compared to the fixed microscopic frequency. At several times during the evolution, the three macroscopic quantities $\mathrm{Re}(u)$, v, w, and y were extracted, and plotted as functions of X. Here and in other figures, in order to calculate y, we take $h = 1/N$, where N is the number of points in the simulation. The initial conditions are indicated by dashed curves.

[6] These inequalities will force the branch point configuration (see Section 3) to be of the form $\{z_1, z_2, z_3, z_4\} = \{i\alpha, i/\alpha, i\beta, i/\beta\}$, with α and β real and $\alpha\beta < 0$.

The small oscillations that can be seen upon close inspection of some of the curves in this and subsequent figures arise from the slight deviation of the local fields from an exact antiperiod 2 solution and are clearly not amplified on the macroscopic time scale $T = ht$. Moreover, they become smaller and even less consequential as the lattice is refined and hence, will not be part of the theory developed below. In order to support this supposition, we present in Fig. 2 the results of an additional numerical experiment using 400 rather than 1600 lattice points and using the same macroscopic initial conditions as those shown in Fig. 1. Taking snapshots at the same macroscopic times T, we obtain curves that are essentially the same as those shown in Fig. 1, except with binary oscillations that are approximately four times the size of the ones visible in Fig. 1. Thus, if these small oscillations affect the macroscopic dynamics at all, it will be on a macroscopic time scale T of at least order $O(1/h)$, which is beyond the time scales we shall consider here.

Similar experiments can be conducted in the defocusing case, where we expect to find two parameter regimes of modulational stability and one parameter regime of modulational instability. We present one numerical experiment for each of the two stable regimes. Fig. 3 demonstrates a case in which $|Q_n(t)|^2 > 1$ for all n, whereas in Fig. 4 we give the results of a numerical experiment conducted in the other stable component of the defocusing antiperiod 2 phase space corresponding to $|Q_n(t)|^2 < 1$ for all n. In the unstable defocusing cases when some magnitudes are greater than 1 and some magnitudes are less than 1, the behavior is so bad that it is not useful to present a picture. We suspect that the corresponding solutions to the underlying Ablowitz-Ladik equations achieve infinite amplitude in finite time [13,11].

Thus, we have numerically observed modulational stability of the microscopic antiperiod 2 ansatz in the parameter regimes where the fixed-parameter ansatz is linearly stable. Related phenomena have been observed by Bloch and Kodama [3], who found a class of solutions of the semi-infinite Toda lattice satisfying zero boundary conditions at the fixed end and enjoying modulational stability in fixed neighborhoods of the origin. The relationship between their observations and ours will be discussed further later.

We may attempt to describe the evolution of the fields $u(X,T)$, $v(X,T)$, $w(X,T)$, and $y(X,T)$ by deriving

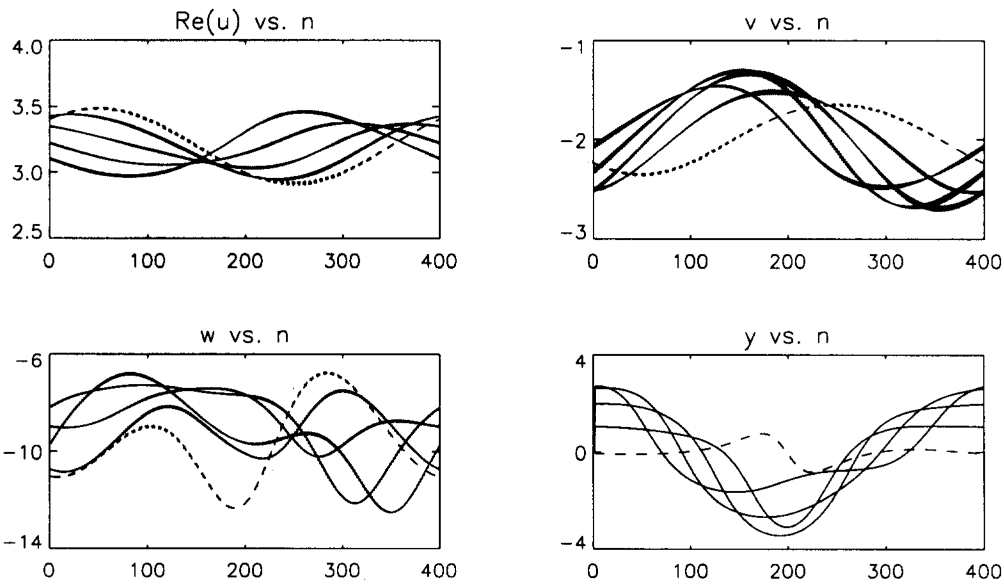

Fig. 2. A numerical simulation of the focusing Ablowitz-Ladik equations, with the same macroscopic initial data as in the experment of Fig. 1, but done with only one fourth as many lattice points. Note that the snapshot curves here differ essentially from those in Fig. 1 only in the size of the superimposed binary oscillations.

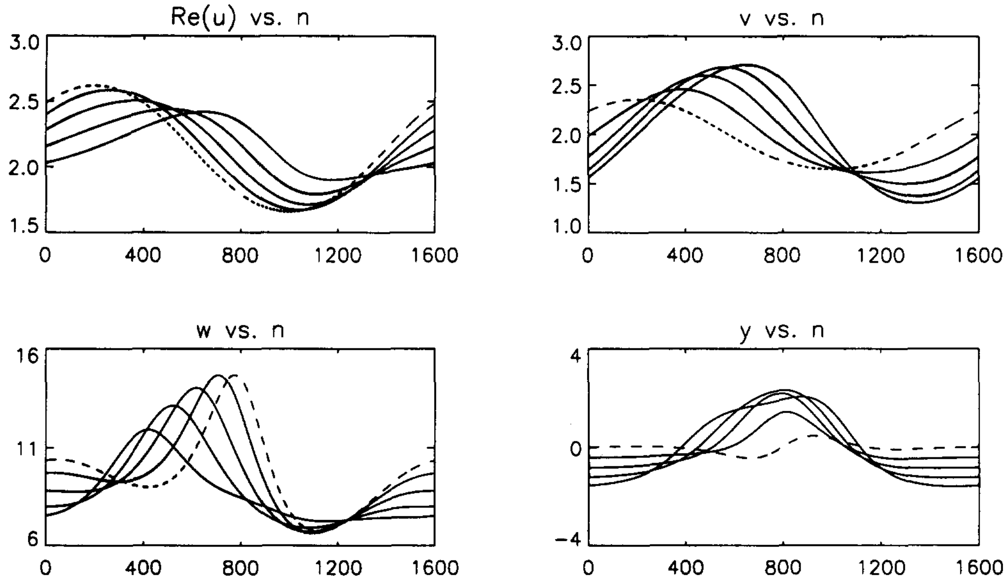

Fig. 3. A numerical simulation of the defocusing Ablowitz-Ladik equations, with initial data corresponding to a spatial modulation of an antiperiod 2 field, in which the amplitudes of Q_e and Q_o are both greater than 1.

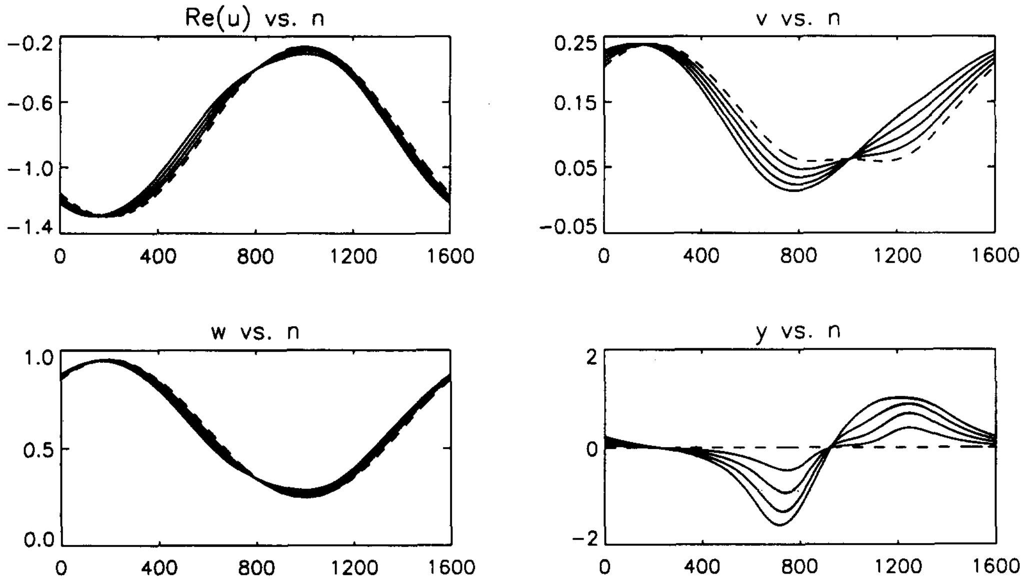

Fig. 4. A numerical simulation of the defocusing Ablowitz-Ladik equations, with initial data corresponding to a spatial modulation of an antiperiod 2 field, in which the amplitudes of Q_e and Q_o are both less than 1.

modulation equations from the Ablowitz-Ladik equations (1). The procedure is a standard multiple scale analysis of (1). Introduce a small parameter h and assume the perturbation expansions

$$\begin{aligned} Q_n(t;X,T) &= Q_n^{(0)}(t;X,T) + hQ_n^{(1)}(t;X,T) + \mathcal{O}(h^2)\,, \\ R_n(t;X,T) &= R_n^{(0)}(t;X,T) + hR_n^{(1)}(t;X,T) + \mathcal{O}(h^2)\,, \end{aligned} \tag{26}$$

where $X = hn$ and $T = ht$, and the leading terms are given by

$$Q_n^{(0)}(t;X,T) = i^n \exp(-2it) \begin{cases} Q_o(X,T), & n \text{ odd}, \\ Q_e(X,T), & n \text{ even}, \end{cases} \tag{27}$$

and

$$R_n^{(0)}(t;X,T) = (-i)^n \exp(2it) \begin{cases} R_o(X,T), & n \text{ odd}, \\ R_e(X,T), & n \text{ even}. \end{cases} \tag{28}$$

In the Ablowitz-Ladik equations (1), make the multiple scale replacements:

$$\partial_t \to \partial_t + h\partial_T, \quad Q_{n\pm 1} \to Q_{n\pm 1} \pm h\partial_X Q_{n\pm 1}, \quad R_{n\pm 1} \to R_{n\pm 1} \pm h\partial_X R_{n\pm 1}. \tag{29}$$

Now take all four variables n, t, X, and T, to be independent, and consider the limit $h \downarrow 0$. At leading order, the equations are solved identically because the pair $(Q_n^{(0)}, R_n^{(0)})$ is an exact antiperiod 2 solution in n and t. At order $\mathcal{O}(h)$, one obtains the following system of equations for the pair $(Q_n^{(1)}, R_n^{(1)})$:

$$\begin{aligned} &i\partial_t Q_n^{(1)} - 2Q_n^{(1)} + (1 - Q_n^{(0)}R_n^{(0)})(Q_{n+1}^{(1)} + Q_{n-1}^{(1)}) = -i\partial_T Q_n^{(0)} - (1 - Q_n^{(0)}R_n^{(0)})\partial_X(Q_{n+1}^{(0)} - Q_{n-1}^{(0)}), \\ &i\partial_t R_n^{(1)} + 2R_n^{(1)} - (1 - Q_n^{(0)}R_n^{(0)})(R_{n+1}^{(1)} + R_{n-1}^{(1)}) = -i\partial_T R_n^{(0)} + (1 - Q_n^{(0)}R_n^{(0)})\partial_X(R_{n+1}^{(0)} - R_{n-1}^{(0)}). \end{aligned} \tag{30}$$

These linear equations have homogeneous solutions of exactly the same form as the pair $(Q_n^{(0)}, R_n^{(0)})$. The driving terms on the right side are also antiperiod 2, with frequency 2. In order to keep the asymptotic expansion well ordered for times $T = \mathcal{O}(1)$, we must thus insist that the right side vanishes identically in n. Evaluating at odd and even sites gives two conditions on $Q_n^{(0)}$ and two conditions on $R_n^{(0)}$, namely,

$$\begin{aligned} &\partial_T Q_o + 2(1 - Q_oR_o)\partial_X Q_e = 0, \quad \partial_T Q_e + 2(1 - Q_eR_e)\partial_X Q_o = 0, \\ &\partial_T R_o + 2(1 - Q_oR_o)\partial_X R_e = 0, \quad \partial_T R_e + 2(1 - Q_eR_e)\partial_X R_o = 0. \end{aligned} \tag{31}$$

To check our work, we have integrated the modulation equations (31) numerically. There is such good agreement with the dynamics shown in the preceding figures that it is not instructive to show plots of the results. From the equations (31), and the limiting expressions (21) and (22), it is easy to see that

$$\partial_T u + \partial_X(4v) = 0, \tag{32}$$

$$\partial_T v + \partial_X(\exp u) = 0. \tag{33}$$

The field $v(X,T)$ may be eliminated to show that $u(X,T)$ satisfies the scalar wave equation

$$\partial_T^2 u = \partial_X^2(4 \exp u). \tag{34}$$

If this equation is hyperbolic, then the Cauchy problem at $T = 0$ will be well-posed, and this is interpreted as the modulational stability of the underlying (almost) antiperiod 2 solution. On the other hand, if (34) is elliptic, then the Cauchy problem at $T = 0$ will be ill-posed, indicating the modulational instability of the underlying solution. In fact, the nonlinear wave equation (34) is hyperbolic if $4\exp u$, the square of the local characteristic speeds, is positive. Unraveling this condition in terms of $Q_{e,o}$ and $R_{e,o}$ shows that hyperbolicity of (34) is equivalent to the linear stability condition derived above.

The nonlinear wave equation (34) can be placed in Riemann invariant form. Define the quantities

$$r_{\pm}(u,v) = \exp(u/2) \pm v\,. \tag{35}$$

Then, (34) takes the form of the diagonalized system

$$\partial_T r_{\pm}(X,T) \pm \big(r_+(X,T) + r_-(X,T)\big)\partial_X r_{\pm}(X,T) = 0\,. \tag{36}$$

It is also possible to derive modulation equations for $w(X,T)$ and $y(X,T)$ from the equations (31). First, define the quantity

$$Y = \frac{1}{4i}\log\left(\frac{Q_eQ_o}{R_eR_o}\right), \tag{37}$$

so that $y = \partial_X Y$. For reasons we have mentioned above, our description of the macroscopic dynamics must not depend on Y, which cannot be accurately measured, but it is convenient to use Y as an intermediate variable. We will need the Jacobian of the system of equations (21), (22), (23), and (37), which expresses the quantities u, v, w, and Y in terms of the quantities Q_e, Q_o, R_e, and R_o, whose dynamics are described by (31). The Jacobian is

$$\frac{\partial(u,v,w,Y)}{\partial(Q_e,Q_o,R_e,R_o)} = \begin{bmatrix} \dfrac{-R_e}{1-Q_eR_e} & \dfrac{-R_o}{1-Q_oR_o} & \dfrac{-Q_e}{1-Q_eR_e} & \dfrac{-Q_o}{1-Q_oR_o} \\ -\dfrac{R_o}{2} & -\dfrac{R_e}{2} & -\dfrac{Q_o}{2} & -\dfrac{Q_e}{2} \\ R_e & R_o & Q_e & Q_o \\ \dfrac{1}{4iQ_e} & \dfrac{1}{4iQ_o} & -\dfrac{1}{4iR_e} & -\dfrac{1}{4iR_o} \end{bmatrix}. \tag{38}$$

Inverting this matrix gives derivatives with respect to u, v, w, and Y. In particular, we will need the derivatives with respect to Y:

$$\begin{aligned} &\partial_Y Q_e = iQ_e\,, \quad \partial_Y Q_o = iQ_o\,, \\ &\partial_Y R_e = -iR_e\,, \quad \partial_Y R_o = -iR_o\,. \end{aligned} \tag{39}$$

Differentiation of (23) and (37) with respect to T, substitution from (31), and elimination of the spatial derivatives of Q_e, Q_o, R_e, and R_o using the inverse of the Jacobian (38) then yields equations for $w(X,T)$ and $Y(X,T)$ in the form:

$$\begin{aligned} \partial_T w &= a^{(u)}\partial_X u + a^{(v)}\partial_X v + a^{(w)}\partial_X w + a^{(Y)} y\,, \\ \partial_T Y &= b^{(u)}\partial_X u + b^{(v)}\partial_X v + b^{(w)}\partial_X w + b^{(Y)} y\,, \end{aligned} \tag{40}$$

where we have used the definition $\partial_X Y = y$. The coefficients in the equation for $w(X,T)$ are:

$$a^{(u)} = \frac{(1-Q_eR_e)(1-Q_oR_o)}{2(Q_oR_e-Q_eR_o)}\left[\frac{Q_oR_e^2}{R_o}+\frac{Q_o^2R_e}{Q_e}-\frac{Q_e^2R_o}{Q_o}-\frac{Q_eR_o^2}{R_e}\right], \tag{41}$$

$$a^{(v)} = 4-2Q_eR_e-2Q_oR_o, \tag{42}$$

$$a^{(w)} = \frac{1}{2(Q_oR_e-Q_eR_o)}\left[\frac{Q_oR_e^2}{R_o}+\frac{Q_o^2R_e}{Q_e}-\frac{Q_eR_o^2}{R_e}-\frac{Q_e^2R_o}{Q_o}+2Q_e^2R_o^2-2Q_o^2R_e^2\right], \tag{43}$$

$$a^{(Y)} = 2i(Q_eR_e-Q_oR_o)(Q_oR_e-Q_eR_o). \tag{44}$$

The coefficients in the equation for $Y(X,T)$ are:

$$b^{(u)} = \frac{i(1-Q_eR_e)(1-Q_oR_o)}{8(Q_eR_o-Q_oR_e)(Q_oR_o-Q_eR_e)}\left[(2-Q_oR_o-Q_eR_e)\left(\frac{Q_eR_o}{Q_oR_e}+\frac{Q_oR_e}{Q_eR_o}-6\right)\right.$$
$$\left.+(Q_eR_o+Q_oR_e)^2\left(\frac{1-Q_oR_o}{Q_o^2R_o^2}+\frac{1-Q_eR_e}{Q_e^2R_e^2}\right)\right], \tag{45}$$

$$b^{(v)} = \frac{i}{2(Q_oR_e-Q_eR_o)}\left[\frac{Q_e}{Q_o}+\frac{R_e}{R_o}-\frac{Q_o}{Q_e}-\frac{R_o}{R_e}\right], \tag{46}$$

$$b^{(w)} = \frac{i}{8(Q_oR_o-Q_eR_e)(Q_oR_e-Q_eR_o)}\left[(2-Q_oR_o-Q_eR_e)\left(\frac{Q_eR_o}{Q_oR_e}+\frac{Q_oR_e}{Q_eR_o}-6\right)\right.$$
$$+2\frac{Q_eR_e}{Q_oR_o}+2\frac{Q_oR_o}{Q_eR_e}+2Q_eR_e+2Q_oR_o+\left(\frac{Q_e}{Q_o}\right)^2+\left(\frac{Q_o}{Q_e}\right)^2+\left(\frac{R_e}{R_o}\right)^2+\left(\frac{R_o}{R_e}\right)^2$$
$$\left.-3\frac{Q_e^2R_o}{Q_o}-3\frac{Q_oR_e^2}{R_o}-3\frac{Q_o^2R_e}{Q_e}-3\frac{Q_eR_o^2}{R_e}+4(Q_oR_e-Q_eR_o)^2\right], \tag{47}$$

$$b^{(Y)} = -\frac{1-Q_oR_o}{2}\left[\frac{Q_e}{Q_o}+\frac{R_e}{R_o}\right]-\frac{1-Q_eR_e}{2}\left[\frac{Q_o}{Q_e}+\frac{R_o}{R_e}\right]. \tag{48}$$

These eight coefficients, given above in terms of Q_e, Q_o, R_e, and R_o, are in fact well-defined functions of u, v, and w (we will see below that they do not depend on Y). For fixed values of u, v, w, and Y, the possible values of Q_e, Q_o, R_e, and R_o are permuted by the elements of the symmetry group of (21), (22), (23), and (37), which is generated by the following transformations:

$$(Q_e,Q_o,R_e,R_o)\mapsto(Q_o,Q_e,R_o,R_e), \tag{49}$$

$$(Q_e,Q_o,R_e,R_o)\mapsto(iQ_e,iQ_o,-iR_e,-iR_o). \tag{50}$$

All eight coefficients are invariant under these transformations, and hence do not depend on any particular choice of branch in the inversion of (21), (22), (23), and (37). This means that we have derived closed modulation equations for u, v, w, and Y.

The quantity Y, however, is impossible to measure accurately. In order to obtain a description of the dynamics that only depends on quantities that can be unambiguously observed, all reference to Y must be eliminated from the modulation equations in favor of the spatial derivative y. This can be done since none of the eight coefficients in the expressions (40) depend on Y, which can be easily seen by directly differentiating them with respect to Y, and substituting from (39). The resulting derivatives vanish identically. So, differentiation of the equation for Y with respect to X eliminates all reference to the unmeasurable antideriviative Y, and we obtain the equations for w and y:

$$\partial_T w = a^{(u)}(u,v,w)\partial_X u + a^{(v)}(u,v,w)\partial_X v + a^{(w)}(u,v,w)\partial_X w + a^{(Y)}(u,v,w)y, \tag{51}$$

$$\partial_T y = \partial_X\left[b^{(u)}(u,v,w)\partial_X u + b^{(v)}(u,v,w)\partial_X v + b^{(w)}(u,v,w)\partial_X w + b^{(Y)}(u,v,w)y\right]. \tag{52}$$

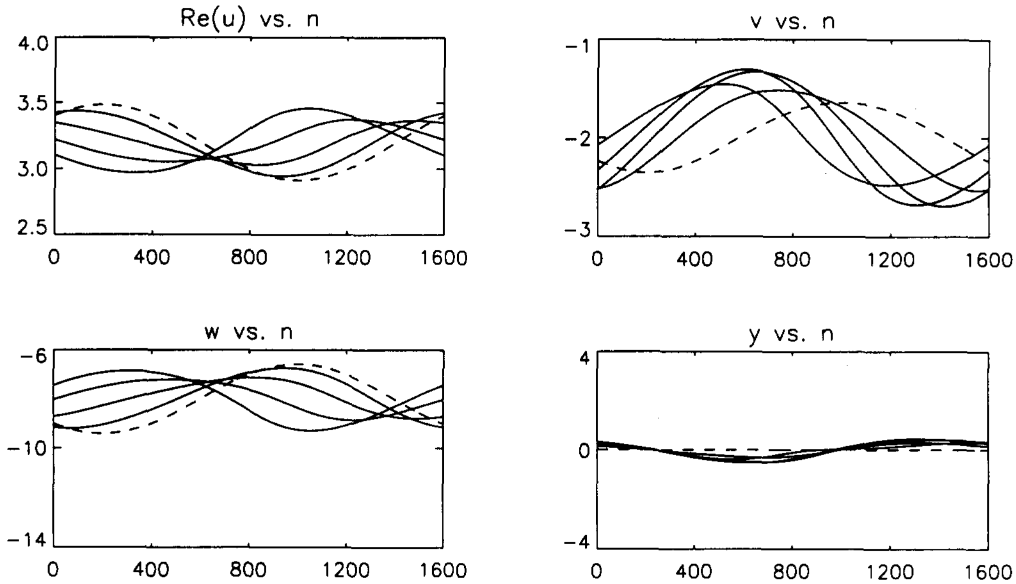

Fig. 5. A numerical simulation of the focusing Ablowitz-Ladik equations. Compare this simulation to that given in Fig. 1.

The four equations (32), (33), (51), and (52) make up a closed description of all the macroscopic observables associated with a modulated antiperiod 2 wavetrain.

We have already presented numerical experiments corroborating the modulational stability criteria that have been reproduced by these modulation equations. However, we now have another feature predicted by the theory that we may attempt to verify numerically, namely the closure of the equations for u and v. In Figs. 5, 6, and 7, we present the results of numerical experiments similar to the three already given, the only differences being the initial conditions on the slow fields $w(X)$ and $y(X)$. Observe that the evolutions of the fields u and v in Figs. 5, 6, and 7 are identical to the evolutions of u and v in Figs. 1, 3, and 4 respectively, while the evolutions of the fields w and y differ. These experiments support the conclusion that the fields u and v evolve independently of w and y.

We have derived four modulation equations for the four independent observable parameters of the antiperiod 2 local ansatz. We will see below that the antiperiod 2 solution can be written in the form

$$Q_n(t) = \exp\big(i(\tfrac{1}{2}\pi n - 2t - \theta_0^{(0)}(u,v,w,y))\big)\tilde{Q}(\pi n - \theta_1^{(0)}(u,v,w,y);u,v)\,, \tag{53}$$

$$R_n(t) = \exp\big(-i(\tfrac{1}{2}\pi n - 2t - \theta_0^{(0)}(u,v,w,y))\big)\tilde{R}(\pi n - \theta_1^{(0)}(u,v,w,y);u,v)\,, \tag{54}$$

where $\tilde{Q}(\theta;u,v)$ and $\tilde{R}(\theta;u,v)$ are given functions having period 2π in θ, and whose shape is determined by the parameters u and v. Thus, we see that we have derived modulation equations not only for the action variables that describe the wave shape, *but also for the arbitrary phase constants.* Moreover, these phase constants are not hidden variables, in the sense that they may be unambiguously observed, but their values do not affect the self-consistent modulational evolution of u and v.

3. Connection with the finite genus theory

The Ablowitz-Ladik equations (1) are integrable. Concretely, this means that the nonlinear system (1) is merely the consistency condition for two linear problems in which the functions $Q_n(t)$ and $R_n(t)$ appear as

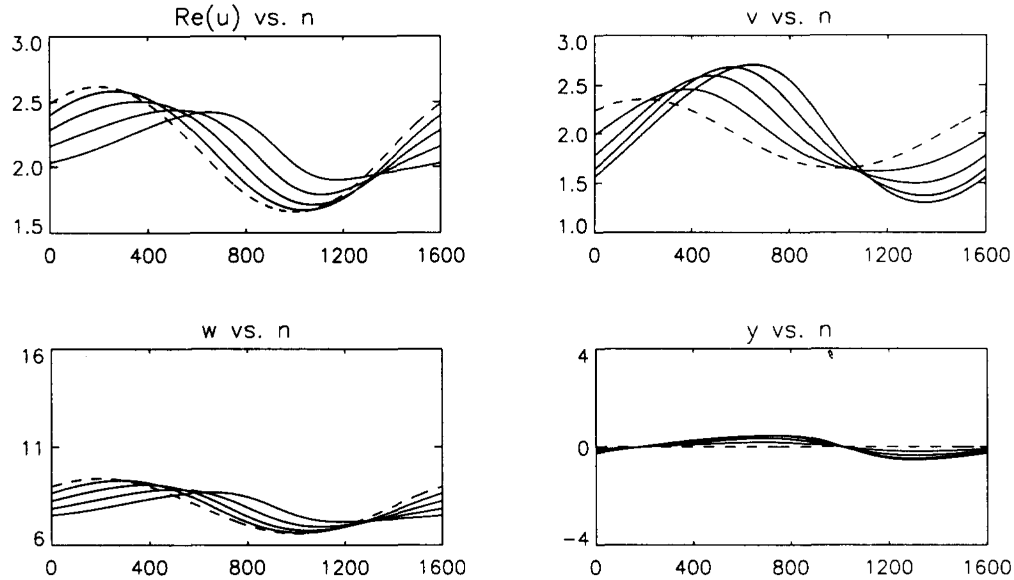

Fig. 6. A numerical simulation of the defocusing Ablowitz-Ladik equations, with large amplitude modulated antiperiod 2 initial data. Compare this simulation with that given in Fig. 3.

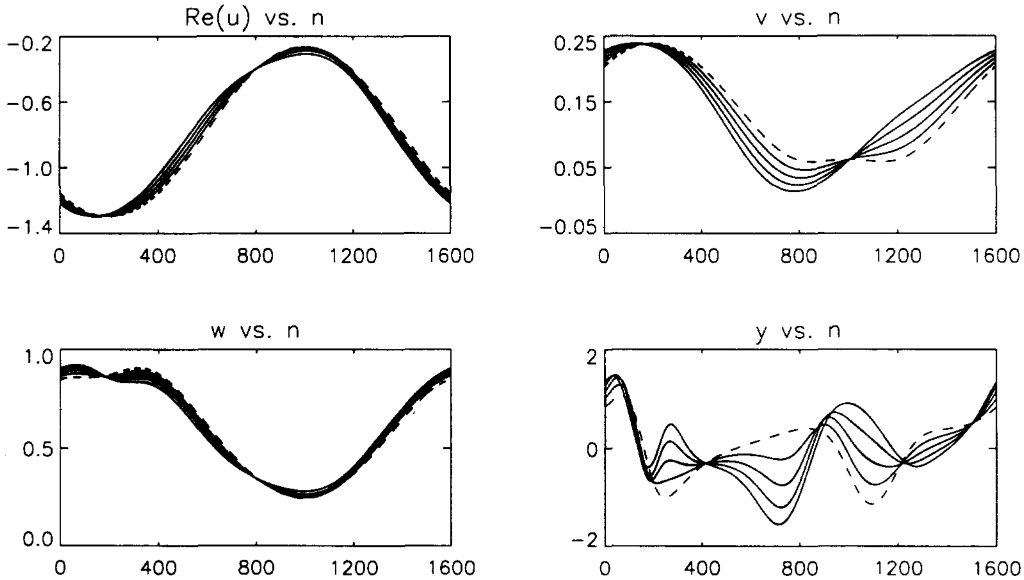

Fig. 7. A numerical simulation of the defocusing Ablowitz-Ladik equations, with small amplitude modulated antiperiod 2 initial data. Compare this simulation with that given in Fig. 4.

coefficients. The two linear problems further involve an arbitrary complex parameter z, and are said to make up a *Lax pair*. In the case when $Q_n(t)$ and $R_n(t)$ solve (1), there exists a basis of simultaneous solutions to both problems of the Lax pair for almost all values of the parameter z. The values of z for which the basis of simultaneous solutions degenerates are called *branch points*; the branch points do not depend on n or t. The Lax pair consists of a difference equation (in n) for a vector function $\boldsymbol{u}$

$$u_{n+1}(t,z) = \begin{bmatrix} z & Q_n(t) \\ zR_n(t) & 1 \end{bmatrix} u_n(t,z)\,, \tag{55}$$

and a differential equation in t for $\boldsymbol{u}$

$$-i\partial_t u_n(t,z) = \begin{bmatrix} z-1-Q_n(t)R_{n-1}(t) & Q_n(t)-z^{-1}Q_{n-1}(t) \\ zR_{n-1}(t)-R_n(t) & 1-z^{-1}+R_n(t)Q_{n-1}(t) \end{bmatrix} u_n(t,z)\,. \tag{56}$$

There are many solutions $Q_n(t)$ and $R_n(t)$ of the Ablowitz-Ladik equations (1) for which the number of branch points z_k is finite. This class of solutions (studied in [21]) includes all spatially periodic solutions and in particular the antiperiod 2 solutions described in Section 2. It can be shown that there is a special basis of functions, $u_n^+(t,z)$ and $u_n^-(t,z)$, such that analytic continuation (in z) of the two basis elements through a closed loop around one of the branch points results in permutation of the functions. Thus, the functions $u_n^+(t,z)$ and $u_n^-(t,z)$ can be considered to be stereographic projections of a single function $u_n(t,P)$ where P is a point on a finite-genus hyperelliptic Riemann surface that is a two-sheeted branched cover of the z-plane. In this way, we may associate a fixed Riemann surface with each such solution to the Ablowitz-Ladik equations (1). The branch points z_k are constants of motion that play the role of action variables.

So, choose $2g+2$ nonzero distinct complex numbers z_k and let Γ be the Riemann surface of the relation

$$r^2 = \prod_{k=1}^{2g+2} (z-z_k)\,. \tag{57}$$

The values of z and r, linked by (57) provide coordinates for points on Γ. Define ∞^+ (respectively ∞^-) to be the point on Γ over $z=\infty$ near which $r = z^{g+1}(1+\mathcal{O}(z^{-1}))$ (respectively $r = -z^{g+1}(1+\mathcal{O}(z^{-1}))$) for large z. Choose a number η satisfying

$$\eta^2 = \prod_{k=1}^{2g+2} z_k\,, \tag{58}$$

and define 0^+ (respectively 0^-) to be the point on Γ over $z=0$ near which $r=\eta(1+\mathcal{O}(z))$ (respectively $r=-\eta(1+\mathcal{O}(z))$) for small z. Finally, choose almost any set of g points on Γ, P_j. Then, according to Lemmas 2.3 and 2.4 of [21], there is a unique function $\boldsymbol{u}$ having asymptotic expansions of the form

$$u_{1,n}(t,P) = \begin{cases} z^n\left(c_{1,n}^{(1)}(t)+c_{1,n}^{(2)}(t)z+\mathcal{O}(z^2)\right), & P\to 0^-, \\ \exp\left(i(1-z^{-1})t\right)\left(d_{1,n}^{(1)}(t)+d_{1,n}^{(2)}(t)z+\mathcal{O}(z^2)\right), & P\to 0^+, \\ z^n\exp(i(z-1)t)\left(1+a_{1,n}^{(2)}(t)z^{-1}+\mathcal{O}(z^{-2})\right), & P\to\infty^+, \\ b_{1,n}^{(1)}(t)+b_{1,n}^{(2)}(t)z^{-1}+\mathcal{O}(z^{-2}), & P\to\infty^-, \end{cases} \tag{59}$$

$$u_{2,n}(t,P) = \begin{cases} z^{n+1}\left(c_{2,n}^{(1)}(t) + c_{2,n}^{(2)}(t)z + \mathcal{O}(z^2)\right), & P \to 0^-, \\ \exp(i(1-z^{-1})t)\left(1 + d_{2,n}^{(2)}(t)z + \mathcal{O}(z^2)\right), & P \to 0^+, \\ z^n \exp(i(z-1)t)\left(a_{2,n}^{(1)}(t) + a_{2,n}^{(2)}(t)z^{-1} + \mathcal{O}(z^{-2})\right), & P \to \infty^+, \\ z\left(b_{2,n}^{(1)}(t) + b_{2,n}^{(2)}(t)z^{-1} + \mathcal{O}(z^{-2})\right), & P \to \infty^-, \end{cases} \tag{60}$$

and otherwise holomorphic, except at the points P_j where both components have simple poles. Moreover, as shown in Lemmas 3.1 and 3.2 of [21], this function is a simultaneous solution of (55) and (56), provided that the functions $Q_n(t)$ and $R_n(t)$ are taken to be determined by the function $\boldsymbol{u}$ according to

$$Q_n(t) = d_{1,n+1}^{(1)}(t), \quad R_n(t) = a_{2,n+1}^{(1)}(t). \tag{61}$$

It then follows from the consistency that these two functions solve the Ablowitz-Ladik equations. Further solutions may be obtained from (61) through the scaling symmetry:

$$(Q_n(t), R_n(t)) \mapsto (\xi Q_n(t), \xi^{-1} R_n(t)). \tag{62}$$

The data involved in specifying the solution were the $2g+2$ branch points, the number η, the $g+1$ points P_j, and the complex scaling constant ξ.

If one chooses a canonical basis of homology cycles $(a_1, \ldots, a_g, b_1, \ldots, b_g)$ for a Riemann surface Γ of genus g, then there is associated with Γ a well-defined *Riemann theta function*, $\Theta(w_1, \ldots, w_g)$. The function $\boldsymbol{u}$ as well as the corresponding solution $Q_n(t)$ and $R_n(t)$ to (1) are given in terms of this theta function, which is a function on $\mathbb{C}^g$ that is periodic in each (real) direction[7] with period 2π. The function $Q_n(t)$ is given by

$$Q_n(t) = A\exp(i\theta_0)\frac{\Theta(\theta_1, \ldots, \theta_g)}{\Theta(\theta_1 + \beta_1, \ldots, \theta_g + \beta_g)} \tag{63}$$

(compare with Eq. (3.12) of [21]) where $\theta_j = k_j n - \omega_j t + \alpha_j$, and β_j and A are complex constants. The wavenumbers k_j, frequencies ω_j, and the numbers β_j are determined by the Riemann surface Γ with choice of homology cycles. In the focusing and defocusing cases, it is always possible to choose a canonical basis of homology cycles so that all frequencies and wavenumbers are real. With this choice, it is evident from the formula (63) and the periodicity of the theta function that $Q_n(t)$ will be a multiperiodic function of n and t individually referred to as a *multiphase solution*. On the other hand, the initial phases α_j are *not* determined by the branch points. They represent the location of the orbit in the $g+1$ torus when $n=0$ and $t=0$.

A generic Riemann surface Γ will give rise to frequencies and wavenumbers that cause the evolution in either n or t to cover the $g+1$ torus. In this case, the initial phase information (the values of α_j) will not be accessible to a macroscopic observer of the multiphase solution $Q_n(t)$, since the motion that fills the torus is fast on the scales X and T. The only information available to the macroscopic observer of the multiphase solution $Q_n(t)$ is the set of frequencies ω_j and wavenumbers k_j, which could be recovered by Fourier analysis at fixed n (for the frequencies) and fixed t (for the wavenumbers). The $2g+2$ branch points z_k are, in this generic case, equivalent information to the $g+1$ frequencies and the $g+1$ wavenumbers. Thus, the astute macroscopic observer of a typical multiphase solution $Q_n(t)$ sees a fixed Riemann surface Γ of some genus

[7] See Dubrovin [6] for a discussion of theta functions. Note that Dubrovin uses a normalization in which the theta function has *imaginary* periods.

g, *and nothing else*, all other information being higher order in the limit $h \downarrow 0$. This statement was put on a rigorous basis by Lax and Levermore in the context of the zero-dispersion limit, and it was used as an assumption in the constructive modulation theory of Flaschka, Forest, and McLaughlin.

The theory of modulated generic multiphase solutions in integrable systems, and in particular in the Ablowitz-Ladik equations, is thus a theory of slowly deforming Riemann surfaces $\Gamma(X,T)$. This theory is described in Section 7 of [21], in which several of the formulas appearing below are derived. If the branch points are distinct for all X at $T = 0$, then any infinitesimal deformation will preserve the genus, and the evolution in T may be constrained to any fixed genus g. From formulas for the function $\boldsymbol{u}$ that solves the Lax pair, it is possible to show that in the focusing and defocusing cases (to which we restrict attention from this point onward), the modulation equations describing the evolution of $\Gamma(X,T)$ can be written in the compact form

$$\partial_T \Omega_{(3)} = \partial_X \Omega_{(2)}, \tag{64}$$

where $\Omega_{(2)}$ and $\Omega_{(3)}$ are unique differentials on Γ of the second and third kind respectively. The wavenumbers k_j and frequencies ω_j are expressed in terms of these two differentials:

$$k_0 = i \int_{0^+}^{\infty^+} \left[\Omega_{(3)} - \frac{dz}{z+1} \right], \tag{65}$$

$$k_j = -i \int_{b_j} \Omega_{(3)}, \tag{66}$$

$$\omega_0 = 2 - i \int_{0^+}^{\infty^+} \left[\Omega_{(2)} - i\,dz - i\frac{dz}{z^2} \right], \tag{67}$$

$$\omega_j = i \int_{b_j} \Omega_{(2)} \tag{68}$$

(compare with Eqs. (2.15), (3.14) and (3.15) of [21]) where $j = 1, \ldots, g$. From these formulas, it is easy to see that spatio-temporal modulation described by (64) implies *conservation of waves* in each phase variable independently:

$$\partial_T k_j + \partial_X \omega_j = 0, \tag{69}$$

where $j = 0, \ldots, g$. As Γ evolves under (64), the branch points z_k also evolve, and they are described by a closed system of quasilinear partial differential equations. Moreover the branch points themselves are *Riemann invariants*, that is, the modulation equations take the form

$$\partial_T z_k + c(z_k; z_1, \ldots, z_g) \partial_X z_k = 0, \tag{70}$$

for $k = 1, \ldots, 2g + 2$. The characteristic speeds c are real (and thus the modulation equations are hyperbolic) if the branch points lie on the unit circle in the z-plane, as is known to hold in the defocusing case when $|Q_n| < 1$ for all n.

The antiperiod 2 waves studied in Section 2 correspond to surfaces of genus $g = 1$, so we should go into a little more detail for this special case. The general modulation equations for the slow deformation of a genus $g = 1$ Riemann surface $\Gamma(X,T)$ under (1) can be written in the form (64) with

$$\Omega_{(3)} = \left[\frac{1}{2}\left(\frac{1}{z} - \frac{\eta}{zr}\right) + \frac{1}{2}\left(\frac{1}{z} + \frac{z^2}{zr}\right) - \frac{1}{2z} + \frac{C^{(3)}}{r}\right] dz\,, \tag{71}$$

$$\Omega_{(2)} = i\left[\frac{1}{2}\left(1 + \frac{z^2}{r}\right) + \frac{1}{2}\left(\frac{1}{z^2} + \frac{\eta}{z^2 r}\right) - \frac{\eta}{4zr}\sum_{j=1}^{4}\frac{1}{z_j} - \frac{z}{4r}\sum_{j=1}^{4} z_j + \frac{C^{(2)}}{r}\right] dz\,, \tag{72}$$

where the constants $C^{(2)}$ and $C^{(3)}$ are chosen so that integrals of the two differentials over the a cycle of Γ vanish. From Eqs. (64) one obtains four closed modulation equations in Riemann invariant form (70) by expanding near the branch points z_k.

The modulation theory of the antiperiod 2 waves, as developed in Section 2, does not fit this generic description because only the quantities u and v depend on the four branch points, whereas the quantities w and y have nothing to do with $\Gamma(X,T)$ at all. Furthermore, the two remaining degrees of freedom among the four branch points are fixed throughout the modulation. To describe the situation in more detail, let us begin with

Proposition 1. Exact solutions of the Ablowitz-Ladik equations satisfying antiperiod 2 boundary conditions are described by Riemann surfaces of genus $g = 1$ satisfying the constraints

$$\eta = -1\,, \qquad \sum_{k=1}^{4}\left(z_k + \frac{1}{z_k}\right) = 0\,. \tag{73}$$

The remaining degrees of freedom among the four branch points of Γ are described in terms of u and v by

$$\sum_{k=1}^{4} z_k = 4iv\,, \qquad \sum_{k<j} z_k z_j = 2(2\exp(u) - 1 - 2v^2)\,. \tag{74}$$

Proof. First, we will show how antiperiod 2 boundary conditions lead to (73) and (74). In order to do this, we need to obtain the description of the antiperiod 2 exact solution in terms of the spectral data (Γ, η) from the values of Q_n and R_n at some fixed time. The tool for this translation is the *monodromy matrix*:

$$S_n(t) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} z & Q_{n+1}(t) \\ zR_{n+1}(t) & 1 \end{bmatrix} \begin{bmatrix} z & Q_n(t) \\ zR_n(t) & 1 \end{bmatrix}, \tag{75}$$

which encodes all dynamical information about the antiperiod 2 solution. It is easy to see from the Lax pair (55) and (56) and the antiperiod 2 boundary conditions that the polynomial

$$\begin{aligned} P(z) &= (\operatorname{Tr} S)^2 - 4\operatorname{Det} S \\ &= z^4 + 2(Q_{n+1}R_n - Q_nR_{n+1})z^3 + ((Q_{n+1}R_n - Q_nR_{n+1})^2 - 2 \\ &\quad +4(1 - Q_nR_n)(1 - Q_{n+1}R_{n+1}))z^2 - 2(Q_{n+1}R_n - Q_nR_{n+1})z + 1 \\ &= z^4 - 4iv_nz^3 + 2(2\exp(u_n) - 1 - 2v_n^2)z^2 + 4iv_nz + 1 \end{aligned} \tag{76}$$

is independent of n and t. In writing the last line we have used the definitions of u_n and v_n from Section 2. This polynomial defines the Riemann surface Γ associated with the exact solution through the relation $r^2 = P(z)$. The value of η that we should choose is also built into the monodromy matrix. It is given by

$$\eta = -(S_{11} - S_{22})\big|_{z=0} = -1\,. \tag{77}$$

Thus, we see that the spectral data (Γ, η) of an exact solution of (1) satisfying antiperiod 2 boundary conditions gives rise to a Riemann surface of genus $g = 1$ satisfying (73), and that the remaining branch point information is described by u and v according to (74).

Now, we will prove the converse, that the constraints (73) imply that the solution given by the formula (63) satisfies antiperiod 2 boundary conditions. We need to calculate the wavenumbers k_0 and k_1 according to the formulas (65) and (66). Note that the constraints (73) imply that there is a holomorphic involution on Γ:

$$H : (z, r) \mapsto \left(-\frac{1}{z}, \frac{r}{z^2}\right) . \tag{78}$$

The a cycle on Γ, chosen so that the modulation equations have the form (64), will be transformed by H into $H(a)$ which may be smoothly deformed into $-a$. Then the b cycle can be chosen so that $H(b)$ may be smoothly deformed into b. Thus we may replace integrals of $\Omega_{(3)}$ over $H(b)$ by integrals over b at the cost of the integral residues of $\Omega_{(3)}$. Also, we have $H(\infty^{\pm}) = 0^{\mp}$. The pull-back of H operates on $\Omega_{(3)}$ as

$$H^* \Omega_{(3)} = -\Omega_{(3)} . \tag{79}$$

We compute

$$\int_b \Omega_{(3)} = \int_{H(b)} H^* \Omega_{(3)} = - \int_{H(b)} \Omega_{(3)} = - \int_b \Omega_{(3)} + 2\pi i , \tag{80}$$

and thus $k_1 = \pi$ modulo 2π. Similar arguments, again using the involution H, show that (73) leads to $k_0 = \pi/2$. The periodicity of the theta function and the complex exponential then shows that the formula (63) satisfies antiperiod 2 boundary conditions. □

These arguments, along with the formula (63), yield the representation of the class of antiperiod 2 solutions given at the end of Section 2. An unusual feature of the antiperiod 2 solutions in the context of general genus $g = 1$ solutions is that making the choice $k_0 = \pi/2$ and $k_1 = \pi$ *fixes the frequencies* to the values $\omega_0 = 2$ and $\omega_1 = 0$. Choosing arbitrary values of the two wavenumbers introduces two constraints on the four branch points. Now, for a generic choice of the wavenumbers the two frequencies parametrize the possible surfaces Γ subject to these constraints. The fact that antiperiod 2 waves are not generic in the sense that the frequencies are independent of the remaining degrees of freedom in Γ is related to the fact that y is a macroscopic observable and is responsible for preserving the conditions $k_0 = \pi/2$ and $k_1 = \pi$ under modulation.

So only two of the variables u, v, w, and y making up the macroscopic description of a modulated antiperiod 2 wavetrain contain information about the Riemann surface Γ, the remaining two degrees of freedom of Γ being fixed by the constraints (73). As we have seen, generic self-consistent modulations of genus $g = 1$ wavetrains are described by (64) (or equivalently (70)) and involve slow motion of all four independent branch points of Γ. In spite of these structural features that set the special case of modulation theory for antiperiod 2 waves in stark contrast to the general modulation theory of genus $g = 1$ waves, *the general modulation equations for the Riemann surface continue to hold.* These equations are consistently constrained by (73), and are equivalent to the system of equations for u and v, (32) and (33). We make these facts precise in

Proposition 2. Let the $g = 1$ Riemann surface Γ satisfy the constraints (73) identically in X at $T = 0$. Then,
- The general $g = 1$ modulation equations (64) preserve the constraints (73) for all $T > 0$.
- The general $g = 1$ modulation equations (64), subject to the consistent constraints (73) take the simple form of Eqs. (32) and (33).

Proof. The fact that the constraints (73) are preserved by (64) follows from conservation of waves (69) and the fact that the frequencies are constant over the whole class of antiperiod 2 exact solutions. Suppose that at $T = 0$ the constraints are satisfied so that for all X, $k_0 = \pi/2$ and $k_1 = \pi$. Then, also $\omega_0 = 2$ and $\omega_1 = 0$ for all X. Conservation of waves then guarantees that $\partial_T k_0 = \partial_T k_1 = 0$, and hence the antiperiod 2 boundary conditions (equivalently the constraints (73)) are preserved.

Proving the second part of the proposition requires imposing the constraints (73) on the modulation equations (64). The holomorphic involution H used in the proof of Proposition 1 assists in the simplification of the resulting equations. One first expands the differentials $\Omega_{(2)}$ and $\Omega_{(3)}$ in asymptotic series in z near the point 0^+ on Γ. Equating powers of z in (64) into which the expansions have been substituted leads to equations of the form

$$\partial_T \boldsymbol{v} + A(\boldsymbol{v})\partial_X \boldsymbol{v} = 0\,, \tag{81}$$

where $\boldsymbol{v} = (\eta, \sigma_1, \sigma_2, \sigma_3)$ and

$$\begin{aligned} \sigma_1 &= -z_1 z_2 z_3 - z_1 z_2 z_4 - z_1 z_3 z_4 - z_2 z_3 z_4\,, \\ \sigma_2 &= z_1 z_2 + z_1 z_3 + z_1 z_4 + z_2 z_3 + z_2 z_4 + z_3 z_4\,, \\ \sigma_3 &= -z_1 - z_2 - z_3 - z_4\,. \end{aligned} \tag{82}$$

The constraints (73) require that we impose $\eta = -1$ and $\sigma_1 + \sigma_3 = 0$ on the matrix $A(\boldsymbol{v})$. Doing this, and evaluating the elliptic integrals using the involution H yields that

$$\partial_T(\sigma_1 + \sigma_3) = 0\,, \quad \partial_T \eta = 0\,. \tag{83}$$

These equations express what we already knew from conservation of waves, namely that the constraints (73) are preserved under modulation. The remaining two equations can be rewritten by using (76) to find

$$\sigma_1 = 4iv\,, \quad \sigma_2 = 2(2\exp(u) - 1 - 2v^2)\,, \quad \sigma_3 = -4iv\,, \tag{84}$$

and then applying the chain rule to obtain equations for $u(X,T)$ and $v(X,T)$. The resulting equations are exactly (32) and (33). □

So, the modulation theory obtained by multiphase averaging continues to hold in the special case of self-consistent modulations of antiperiod 2 wavetrains. However, comparing the modulational description obtained in Section 2 to the two propositions leads to

Corollary 1. Modulations of the fields w and y are not described by the modulation equations (64) obtained by multiphase averaging.

The description of the macroscopic dynamics provided by multiphase averaging is valid for modulated antiperiod 2 waves, but it is not complete because it does not describe the evolution of $w(X,T)$ and $y(X,T)$, quantities that can be unambiguously observed on the macroscopic scales and whose modulations, as we have seen in Section 2, can be modeled by partial differential equations. The variables w and y are independent of the Riemann surface Γ as it varies in the class of antiperiod 2 waves. They contain information about, on the one hand the phase constants $\theta_0^{(0)}$ and $\theta_1^{(0)}$ in (53) and (54), and on the other hand the point P_1 and the constant ξ that are involved in the specification of the solution from algebro-geometric data.

Near fixed values of X and T, the evolution of a modulated genus $g = 1$ wavetrain is described by a flow on the torus:

$$(n,t) \mapsto (k_0 n - \omega_0 t - \theta_0^{(0)}, k_1 n - \omega_1 t - \theta_1^{(0)}) . \tag{85}$$

Generally the wavenumbers, frequencies, and phase constants will depend on X and T. However, for general values of the wavenumbers and frequencies, the fast motion will cover the torus, making any macroscopic measurements of the phase constants imprecise, and thus the multiphase averaging theory that leads to (64) does not involve the phase constants.

But in the antiperiod 2 case, the fast motion described by the map (85) does not cover the torus. Instead, the motion only covers two parallel circles, as illustrated in Fig. 8. The location of this pair of circles in the torus is determined by the phase constant $\theta_1^{(0)}$ (modulo π). Macroscopic observation of this quantity is unambiguous, and this phase constant contains information about the quantity w. This situation, where the fast motion in space and time fails to cover the torus and hence makes some phase information observable, is a *classical resonance* that is a direct generalization of the resonance condition for completely integrable systems of ordinary differential equations. In the ODE case, each resonance among the phase variables $\theta_j = \omega_j t - \theta_j^{(0)}$ is characterized by a vector $\boldsymbol{m}$ of integers. The resonance occurs for those values of ω_j for which

$$(\omega_0, \ldots, \omega_g)\boldsymbol{m} = 0 . \tag{86}$$

A classical resonance of a multiphase wavetrain in a spatio-temporal system can be defined analogously. The resonance conditions for the phase variables $\theta_j = k_j n - \omega_j t - \theta_j^{(0)}$ are the condition (86) *and* the spatial resonance condition

$$(k_0, \ldots, k_g)\boldsymbol{m} = 0 , \tag{87}$$

which should hold modulo 2π when the spatial variable is restricted to integer values in spatially discrete systems like the Ablowitz-Ladik equations. For modulated antiperiod 2 waves, the vector $\boldsymbol{m} = (0,2)^T$ characterizes the classical resonance that permits the observation of the quantity w.

Now, consider another integer vector, $\boldsymbol{m}^g = (4,0)^T$. The phase of a modulating antiperiod 2 solution in the direction of this vector (which is tangent to the flow along the two circles in the torus) is given by

$$(\theta_0, \theta_1)\boldsymbol{m}^g = 4k_0 n - 4\omega_0 t - 4\theta_0^{(0)} = 2\pi n - 4t - 4\theta_0^{(0)} = -4t - 4\theta_0^{(0)} \quad \text{modulo } 2\pi , \tag{88}$$

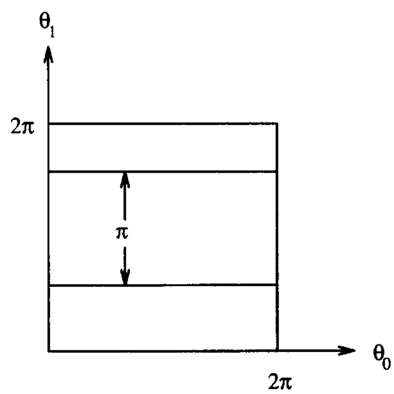

Fig. 8. The subtorus of the antiperiod 2 resonance consists of two parallel circles separated by a distance of π in the θ_1 direction. The vector $\boldsymbol{m}$ points in the θ_1 direction.

The vector $\boldsymbol{m}^g$ does not characterize a classical resonance as described above because the phase in this direction grows linearly with fast time t. However, since the frequency $\omega_0 = 2$ is fixed throughout the modulation, the fast growth of this phase variable can be removed by considering the difference between the values of $(\theta_0, \theta_1)\boldsymbol{m}^g$ at two neighboring lattice points. Scaling by the lattice spacing h yields a quantity $\partial_X(\theta_0^{(0)}, \theta_1^{(0)})\boldsymbol{m}^g$ which can be macroscopically observed. This kind of *generalized resonance* occurs precisely because the frequencies do not modulate. The macroscopic quantity made visible by the generalized resonance characterized by $\boldsymbol{m}^g = (4,0)^T$ is proportional to y.

The modulation theory of antiperiod 2 waves can be characterized as a closed, consistently constrained motion of the four branch points under the equations (64), which drives the slow motion of the phase constants. The above analysis for modulation of antiperiod 2 waves suggests that when the usual modulational prescription breaks down due to resonances, the partial dynamical description (32) and (33) that results may be augmented by appropriate phase equations (51) and (52) which then provide a complete description of the constrained modulated dynamics. It is this structure that we expect to be mirrored in more general cases.

4. Critical wavenumbers and multiphase resonances

Here, we will suggest a general framework that embraces the specific phenomena described in Section 2 and placed in context in Section 3, and allows extension to the modulation theory of multiphase waves in arbitrary integrable spatio-temporal systems. In the theory of multiphase solutions in integrable systems, each solution is associated with a Riemann surface of finite genus, whose branch points are constants of motion that play the role of action variables. In the multiphase averaging theory, the macroscopic description of the genus g solutions is given by a closed system of first-order quasilinear partial differential equations in the slow spatial variable X and the slow temporal variable T for which the branch points z_k of the Riemann surface Γ are Riemann invariants. As we have seen, a fundamental consequence of this theory is that the modulation equations imply *conservation of waves* in each phase variable independently. That is, if $\theta_j = k_j x - \omega_j t$ is one of the phase variables in the modulating multiphase wavetrain, then k_j and ω_j are functions of the branch points that satisfy (69).

It has been pointed out by Forest and Lee [10] that when periodic boundary conditions are enforced on the microscopic oscillations, one should generally expect *no spatial or temporal modulation* in the absence of any external perturbations. Another way of saying this that both generalizes the statement and also avoids reference to any external enforcement of boundary conditions is the following. If at time $T = 0$ the wavenumbers $k_j(X,0)$ of a modulated multiphase wavetrain are independent of X, then generally the wavenumbers will not remain constant for $T > 0$ unless *all* macroscopic quantities are independent of X at $T = 0$.

This is because the requirement that the wavenumbers k_j not evolve in time allows one to deduce from conservation of waves (69) that the frequencies ω_j do not depend on X at $T = 0$. From a generic set of wavenumbers k_j and frequencies ω_j it is possible to solve for the branch points z_k. So, if the initial wavenumbers $k_j(X,0)$ are also taken to be independent of X, then in fact the whole set of branch points z_k must be independent of X at $T = 0$. The modulation equations in Riemann invariant form (70) then imply that no modulation occurs at all.

This argument depends crucially on our ability to solve for the branch points in terms of the frequencies and wavenumbers. As was demonstrated by the example of the modulated antiperiod 2 wavetrains, it is in fact possible to have self-consistent spatio-temporal modulation of a solution with fixed wavenumbers on the microscopic level, as long as the wavenumbers are carefully chosen. This kind of modulation will occur *even in absence of any external perturbation.*

Let us describe the wavenumbers for which nontrivial modulation is possible. In general, the $g+1$ wavenumbers and frequencies are determined from the $2g+2$ branch points

$$k_j = k_j(z_1, \dots, z_{2g+2}), \tag{89}$$

$$\omega_j = \omega_j(z_1, \dots, z_{2g+2}). \tag{90}$$

We introduce $g+1$ additional functions of the branch points

$$\rho_j = \rho_j(z_1, \dots, z_{2g+2}), \tag{91}$$

which we will call *amplitudes*. The purpose of these functions is to allow us to solve for the branch points without involving the frequencies:

$$z_j = z_j(k_1, \dots, k_{g+1}, \rho_1, \dots, \rho_{g+1}). \tag{92}$$

In the $g = 1$ case described in Section 3, the functions $u(z_1, z_2, z_3, z_4)$ and $v(z_1, z_2, z_3, z_4)$ defined by (74) served as amplitudes. Then, using the formulas (90) for the frequencies as functions of the branch points, one obtains the *dispersion relations*

$$\omega_j = \omega_j(z_1, \dots, z_{2g+2}) = f_j(k_1, \dots, k_{g+1}, \rho_1, \dots, \rho_{g+1}). \tag{93}$$

We define a *critical set of wavenumbers* $k_1^c, \dots, k_{g+1}^c$ as a set of wavenumbers for which the frequencies do not depend on the amplitudes:

$$\partial_{\rho_i} f_j(k_1^c, \dots, k_{g+1}^c, \rho_1, \dots, \rho_{g+1}) \equiv 0, \tag{94}$$

for $i, j = 1, \dots, g+1$. This definition is independent of exactly how we chose the amplitude functions $\rho_j(z_1, \dots, z_{2g+2})$. The wavenumbers $k_0 = \pi/2$ and $k_1 = \pi$ are critical for the genus $g = 1$ solutions of the Ablowitz-Ladik equations, because fixing these wavenumbers automatically determines the frequencies $\omega_0 = 2$ and $\omega_1 = 0$ independently of the amplitudes u and v.

If the wavenumbers of a modulated multiphase wavetrain are taken to be spatially constant at $T = 0$, and are taken to be critical, then the frequencies will be spatially constant at $T = 0$ *regardless of the values of the amplitudes.* Conservation of waves (69) then guarantees that the criticality will be preserved, and the amplitudes ρ_j, if initially varying in space, will undergo nontrivial self-consistent modulation. This is why the quantities u and v are able to experience modulation in the case of antiperiod 2 waves in the Ablowitz-Ladik equations. The point is that modulation of waves with critical wavenumbers is a self-consistent modulation of the amplitudes only (and possibly the phase constants, if resonances are present). All wavenumbers and frequencies are independent of the macroscopic independent variables X and T.

Aside from allowing the possibility of self-consistent spatio-temporal modulation, criticality of wavenumbers requires a generalization of the notion of a classical resonance. First, let us describe classical resonance in some detail. Multiphase wavetrains in general spatio-temporal systems are described by functions of space x and time t of the form

$$\psi(x,t) = F(\theta_1, \dots, \theta_N), \tag{95}$$

where F is periodic with period 2π in each of its arguments, and $\theta_j = k_j x - \omega_j t - \theta_j^{(0)}$. This solution is for fixed x (resp. t) a linear flow in t (resp. x) on the N-torus. For generic choices of the wavenumbers k_j and frequencies ω_j, at least one of these two linear flows will densely cover the N-torus; such a solution $\psi(x,t)$ is called *ergodic* or *nonresonant*.

Now, suppose that we are given a set of C linearly independent N-vectors $\boldsymbol{m}_c$, $0 \le C \le N$, with integer components such that

$$\begin{aligned} (k_1,\dots,k_N)\boldsymbol{m}_c &= 0 \\ (\omega_1,\dots,\omega_N)\boldsymbol{m}_c &= 0 \end{aligned} \qquad 0 \le c \le C\,. \tag{96}$$

In this case, we say that the the solution ψ is *resonant* with codimension C. The span (over the rationals) of the vectors $\boldsymbol{m}_c$ completely characterizes the resonance [8]. Resonances are not generic because the components of the vectors $\boldsymbol{m}_c$ must be integers.

In the absence of resonance, the modulational description given by multiphase averaging is complete because it is a closed system of equations for all quantities that can be accurately measured on the macroscopic length and time scales. The phase constants $\theta_j^{(0)}$ are not included in the description, but this is fine since their precise measurement is hindered by the fast winding on the N-torus. This description changes when there is a resonance that is preserved under modulation. In particular, the equations given by the multiphase averaging prescription remain closed, but the description of the macroscopic dynamics offered by these equations will no longer be complete.

The *complete* modulational description of a resonant wavetrain must a constrained special case of the multiphase averaging theory, extended to model the evolution of slow phases, as was done in Section 2 with the derivation of modulation equations for the quantities w and y. The constraints on the wavenumbers and frequencies that cause resonance imply constraints on the branch points of the associated Riemann surface Γ that must be compatible with the modulation equations; otherwise the resonance will not be preserved. In any case, to have a complete description of all macroscopically observable quantities near some X and T where a resonance occurs, the multiphase averaging equations must be augmented because the solution ψ no longer covers the N-torus as it evolves in space and time; it only covers a subtorus (or in the discrete case a finite family of subtori) of codimension C. The location of this subtorus within the N-torus becomes macroscopically observable information. This information is the vector of slow phases $(\theta_1^{(0)},\dots,\theta_N^{(0)})\boldsymbol{m}_c$. Other equations must be adjoined to the modulation equations to describe these slow phases.

If the wavenumbers k_j are critical, then two new features appear that change the way that resonances affect macroscopic dynamics. First, and most importantly, we have seen from conservation of waves (69), that criticality of wavenumbers is preserved under modulation. Thus, unlike ordinary resonances, those that occur for critical wavenumbers (and their corresponding fixed frequencies) *will persist under modulation*. This means that the presence of slow phases for these cases is a permanent feature of the macroscopic theory, and not merely transient as the action variables pass near a resonance in the course of their evolution. Resonances that follow from critical wavenumbers are permanent, and modeling the slow phases is *essential* for a complete description of the dynamics.

Furthermore, if the wavenumbers are critical it is not necessary to have both conditions (96) satisfied in order for slow phase information to be macroscopically observable. This will lead to the idea of a *generalized resonance*. Suppose that only the spatial resonance condition is satisfied by some integer vector $\boldsymbol{m}^g$:

$$(k_1^c,\dots,k_N^c)\boldsymbol{m}^g = 0\,, \tag{97}$$

but that the temporal resonance condition is not satisfied. The phase in the spatially resonant direction $\boldsymbol{m}^g$ is

[8] The example dispersive wave equation we have addressed in this paper is discrete in space. In such discrete cases, we replace x by n, an integer; the condition $(k_1,\dots,k_N)\boldsymbol{m}_c = 0$ should then hold modulo 2π.

$$(\theta_1,\ldots,\theta_N)\boldsymbol{m}^g = -(\omega_1,\ldots,\omega_N)\boldsymbol{m}^g t + (\theta_1^{(0)},\ldots,\theta_N^{(0)})\boldsymbol{m}^g . \tag{98}$$

In the context of a general modulation, this quantity will be slowly varying in space, and rapidly varying (linearly) in time. However, the criticality of the wavenumbers guarantees that the frequencies do not undergo any spatial or temporal modulation – they do not depend on slow space or time. Thus, we can eliminate the fast dependence on time by differentiating with respect to the slow spatial variable X. For critical wavenumbers satisfying (97), the quantity

$$\partial_X(\theta_1,\ldots,\theta_N)\boldsymbol{m}^g = \partial_X(\theta_1^{(0)},\ldots,\theta_N^{(0)})\boldsymbol{m}^g \tag{99}$$

is a function of slow space X and time T that must be modeled in order to provide a complete modulational description. In this way, the notion of resonance is generalized when the wavenumbers are critical.

Although we have presented only one example of multiphase modulation theory with critical wavenumbers, we believe that other examples must surely exist. We plan to study the microscopic behavior that locally develops when the antiperiod 2 modulation equations break down due to the appearance of hyperbolic shocks in order to look for other cases of modulated wavetrains with critical wavenumbers.

5. Discussion

In this paper, we have started an investigation of some of the phenomena associated with resonances in spatio-temporal systems. Aided by the criticality of the wavenumbers in the antiperiod 2 case, we were able to explore phase modulation in a convenient mathematical laboratory where the resonance of interest was preserved by the macroscopic dynamics. We found that the multiphase averaging prescription continued to hold in the presence of resonance, but became incomplete as it did not model the slow evolution of the phase constants. In this case, we were able to supplement the modulation equations with PDEs describing these phase constants by a straightforward asymptotic analysis of the microscopic system.

However, we have avoided the truly difficult problems associated with generalizing the KAM theory of ordinary differential equations to the infinite dimensional context. Multiphase averaging in integrable wave systems is a generalization of the averaging theory used to handle small perturbations of integrable Hamiltonian systems [2]. It is known that in such systems with one degree of freedom, the averaging prescription always gives the correct description of the perturbed dynamics to first-order in the perturbation parameter. However, when the unperturbed system has two or more degrees of freedom, resonances can arise that require one to use techniques of partial averaging in order to avoid the small denominators that spoil the ordering of terms in the asymptotic expansion of the solution. Unfortunately, exact resonances are dense in the phase space of the unperturbed system; however, as long as one avoids parts of phase space near serious resonances (that is, those corresponding to vanishing rational combinations of frequencies, with coefficients having small integer denominators), the KAM theory predicts that the full averaging prescription should be locally valid.

The corresponding problems of resonance looming on the horizon of the multiphase averaging theory are both more complicated and less understood. Some of the problems have to do with finding the regions of validity in phase space (if any) of the multiphase averaging theory, which assumes that the multiple phases of the modulated solution are independent over the rationals. Although dense, the resonances form a set of measure zero in the phase space, and one would like to think that it is possible to ignore them most of the time, as was guaranteed in finite dimensional problems by the KAM theory. The problem is that in the infinite dimensional context, it is harder to avoid the resonances. When a spatially modulated wavetrain (wavenumbers and frequencies varying smoothly) is given as an initial condition for a dispersive wave system, resonances will be densely distributed in physical space, and not just in phase space. In an attempt to justify the nonresonant modulation theory obtained from multiphase averaging, Chierchia, Ercolani, and McLaughlin have given a KAM-like estimate of the allowable resonances in wavenumber distributions that are functions of physical space by defining a "local nonresonance" condition that is generic in an appropriate sense [4]. Dobrokhotov, on the other hand, reasonably prefers a derivation of modulation equations in which one does not have to make such restrictions on the initial data [5]. He takes the point of view that the density of resonances in space causes the nonresonant multiphase averaging theory to fail in all genuinely multiphase cases. In his words, "Since these [resonant] points are placed in the [physical] space, it is impossible to throw them away together with their small neighborhoods as was done in the KAM-theory."

In spite of these arguments, the nonresonant modulation equations of multiphase averaging seem to be valid in several contexts where no particular nonresonance constraints are imposed on initial conditions. Indeed, the modulation equations derived formally in [9] were simultaneously derived rigorously by Lax and Levermore [16,17] as describing the zero-dispersion limit, generally a weak limit, of solutions of the Korteweg-deVries equation. This was done by analyzing the associated inverse spectral problem rather than by averaging explicit multiphase solutions, and considered a broad class of initial data that certainly includes functions that have resonances densely occupying physical space. Similar results have been found in other integrable systems for which a zero-dispersion limit makes sense (see [14] for the defocusing cubic nonlinear Schrödinger equation [8], for the focusing modified Korteweg-de Vries equation, and [20] for the Toda lattice). This theory and its applications have been recently reviewed in [18]. The reason that the multiphase averaging theory *appears* to have trouble with resonances is that it derives the modulation equations from a local N-phase ansatz for the microstructure of the oscillations, and the validity of the ansatz itself is called into question near a resonance. On the other hand, the Lax-Levermore theory avoids making any assumptions about the microstructure of the solution, and thereby manages to characterize the limit globally in terms of quantities that can be interpreted as well-defined parameters of oscillation either through the higher-order theory of Venakides [24] or when combined with a consistent local multiphase ansatz. We remark that the theory of Venakides captures the microstructure of the multiphase wavetrain, but does not model the evolution of the phase constants, as those dynamics appear at yet higher order. Finally, note that the problem of *formally* deriving modulation equations in resonant cases disappears if one considers not torus averages as in [9], but space-time averages, as is done in [21] for the Ablowitz-Ladik equations. From these considerations, it is not at all surprising that the modulation equations hold in spite of resonance, since they may be derived in several ways without any reference to multiphase torus averages.

Other problems have to do with dealing with resonances rather than learning when it is possible to ignore them. In the perturbation theory of ODEs, the solution to the problem of constructing the first-order correction to a resonance is given by the prescription of partial averaging, which describes temporal modulation of phase information made visible by resonance, as well as modulation of action variables. We have been able to describe phase modulation in a spatio-temporal context, but our derivation did not make use of any of the beautiful geometrical structure of the integrable system we were studying. We would like to understand how to lift our derivation to the level of the geometrical data $(\Gamma,\eta;P_1,\ldots,P_g,\xi)$ used to describe the multiphase wavetrains. It is at this level that the modulation equations obtained from multiphase averaging take their most elegant form, involving deformations of an invariant object $\Gamma(X,T)$. We expect that an equally elegant prescription must exist, certainly generalized from the partial averaging idea for ODEs, that describes modulations of resonant combinations of the points $P_1,\ldots,P_g$ and the scaling constant ξ (which determine the phase constants).

Let us conclude with a few observations. First, we want to point out that the phenomenon of algebraic constraints being consistent with modulation equations (as is the case for critical wavenumbers) has been observed in other models. For example, in a neighborhood of the initial location of a shock in the Toda lattice, the long time behavior is given by a degenerate case of modulated genus $g = 2$ oscillations, the degeneracy being described by two algebraic conditions on the six branch points. This degeneracy is required in order to allow a globally odd shock solution in which the particle at the origin is fixed for all time. In this way, the degeneracy constraints effectively introduce boundary conditions on a semi-infinite chain, much as the constraints (73) introduce antiperiod 2 boundary conditions in the Ablowitz-Ladik equations. Bloch and Kodama [3] found in the Toda shock problem that the full genus $g = 2$ modulation equations respect these constraints. In that problem the constraints persist because they amount to simply fixing the values of two of the branch points, and it is easy to see from the modulation equations in Riemann invariant form (70) that such constraints are always preserved under modulation. In another example, Ercolani, Forest, McLaughlin, and Sinha study the perturbed sine-Gordon equation with periodic boundary conditions imposed on the microscopic level [7]. As these wavenumbers are not critical, there is no spatial modulation at all, and temporal modulations are only present due to the presence of non-autonomous perturbations. The spatial periodicity of the perturbation matches that of the modulating wavetrain, and the modulation PDEs are consistently constrained to become ordinary differential equations.

Finally, we should point out that the modulational stability of the antiperiod 2 solutions in the focusing case surprised us at first. This was because in the focusing case there is no spectral symmetry argument for the modulation equations that guarantees the reality of the characteristic velocities (unlike the defocusing case when $|Q_n| < 1$ for all n, where the velocities must be real). Indeed, one lesson to be learned from the example of antiperiod 2 waves is that there may be modulationally stable solutions of integrable problems even in the absence of any spectral symmetry to guarantee reality of the characteristic velocities. The facts for the Ablowitz-Ladik equations are (see [21]):

- the branch points of focusing solutions cannot lie on the unit circle in the z-plane without being double, and
- if a branch point z_k lies on the unit circle, then its corresponding characteristic velocity is real.

It simply does not follow that focusing configurations cannot have real characteristic velocities. Now, it is known that the focusing modulation equations are elliptic for $g = 0$ [12], but aside from the antiperiod 2 waves which we have seen to be stable, more complicated focusing solutions have not been specifically analyzed for their stability properties. It would seem that stability of focusing waves must be determined on a case by case basis. If there are other stable wavetrains in the focusing case, perhaps they will appear when the antiperiod 2 modulation equations develop singularities. Thus, we plan to investigate the local behavior in the focusing problem near points X where infinite derivatives appear in the fields u, v, w, and y for antiperiod 2 oscillations. We would like to know if other microscopic oscillations can spontaneously appear that also may be smoothly modulated.

Acknowledgements

The authors gratefully acknowledge the support provided to them: P.D. Miller and C.D. Levermore were supported by the Air Force under grant AFOSR-90-021 and the NSF under grant DMS-9404570; N.M. Ercolani was supported by the Air Force under grant AFOSR-FQ8671-9000589 and the NSF under grant DMS-8942390. P.D. Miller was also supported in part by the Australian Photonics Cooperative Research Centre.

References

[1] M.J. Ablowitz and J.F. Ladik, Nonlinear differential-difference equations and Fourier analysis, J. Math. Phys. 17 (1976) 1011–1018.

[2] V.I. Arnold, V.V. Kozlov and A.I. Neishtadt, Mathematical aspects of classical and celestial mechanics, in: V.I. Arnold, ed., Dynamical Systems III, Volume 3 of Encyclopaedia of Mathematical Sciences (Springer Verlag, Berlin, 1988).

[3] A.M. Bloch and Y. Kodama, Dispersive regularization of the Whitham equation for the Toda lattice, SIAM J. Appl. Math. (4) 52 (1992) 909-928.

[4] L. Chierchia, N. Ercolani and D. McLaughlin, On the weak limit of rapidly oscillating waves, Duke Math. J. (4) 55 (1987) 759-764.

[5] S.Yu. Dobrokhotov, Resonances in multifrequency averaging theory, in: N. Ercolani, I. Gabitov, D. Levermore and D. Serre, eds., Singular Limits of Dispersive Waves, Plenum NATO ARW series, New York (1994) pp. 203-217.

[6] B.A. Dubrovin, Theta functions and non-linear equations, Usp. Mat. Nauk (2) 36 (1981) 11-80.

[7] N.M. Ercolani, M.G. Forest, D.W. McLaughlin and A. Sinha, Strongly nonlinear modal equations for nearly integrable pdes, J. Nonlin. Sci. 3 (1993) 393-426.

[8] N.M. Ercolani, S. Jin, C.D. Levermore and W. MacEvoy, The zero dispersion limit of the NLS/mKdV hierarchy for the nonselfadjoint ZS operator, University of Arizona preprint (1992).

[9] H. Flaschka, M.G. Forest and D.W. McLaughlin, Multiphase averaging and the inverse spectral solution of the Korteweg-de Vries equation, Commun. Pure Appl. Math. 33 (1980) 739-784.

[10] M.G. Forest and J.E. Lee, Geometry and modulation theory for the periodic nonlinear Schrödinger equation, in: C. Dafermos, J.L. Erickson, D. Kinderleher and M. Slemrod, eds., Oscillation Theory, Computation, and Methods of Compensated Compactness, Volume 3 of IMA Volumes on Mathematics and its Applications (Springer Verlag, New York, 1986) pp. 35-69.

[11] M.H. Hays, Classical and Quantum Mechanical Studies of Nonlinear Lattices, PhD thesis, University of Arizona (1995).

[12] M.H. Hays, C.D. Levermore and P.D. Miller, Macroscopic lattice dynamics, Physica D 79 (1994) 1-15.

[13] M.H. Hays and Yixia Lu, private communication.

[14] Shan Jin, C. David Levermore and David W. McLaughlin, The behavior of solutions of the NLS equation in the semiclassical limit, in: N. Ercolani, I. Gabitov, D. Levermore and D. Serre, eds., Singular Limits of Dispersive Waves, Plenum NATO ARW series, New York (1994) pp. 235-255.

[15] I.M. Krichever, Methods of algebraic geometry in the theory of nonlinear equations, Russian Math. Surveys (6) 32 (1977) 185-213.

[16] P.D. Lax and C.D. Levermore, The zero dispersion limit for KdV equation, Proc. Natl. Acad. Sci. USA (1979) pp. 3602-3606.

[17] P.D. Lax and C.D. Levermore, The small dispersion limit of the Korteweg-de Vries equation I, II, III, Commun. Pure Appl. Math., 36 (1983) 253-290, 571-593, 809-829.

[18] P.D. Lax, C.D. Levermore and S. Venakides, The generation and propagation of oscillations in dispersive initial value problems and their limiting behavior, in: A.S. Fokas and V.E. Zakharov, eds., Important Developments in Soliton Theory (Springer Verlag, Berlin, 1993) pp. 205-241.

[19] C.D. Levermore and J.G. Liu, Large oscillations arising in a dispersive numerical scheme, Submitted to Physica D (1995).

[20] K. McLaughlin, PhD thesis, Courant Institute of Mathematical Sciences, NYU (1994).

[21] P.D. Miller, N.M. Ercolani, I.M. Krichever and C.D. Levermore, Finite genus solutions to the Ablowitz-Ladik equations, to appear in Comm. Pure Appl. Math.

[22] A. Newell, T. Passot, N. Ercolani and R. Indik, Elementary and composite defects of striped patterns, to appear in J. Phys. (Paris).

[23] Colin J. Thompson, Classical Equilibrium Statistical Mechanics (Oxford University Press, Oxford, 1988).

[24] S. Venakides, Higher order Lax-Levermore theory, Commun. Pure Appl. Math. 43 (1990) 335-362.

[25] G. Whitham, Linear and Nonlinear Waves (Wiley, New York, 1974).