# PHYSICA D

DTIC
ELECTE
APR 1 3 1994
S
C
D

# NONLINEAR PHENOMENA

## Future Directions of Nonlinear Dynamics in Physical and Biological Systems

*in honor of Alwyn C. Scott*

A selection of articles presented at the
conference honoring Alwyn C. Scott
on the occasion of his 60th birthday
The Technical University of Denmark, Lyngby, Denmark
23 July – 1 August 1992

Editors:

P.L. Christiansen
J.C. Eilbeck
R.D. Parmentier

94-11246

DTIC QUALITY ... 3

94  4  12  173

# PHYSICA D
# NONLINEAR PHENOMENA

| Accesion For | | |
|---|---|---|
| NTIS CRA&I | ☑ | |
| DTIC TAB | ☐ | |
| Unannounced | ☐ | |
| Justification | | |
| By | | |
| Distribution / | | |
| Availability Codes | | |
| Dist | Avail and / or Special | |
| A-1 | | |

## NORTH-HOLLAND

# Future Directions
# of Nonlinear Dynamics
# in Physical and Biological Systems

*in honor of Alwyn C. Scott*

A selection of articles presented at the
conference honoring Alwyn C. Scott
on the occasion of his 60th birthday
The Technical University of Denmark, Lyngby, Denmark
23 July – 1 August 1992

Editors:

## P.L. Christiansen

*Laboratory of Applied Mathematical Physics*
*The Technical University of Denmark*
*DK-2800 Lyngby, Denmark*

## J.C. Eilbeck

*Department of Mathematics*
*Heriot-Watt University*
*Edinburgh EH15 4AS, UK*

## R.D. Parmentier

*Department of Physics*
*University of Salerno*
*I-84081 Baronissi SA, Italy*

N·H

1993

# NORTH-HOLLAND

# Preface

Early in 1990 a scientific committee was formed for the purpose of organizing a high-level scientific meeting on Future Directions of Nonlinear Dynamics in Physical and Biological Systems, in honor of Alwyn Scott's 60th birthday (December 25, 1991). As preparations for the meeting proceeded, they were met with an unusually broad-scale and high level of enthusiasm on the part of the international nonlinear science community, resulting in a participation by 168 scientists from 23 different countries in the conference, which was held July 23 to August 1 1992 at the Laboratory of Applied Mathematical Physics and the Center for Modelling, Nonlinear Dynamics and Irreversible Thermodynamics (MIDIT) of the Technical University of Denmark. During the meeting about 50 lectures and 100 posters were presented in 9 working days.

The contributions to this present volume have been grouped into the following chapters:

1. Solitons
2. Superconducting devices
3. Biomolecular dynamics
4. Nonlinear optics
5. Classical and quantum mechanical lattice dynamics,

reflecting some of the many different areas of nonlinear science to which Al Scott has made early and fundamental contributions. They form a subset of the total number of the papers and posters presented at the meeting, judged by the editors to be most representative of a cross-section of Al's own research. Other papers from the meeting, which cover an even broader spectrum, have been selected for publication in "Future Directions of Nonlinear Dynamics in Physical and Biological Systems", NATO Advanced Study Institute Series, to appear with Plenum Press.

People who have worked with Al know that he insists that everyone should do what he or she likes to do; this is certainly true, but it is also true that from the very early days he has demonstrated a unique ability to stimulate research in new directions. Moreover, his ability to interact with people and to transmit his optimism and enthusiasm to those around him has furnished a solid launching pad for these new lines of research. Many friends and colleagues around the world are thankful to Al for having received such a stimulus. This was the essential motivation for our meeting and for the publication of these Proceedings. It is a small way of saying "Thank you, Al, and many happy returns".

Peter Christiansen
Chris Eilbeck
Bob Parmentier

*Editors*

# Contents

## 3b. *Davydov solitons*

## 3c. *DNA*

## Chapter 4. Nonlinear optics

## Chapter 5. Classical and quantum mechanical lattice dynamics

PHYSICA D

# Solitons on lattices

D.B. Duncan, J.C. Eilbeck, H. Feddersen and J.A.D. Wattis

*Department of Mathematics, Heriot-Watt University, Edinburgh, EH14 4AS, UK*

We examine a variety of numerical and approximate analytical methods to study families of solitary waves on lattices. Such waves, when they exist, travel through the lattice without loss of energy, and have approximate soliton properties on collision. Corresponding quantum problems are also briefly described.

## 1. Introduction

The study of solitons on discrete lattices date back to the early days of soliton theory. The famous paper by Fermi, Pasta and Ulam (FPU) which inspired the pioneering work on the KdV problem by Kruskal and Zabusky (cf. [12] for references) was a numerical study of a crystal lattice model with an anharmonic interatomic potential. The Hamiltonian of such a system could be written as

$$H = \sum_n [\tfrac{1}{2}\dot{q}_n^2 + V(q_{n+1} - q_n)], \qquad (1.1)$$

where $q_n$ is the displacement of the $n$th particle from its equilibrium position, and $V(q_{n+1} - q_n)$ is the anharmonic interaction potential. In the FPU study, $V'(q) = \gamma q + \alpha q^m$, $m = 2$ or 3. If the relative displacement of the $n$th bond is defined to be $\phi_n = q_{n+1} - q_n$, then the equation of motion becomes

$$\frac{\mathrm{d}^2\phi_n}{\mathrm{d}t^2} = V'(\phi_{n+1}) - 2V'(\phi_n) + V'(\phi_{n-1}). \qquad (1.2)$$

A continuum approximation to the FPU problem leads to the KdV equation ($m = 2$) or the MKdV equation ($m = 3$), (see [12] for details). The analytic solution of such equations through the inverse spectral transform, and the

associated set of exact soliton properties, is now well known.

For the discrete problem, Toda [31] showed that the special choice of $V(\phi) = a[1 - \exp(-b\phi)]$ let to a set of equations which also had exact multi-soliton solutions. Ablowitz and Ladik [1] discovered another integrable lattice which was a discrete version of the Nonlinear Schrödinger equation, and a number of other lattice equations, either continuous or discrete in time, have been shown to have this property. However the basic model investigated by FPU is not integrable, and unfortunately little theory is possible once this property is lost.

In this paper we survey some recent investigations of nonintegrable lattice problems which have soliton or soliton-like solutions (we will use the word "soliton" in the loose sense as a localized travelling wave which survives collision with other such objects unchanged, or almost unchanged). Such studies are by necessity approximate, although we present both numerical and analytic approximations (here by "analytic approximation" we mean a simple analytic formula with parameters derived by numerically or analytic means).

There are two basic numerical approaches we have used in our study. One is to look for travelling waves, i.e., solutions of the form $\phi_n(t) = \phi(n - ct) = \phi(z)$, with $\phi(z) \to 0$ as $|z| \to \infty$, which leads to the equation
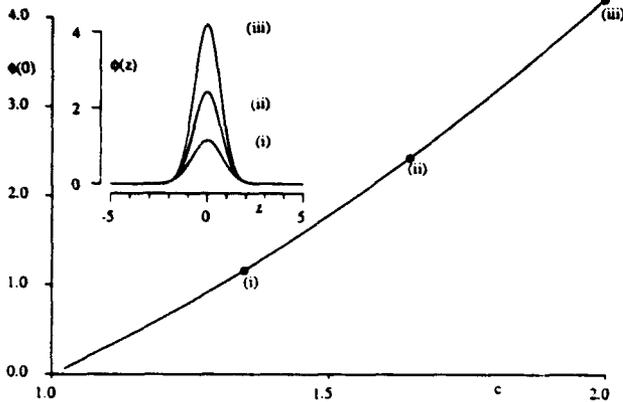
Fig. 1. Travelling waves for the case $V'(q) = q + q^2$.

$$c^2\phi''(z) = V'(\phi(z+1)) - 2V'(\phi(z))$$
$$+ V'(\phi(z-1)). \tag{1.3}$$

We solve this equation using pseudo-spectral methods [17]. By treating $c$ as a parameter we can use path-following methods to trace out the whole curve of solutions as $c$ varies.

In order to study the stability or interactions of such solitary wave solutions, or to study more arbitrary initial conditions, we have investigated a class of symplectic solvers for the initial value problem (1.2) [15]. Some typical results are described in section 2.

Another alternative approach is to find analytic approximations to travelling waves, either by continuum approximations to (1.3) at various levels of approximation, i.e., low $c$ expansions, or expansions in small values of $1/c$, or by the use of variational or weak formulations to fix parameters in some assumed formula for the travelling wave. This approach is described in section 3.

## 2. Numerical results

Results for the potential $V'(q) = q + 2q^3$ are described in [17]. As another example we show in fig. 1 some results for the potential $V'(q) = q + q^2$. Here the main graph shows the plot of maximum height against the velocity $c$, for $c$ in the range $1 \leqslant c \leqslant 2$. The inset shows the corresponding travelling wave profiles for three points on the curve. We see that the energy of the wave is concentrated on the three most central grid points. Numerical studies with these profiles as initial conditions shows the pulses are stable, at least for medium time scales. We mention in passing that a similar calculation can be used to investigate periodic solutions of (1.3) or indeed solitary wave or periodic solutions of continuous problems.

A related equation where the nonlinearity is in the derivative term, arising in an electrical lattice problem, is discussed in [19]. In this case the curvature of the apex of the solitary wave appears to blow up at a finite value of the amplitude, i.e., the tip develops a cusp. Unfortunately this corresponds to a parameter value for which the model is no longer physical. It would be interesting to find this effect in a physically realisable model.

### 2.1. The discrete sine-Gordon equation

If we try to carry out the same calculation for the discrete sine-Gordon (DSG) equation

$$\ddot{\phi}_n = \phi_{n+1} - 2\phi_n + \phi_{n+1} - \Gamma^2 \sin\phi_n, \tag{2.1}$$

we were unable to find travelling waves with the property that $\phi(z) \to 0 \pmod{2\pi}$ as $|z| \to \infty$. The best we could find is shown in fig. 2, where we have cut out the central part of the kink in order to highlight the behaviour of the tails. It can be seen that the tails appear to be periodic waves with a small constant amplitude. The calculation is set up in such a way that the tails travel at the same velocity as the main kink. We hope to investigate the effect of varying the background wave velocity independently of the kink velocity in the future. It seems that the medium must be "pumped" with background radiation before the kink will travel in a lossless manner. This finding is consistent with the results of Peyrard and Kruskal [26], who showed that a kink in
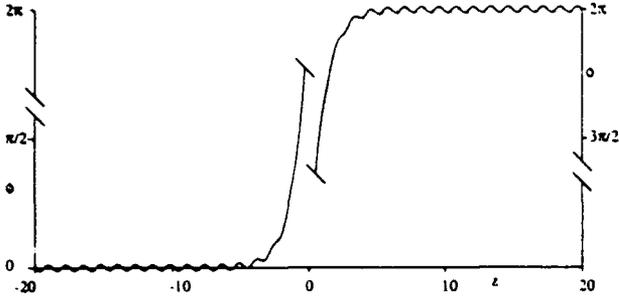
Fig. 2. Travelling kink for the DSG equation, $c = 0.316228$, $\Gamma = 1$.



Fig. 3. Discrete sine-Gordon breather bifurcating from the simple solution (2.3) for $L = 30$ and $\omega = 0.9$. $A = \max f(n, t)$.

a medium with no initial background radiation loses energy as it propagates. If the velocity of the combined wave in fig. 2 is varied, the size of the background radiation increases – the example we have chosen corresponds to a minimum in this amplitude.

*Discrete SG breathers.* Instead of looking for travelling kink solutions of the DSG equation (2.1), we can look for *stationary* breather type solutions of the form $\phi_n(t) = 4 \arctan f(n, t)$. We assume periodic boundary conditions, $f(n, t) = f(n + L, t)$ and consider solutions which are even in both space and time, i.e., $f$ can be expanded as a double Fourier series

$$f(n, t) = \sum_i \sum_j a_{ij} \cos \frac{2\pi i n}{L} \cos j\omega t. \qquad (2.2)$$

One simple solution of this type, which is constant in $n$, is given by [25]

$$f(n, t) = A \operatorname{cd}(\Omega t \mid m), \qquad (2.3)$$

where cd is one of the Jacobian elliptic functions [3] and where $\Omega = \Gamma/(1 + A^2)$ and $m = A^4$.

Now we find a bifurcation diagram where the localised wave, in this case a breather, bifurcates from a simple solution, see fig. 3. The numerical path-following suggests that breather solutions only exist when $\Gamma$ is sufficiently small. By using an accurate and efficient symplectic integration scheme [5,15] on the numerical output from the path-following code it is confirmed that

the breather appears to be stable and does not lose energy through radiation.

## 2.2. The discrete nonlinear Schrödinger equation

The discrete NLS equation appears in several applications in condensed matter physics [4,11,34] and in nonlinear optics [8]. We consider the nonintegrable discrete NLS equation

$$i\frac{d\phi_n}{dt} + \gamma|\phi_n|^2\phi_n + \phi_{n+1} + \phi_{n-1} = 0 \qquad (2.4)$$

with periodic boundary conditions $\phi_{n+L} = \phi_n$, where $L$ is the number of lattice points. Hence, all the solutions we find are periodic with period $L$. For large $L$ we can expect to find good approximations to solitary waves which can be regarded as periodic waves in the lim·· of infinite period. We will look for travelling waves of the form of an envelope travelling wave modulating a carrier wave travelling at a different speed, i.e.,

$$\phi_n(t) = \phi(n - ct)\, e^{i(kn - \omega t)}$$
$$= \phi(z)\, e^{i(kn - \omega t)}, \qquad (2.5)$$

where $c$ is the speed of the envelope of the travelling wave. The periodic boundary condition $\phi(z + L) = \phi(z)$ requires $k$ to be of the form $k = 2\pi m/L$ where $m$ is an integer. With the ansatz (2.5) inserted into (2.4) we find that $\phi(z)$ must satisfy the complex nonlinear differential-advance-delay equation
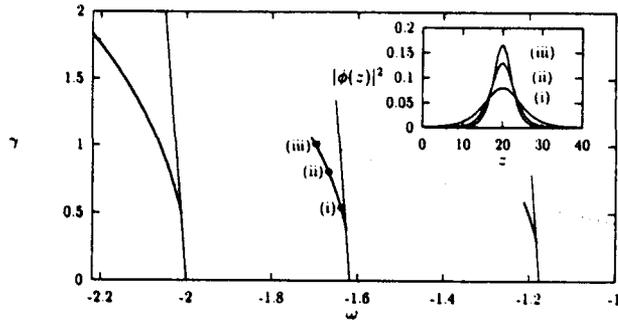
Fig. 4. Paths of travelling waves in the discrete NLS equation (2.4) for $L = 40$ and, from the left, (a) $c = 0.01$, (b) $c = 1.2$, (c) $c = 1.6$. Localized ("solitary") waves bifurcate from stationary solutions (straight lines), but only seem to exist when $\gamma$ is sufficiently small as indicated by the dotted line. The inset shows waveforms along the $c = 1.2$ path.

$$-ic\phi'(z) + \omega\phi(z) + \gamma|\phi(z)|^2\phi(z)$$
$$+ e^{ik}\phi(z+1) + e^{-ik}\phi(z-1)$$
$$= 0. \qquad (2.6)$$

Apart from the trivial solution $\phi(z) \equiv 0$ this equation has the obvious *stationary* solution [18], $\phi(z) = \Phi$ where $\Phi$ is a (complex) constant. The dispersion relation for the normalised (i.e., $\sum_{n=1}^{L} |\phi_n|^2 = 1$) stationary solutions is easily found as

$$\omega = -\frac{\gamma}{L} - 2\cos k. \qquad (2.7)$$

The stationary solutions are interesting in this context because localised travelling ("solitary") waves bifurcate from them. The bifurcation points can be found analytically using standard bifurcation theory [21].

The travelling waves which bifurcate from the stationary solutions are calculated numerically using the path-following method [17]. As we follow the path away from the bifurcation point, i.e., for increasing nonlinearity, we see how the travelling wave becomes more and more localised (cf. fig 4) until (for $c \neq 0$) at some point the path stops.

Until the path stops the bifurcation diagram is very similar to what is found in the continuum limit [35]. However, in the discrete case we also find *dark* solitary waves for positive $\gamma$. In the

continuum limit dark solitary waves only exist for negative $\gamma$ [2,23].

If we consider the discrete NLS equation we can expect to find dark solitary waves for negative $\gamma$ (just as we found bright solitary waves for positive $\gamma$). Assume that $\phi_n(t) = \phi(z)\exp[i(kn - \omega t)]$ is one such dark solitary wave for $\gamma = -\gamma_0 < 0$. Then it is easily shown by insertion that the dark solitary wave

$$\tilde{\phi}_n(t) = \phi(-z)$$
$$\times \exp\left(i\{[(2q+1)\pi - k]n + \omega t\}\right), \qquad (2.8)$$

where $q$ is an integer, satisfies the discrete NLS equation (2.4) with $\gamma = +\gamma_0$.

### 2.3. Davydov's equations

Davydov's equations have been proposed as a model for propagation of vibrational energy in the form of solitary waves in $\alpha$-helix protein and take the form [10,11]

$$i\hbar\frac{d\phi_n}{dt} = \chi(\beta_{n+1} - \beta_n)\phi_n - J(\phi_{n+1} + \phi_{n-1}), \qquad (2.9)$$

$$M\frac{d^2\beta_n}{dt^2} = w(\beta_{n+1} - 2\beta_n + \beta_{n-1})$$
$$+ \chi(|\phi_n|^2 - |\phi_{n-1}|^2). \qquad (2.10)$$

Here $|\phi_n|^2$ is the probability of finding a quantum of amide-I vibrational energy at site $n$ and $\beta_n$ is the longitudinal displacement of the $n$th peptide group of the protein. $\chi$, $J$, $M$ and $w$ are physical constants for which we will use the estimates [30] $w = 50$ N/m, $M = 5.7 \times 10^{-25}$ kg and $J = 1.55 \times 10^{-22}$ J. At present it is unclear what value $\chi$ should take [7] although experiments [6,7] suggest $\chi = 35$–62 pN.

Note that under the "adiabatic" approximation, $d^2\beta_n/dt^2 = 0$, eq. (2.10) is satisfied for $\beta_{n+1} - \beta_n = -\chi|\phi_n|^2/w$ and eq. (2.9) reduces to the discrete NLS equation (2.4) with $\gamma = \chi^2/Jw$ (and $t$ replaced by the scaled time $Jt/\hbar$).

With the above choice of parameters in Davydov's equations (2.9), (2.10) we find a bifur-

cation diagram very similar to that found for the discrete NLS equation [22]; stable solitary waves bifurcate from stationary solutions and the paths of solitary waves stop at some point as $\chi$ is increased. For small values of $c$ this happens for $\chi \approx 120$ pN. We also find dark solitary waves.

## 2.4. 2D lattices

Some results on a generalization of (1.2) to isotropic 2D lattices is given in [19]. An interesting anisotropic lattice, proposed by Zakharov, which reduces to the KP equation in the continuum limit, is discussed in [14].

## 3. Analytic approximations

In this section we outline various analytic approximations to solitary wave solutions of the lattice equation (1.2). We start with the travelling wave equation (1.3) obtined from (1.2) by the travelling wave ansatz. For illustration, we consider the cubic polynomial potential

$$V(\phi) = \tfrac{1}{2}\phi^2 + \tfrac{1}{3}a\phi^3, \tag{3.1}$$

where $a > 0$ is a constant. We examine continuum-like approximations derived from a Fourier transformed version of eq. (1.3) and valid in various wave speed ranges and one method valid in the high speed, large amplitude limit. We also describe methods based on variational and weak formulations of the travelling wave equation (1.3) which are valid at all wave speeds.

## 3.1. Continuum approximations

We start by deriving various continuum approximations from a Fourier transformed version of (1.3) (see [24] for example):

$$c^2 \widehat{\phi}(k) = \left( \frac{4\sin^2(\tfrac{1}{2}k)}{k^2} \right) \widehat{F}(k)$$

$$\stackrel{\text{def}}{=} \widehat{A}(k)\widehat{F}(k), \tag{3.2}$$

where $\widehat{g}(k)$ represents the Fourier transform of $g(z)$ and we use the abbreviated notation $F(z)$ for $V'(\phi(z))$. Inserting Taylor series or Padé approximations of the transformed operator $\widehat{A}(k)$ in the equation above, we then invert the transform to obtain standard nonlinear ODEs for the solution $\phi$ of (1.3).

For example, using the (0,2) Padé approximation $1/(1 + \tfrac{1}{12}k^2)$ of $\widehat{A}(k)$ generates the second order ODE

$$c^2(1 - \tfrac{1}{12}D^2)\phi = V'(\phi),$$

where the operator $D = d/dz$. This has as a solution

$$\phi(z) = \frac{3}{2a}(c^2 - 1)$$

$$\times \operatorname{sech}^2 \left( z\sqrt{\frac{3(c^2 - 1)}{c^2}} \right) \tag{3.3}$$

using the cubic potential (3.1). This ODE is also obtained via an approximation of the original lattice equation (1.2) by the regularised or improved Boussinesq equation as in [28,29] and by a very different approach in [9].

The most accurate approximation of $\widehat{A}(k)$ which leads to a tractible ODE is the (2,2) Padé approximation

$$\widehat{A}(k) \approx \frac{1 - \tfrac{1}{20}k^2}{1 + \tfrac{1}{30}k^2},$$

which, after substitution in (3.2), multiplication by $1 + \tfrac{1}{30}k^2$ and inversion of the transform, gives the ODE

$$c^2(1 - \tfrac{1}{30}D^2)\phi = (1 + \tfrac{1}{20}D^2)V'(\phi),$$

where the differential operator $D = d/dz$. An implicit solution of the ODE can be found and for the cubic potential (3.1) we obtain the speed–height relationship

Height $= \phi(0)$

$$= \frac{1}{9a}[4c^2 - 9 + c\sqrt{5(14c^2 - 9)}]$$

for the travelling wave. The solution $\phi$ is found from

$$\pm z = \frac{2}{\sqrt{5}}\tan^{-1}\Theta$$

$$- \frac{(2c^2 + 3)}{6\sqrt{5a\phi(0)\beta}}\log\left(\frac{1 - \Theta\sqrt{\beta/a\phi(0)}}{1 + \Theta\sqrt{\beta/a\phi(0)}}\right),$$

(3.4)

where the constant $\beta$ and the function $\Theta$ are defined by

$$\beta = \tfrac{1}{9}[9 - 4c^2 + c\sqrt{5(14c^2 - 9)}],$$

$$\Theta^2(z) = \frac{a[\phi(0) - \phi(z)]}{a\phi(z) + \beta}. \qquad (3.5)$$

For fixed $z$, $\phi(z)$ can be found by first solving for $\Theta$ in (3.4) by Newton's method and then for $\phi$ in (3.5). Although the expressions are complicated, the process can be automated quite easily.

The approximations so far derived have relied on expanding $\widehat{A}(k)$ for small $k$, assuming that the waves are mainly composed of the smaller wave numbers. However, for fast waves we expect the pulse to become narrow and so to have an appreciable component from higher wave number modes. This can be accommodated by forming an approximation of $\widehat{A}(k)$ that will be globally valid in $k$. One way to do this is to match an approximation $\widehat{\lambda}(k)$ of $\widehat{A}(k)$ at the low frequency end so that $\widehat{A}(0) = \widehat{\lambda}(0)$ and then to minimise the $L^2$ norm of the difference $(\|\widehat{A}(k) - \widehat{\lambda}(k)\|_2)$. Using a (0,2) Padé form $\widehat{\lambda}(k) = 1/(1 + \alpha k^2)$, the $L_2$ norm is minimised with $\alpha = \tfrac{1}{4}$ and this leads to the ODE

$$c^2(1 - \tfrac{1}{4}D^2)\phi = V'(\phi)$$

with solution

$$\phi(z) = \frac{3}{2a}(c^2 - 1)\operatorname{sech}^2\left(z\sqrt{\frac{c^2 - 1}{c^2}}\right). \qquad (3.6)$$

Comparing this with a similar solution found earlier using the (0,2) Padé approximation

(3.3), we see that this approximation has given us a different length scale for the waves.

All these methods have been applied to the two-dimensional lattice, the KP lattice [14], an electrical transmission lattice of Remoissenet and Michaux [27] and a lattice with second, nearest neighbour interactions [32]. For more details of these results see [33]. The only significant obstacle to its use on more complicated lattice equations is being able to isolate the operator into a single term in Fourier space.

### 3.2. Asymptotics of fast waves

Much has been written about the asymptotic regime $c \to 1$ (see the subsection above). Here the solitary waves are low amplitude and slowly varying. Continuum approximations are applicable and results are often easily obtained in terms of elementary functions.

However in the other limit, $c \to \infty$, very little has been published. A simple approximation has been provided by Druzhinin and Ostrovsky [13], which neglects the coupling between adjacent lattice sites. The result is an approximation in terms of elementary functions, which is valid for the body of the wave, but has the wrong asymptotic behaviour in the tail. A different approach is taken in [14] for the KP lattice where the shape of the high speed travelling waves is shown to be determined by a function from the class $Q_p(z)$ which satisfy the differential-difference equation

$$Q_p''(z) = Q_p(z + 1)^p - 2Q_p(z)^p + Q_p(z - 1)^p. \qquad (3.7)$$

This equation has no solutions in terms of elementary functions, but numerical studies indicate that it is even and pulse shaped. Numerical calculations show that height and width in the first two cases are given by

$$H_2 \approx 1.40, \quad W_2 \approx 1.44,$$

$$H_3 \approx 1.30, \quad W_3 \approx 1.28.$$

We now apply the argument in [14] to the travelling wave equation (1.3) with cubic potential (3.1). We see that since the height of the travelling wave increases with speed $c$, the potential is dominated by the cubic term at high speed and equation (1.3) can be approximated by

$$\frac{c^2}{a}\phi''(z) = \phi^2(z+1) - 2\phi^2(z) + \phi^2(z-1).$$

The solution of this equation is related to the solution $Q_p$ of (3.7) by the transformation

$$\phi(z) = \frac{c^2}{a}Q_p(z). \qquad (3.8)$$

As a consequence, the height of the pulse $\phi$ is prediced to be approximately $1.4c^2/a$ at high speed and the width is fixed at 1.44 lattice spacings.

### 3.3. Variational formulation

A different approach to forming analytic approximations is to make use of the Lagrangian structure of the equations. The Lagrangian formulation gives us an extremum principle for the solution of a system of equations. From the kinetic ($T(\dot{q})$) and potential ($V(q)$) Energies we get the Lagrangian, $L(\dot{q},q) \stackrel{\text{def}}{=} T - V$. Equations of motion are then generated by application of the Euler–Lagrange equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_n}\right) - \frac{\partial L}{\partial q_n} = 0.$$

In this derivation, however, we have skipped over an important point, that the first variation of the action integral, $I = \int L$ vanishes ($\delta I = \delta \int L = 0$). It is this part of the theory which we exploit.

In a fashion reminiscent of finite element techniques, we shall seek such a critical point over a restricted space of functions $X$, for the travelling wave function, $q(z) \in X$. We pick the space $X$ to match important qualitative features of the solution such as symmetry and shape and also

to enable us to carry out integrations explicitly. We use

$$X = \{q(z) = E\tanh(Fz); E, F \in \mathbb{R}_+\},$$

and note that we are working with the variable $q(z)$ which is defined in terms of $q_n(t)$ appearing in (1.1) using the travelling wave ansatz $q_n(t) = q(n - ct) = q(z)$. The function $q(z)$ is related to $\phi(z)$ by

$$\phi(z) = q(z+1) - q(z), \qquad (3.9)$$

and forms a kink soliton, as opposed the pulse shape of $\phi$. This restriction now means that $I$ is a function of $E$ and $F$, so that setting the first variation of $I$ to zero corresponds to the coupled, nonlinear algebraic equations

$$\frac{\partial I}{\partial E} = 0, \quad \frac{\partial I}{\partial F} = 0.$$

As might be expected, the algebra is messy. Using an algebraic manipulation package we can obtain a relationship between $E$ and $c$ and $F$ and $c$ and we can extract information for any speed of solitary wave using numerical routines. Secondly we can use asymptotic analysis to investigate the two limits $c \to 1$ and $c \to \infty$. Both of these have been carried out in detail in [16]. For the cubic potential with $c^2 \to 1$ we recover all of the continuum limit results, since $\phi(z) = q(z+1) - q(z)$ asymptotes to a $\text{sech}^2$ pulse. In the high speed region, the expressions for height and width

$$H_3 = (c^2 - 1)[1.404\,870 + 0.059\,759c^{-2} + \mathcal{O}(c^{-4})], \qquad (3.10)$$

$$W_3 = \sqrt{\frac{c^2}{c^2-1}}[1.433\,565 - 0.346\,955c^{-2} + \mathcal{O}(c^{-4})] \qquad (3.11)$$

show agreement to within about 2% of the $Q_p(z)$ functions described earlier and graphs of the waveform also demonstrate a good fit.

## 3.4. Weak formulation

We now describe a simpler method for the approximation of the travelling wave equation (1.3) which uses a weak formulation of the problem and can be shown to be closely related to the variational formulation described above. More details can be found in [16]. We multiply (1.3) by a smooth test function $\psi(z)$, integrate with respect to $z$ and use integration by parts (assuming appropriate conditions on $\phi(z)$ and $\psi(z)$ as $|z| \to \infty$) to obtain

$$c^2 \int_{-\infty}^{\infty} \psi''(z)\,\phi(z)\,dz$$
$$= \int_{-\infty}^{\infty} V'(\phi(z))$$
$$\times \left[\psi(z+1) - 2\psi(z) + \psi(z-1)\right] dz. \tag{3.12}$$

This equation is satisfied by any resonably well behaved test function $\psi(z)$ when $\phi(z)$ is a solution of (1.3). when $\phi(z)$ is a solution of (1.3).

Now we look for an approximate solution of (1.3) for the cubic potential (3.1). We use the form

$$\phi(z) = A \operatorname{sech}^2(\beta z) \tag{3.13}$$

because it has the expected symmetric pulse shape and satisfies (1.3) in the continuum limit. We substitute this approximate solution in the weak form (3.12), select two simple test functions $\{z^2, z^4\}$ and fix the constants $A, \beta$ by solving the resulting algebraic equations. We find that

$$A = \frac{3}{2a}(c^2 - 1)$$

and

$$\beta^2 = 3\left(\frac{c^2 - 1}{c^2}\right). \tag{3.14}$$

The behaviour as $c \to 1$ is in exact agreement with the approximations (3.3), (3.4) and (3.6).

Unfortunately, numerical tests using this approximate solution as intial data show that it only accurate in producing travelling waves up to speed 1.7 and fails to produce one after that. Further numerical tests using a three parameter approximation

$$\phi(z) = A \operatorname{sech}^n(\beta z)$$

for $n = 4, 6, \ldots$ show that $n$ must be increased to obtain good results at higher speeds. It is possible to obtain an optimal value of $n$ for given speed $c$ by solving the algbraic equations obtained from the weak form (3.12) with three test functions $\{z^2, z^4, z^6\}$. This is quite messy and valid only up to $c \approx 2.5$ where $n \to \infty$ and, taking this limit into account, the approximate solution is

$$\phi(z) = \sqrt{2}(c^2 - 1) \exp\left(-z^2 \frac{3(c^2 - 1)}{2c^2}\right) \tag{3.15}$$

for all higher speeds. The numerical tests indicate that this Gaussian approximate solution gives very good results at high speeds and that it is also reasonably good at low speeds.

This technique is easy to use, particularly when the approximate solution $\phi$ and test functions $\psi$ are chosen such that the integrals in the weak form can be obtained easily.

## 3.5. Numerical results

One way to evaluate the accuracy of the various solutions given above is to compare the speed–height and width–height relationships for the travelling waves they predict with those found by direct numerical approximation as described in section 1. The results for the approximations described in the previous sub-sections are shown in table 1. We see that many methods agree in the low speed ($c \approx 1$) region since they reproduce correctly the simple $\operatorname{sech}^2$ solitary wave predicted by continuum theory. In the high speed limit, we see that the variational approximation (3.9) is very close to the "correct"

**Table 1**
Relationship between height, width and speed predicted by the methods of this section. Width refers to the half height width of the pulse. The last three rows give the leading order behaviour for $c^2 \to 1$ and $c^2 \to \infty$ as appropriate. We use the function $R^2(c^2) = c^2/(c^2 - 1)$.

| Solution/Type | Amplitude | | Width | |
|---|---|---|---|---|
| | low speed | high speed | low speed | high speed |
| (3.3)/Padé | $1.5\,(c^2 - 1)$ | | $1.0177\,R(c^2)$ | |
| (3.6)/Padé | $1.5\,(c^2 - 1)$ | | $1.7627\,R(c^2)$ | |
| (3.13)/weak | $1.5\,(c^2 - 1)$ | | $1.0177\,R(c^2)$ | |
| (3.15)/weak | $\sqrt{2}\,(c^2 - 1)$ | | $1.3596\,R(c^2)$ | |
| (3.4)/Padé | $1.5\,(c^2 - 1)$ | $1.3741\,c^2$ | $1.0177\,R(c^2)$ | $1.4336\,R(c^2)$ |
| (3.8)/fast | not valid | $1.40\,c^2$ | not valid | 1.44 |
| (3.9)/variational | $1.5\,(c^2 - 1)$ | $1.4049\,c^2$ | $1.0177\,R(c^2)$ | $1.3590\,R(c^2)$ |

value $1.40c^2$ found by accurate direct numerical calculation using spectral methods reported in section 1. Also, we see that the pulse width is bounded below at high speeds and the approximate solutions have significant size over only a few lattice spacings.

Another test of the methods is to use the solutions they predict as initial data in the ODE system (1.2) and measure the proportion of energy carried by any resulting travelling wave form if the solution settles down. At low speeds ($1 < c < 1.7$), all the approximations which reproduce the continuum results (3.3), (3.4), (3.9) and (3.13) produce very clean travelling waves carrying most of the input energy. The Padé approximation (3.6) and the weak form approximation (3.15) are designed for high speeds, but perform reasonably well at these low speeds. At speeds above about 1.7 some of the methods fail completely, but the approximations (3.4), (3.6), (3.8), (3.9) and (3.15) are all designed to work at these speeds and perform well.

## 4. Quantum problems

In a volume dedicated to Alwyn Scott, we would be remiss to leave the topic of solitons on lattices without mentioning briefly recent developments in the study of the corresponding quantum version of these problems [20]. For example, the second quantised version of the Hamiltonian for the DNLS equation (2.4) is

$$\hat{H}_1 = -\sum_{j=1}^{f} (b_j^\dagger b_{j+1} + b_j^\dagger b_{j-1} + \tfrac{1}{2}\gamma b_j^\dagger b_j^\dagger b_j b_j),$$

(4.1)

where $b_j^\dagger$ and $b_j$ are bosonic raising and lowering operators satisfying the usual commutation relations. If we calculate the energy levels on a periodic lattice of length $f = 60$, with two quanta (the smallest non-trivial case), the distribution of energy eigenvalues as a function of momentum $k$ looks like fig. 5. The distinct band below the quasi-continuum is the so-called soliton band. A state in this band has a high probability of having both quanta on the same site, hence the name soliton as a localized lump of energy. However, this being quantum mechanics, there is equal probability of finding this double quanta on any site of the lattice.
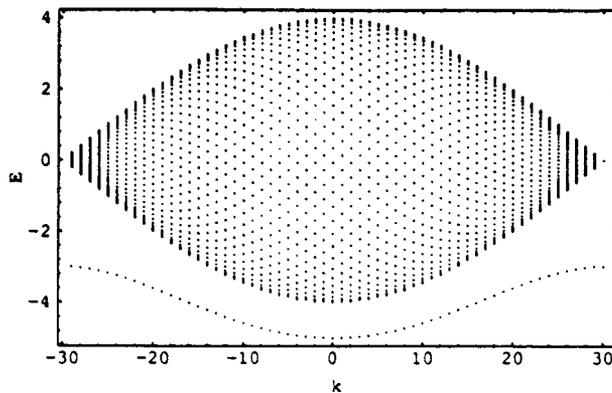
Fig. 5. Quantum energy levels for the DNLS equation.

## Acknowledgements

## References

[1] M.J. Ablowitz and J.F. Ladik, On the solution of a class of nonlinear partial difference equations, Stud. Appl. Math. 57 (1977) 1–12.

[2] M.J. Ablowitz and H. Segur, Solitons and the Inverse Scattering Transform (SIAM, Philadelphia, 1981).

[3] M. Abramowitz and I. Stegun, Handbook of Mathematical Functions (Dover, New York, 1965).

[4] P. Baňacký and A. Zajac, Theory of particle transfer dynamics in solvated molecular complexes: analytic solutions of the discrete time-dependent nonlinear Schrödinger equation. I. Conservative system, Chem. Phys. 123 (1988) 267–276.

[5] J. Candy and W. Rozmus, A symplectic integration algorithm for separable Hamiltonian functions, J. Comp. Phys. 92 (1991) 230–256.

[6] G. Careri, U. Buontempo, F. Galluzi, A.C. Scott, E. Gratton and E. Shyamsunder, Spectroscopic evidence for Davydov-like solitons in acetanilide, Phys. Rev. B 30 (1984) 4689–4702.

[7] P.L. Christiansen and A.C. Scott, Davydov's Soliton Revisited (Plenum, New York, 1990).

[8] D.N. Christodoulides and R.I. Joseph, Discrete self-focussing in nonlinear arrays of coupled waveguides, Opt. Lett. 13 (1988) 794–796.

[9] M.A. Collins, A quasi-continuum approximation for solitons in an atomic chain, Chem. Phys. Lett. 72 (1981) 342–347.

[10] A.S. Davydov, Solitons in Molecular Systems (Reidel, Dordrecht, 1985).

[11] A.S. Davydov and N.I. Kislukha, Solitary excitons in one-dimensional molecular chains, Phys. Stat. Sol. (b) 59 (1973) 465–470.

[12] R.K. Dodd, J.C. Eilbeck, J.D. Gibbon and H.C. Morris, Solitons and Nonlinear Wave Equations (Academic Press, 1984).

[13] O.A. Druzhinin and L.A. Ostrovsky, Solitons in discrete lattices, Phys. Lett. A 160 (1991) 357–362.

[14] D.B. Duncan, J.C. Eilbeck, C.H. Walshaw and V.E. Zakharov, Solitary waves on a strongly anisotropic KP lattice, Phys. Lett. A 158 (1991) 107–111.

[15] D.B. Duncan, C.H. Walshaw and J.A.D. Wattis, A symplectic solver for lattice equations, in: M. Remoissenet and M. Peyrard, eds., Nonlinear Coherent structures in Physics and Biology, Lecture Notes in Physics vol. 393 (Springer, 1991) pp. 151–158.

[16] D.B. Duncan and J.A.D. Wattis, Approximation of solitary waves on lattices using weak and variational formulations, Chaos, Solitons and Fractals 2 (1992) 505–518.

[17] J.C. Eilbeck and R. Flesch, Calculation of families of solitary waves on discrete lattices, Phys. Lett. A 149 (1990) 200–202.

[18] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, The discrete self-trapping equation, Physica D 16 (1985) 318–338.

[19] J.C. Eilbeck, Numerical studies of solitons on lattices, in: M. Remoissenet and M. Peyrard, eds., Nonlinear Coherent Structures in Physics and Biology., Lecture Notes in Physics vol. 393 (Springer, 1991) pp. 143–150.

[20] V.Z. Enol'skii, M. Salerno, A.C. Scott and J.C. Eilbeck, There's more than one way to skin Schrödinger's cat, Physica D 59 (1992) 1–24.

[21] H. Feddersen, Solitary wave solutions to the discrete nonlinear Schrödinger equation, in: M. Remoissenet and M. Peyrard, editors, Nonlinear Coherent structures in Physics and Biology, Lecture Notes in Physics vol. 393 (Springer, 1991) pp. 159–167.

[22] H. Feddersen, Numerical calculations of solitary waves in Davydov's equations, Physica Scripta 47 (1993) 481–483.

[23] A. Hasegawa, Optical Solitons in Fibres (Springer, Berlin, 1990).

[24] D. Hochstrasser, F.G. Mertens and H. Buttner, An iterative method for the calculation of narrow solitary excitations on atomic chains, Physica D 35 (1989) 259–266.

[25] R.D. Parmentier, Fluxons in long Josephson junctions, in: K. Lonngren and A. Scott, eds., Solitons in Action (Academic, New York, 1978) pp. 173–199.

[26] M. Peyrard and M.D. Kruskal, Kink dynamics in the highly discrete sine-Gordon system, Physica D 14 (1984) 88–102.

[27] M. Remoissenet and B. Michaux, Electrical transmission lines and soliton propagation in physical systems, in: G. Maugin, ed., Proc. CMDS6 Conf., Dij 1989 (Longman, 1990) pp. 313–324.

[28] P. Rosenau, Dynamics of nonlinear mass spring chains near the continuum limit, Phys. Lett. A 118 (1986) 222–227.

[29] P. Rosenau, Dynamics of dense lattices, Phys. Rev. A 36 (1987) 5868–5876.

[30] A.C. Scott, Davydov's soliton, Phys. Rep. 217 (1992) 1–67.

[31] M. Toda, Theory of Nonlinear Lattices, Springer Series in Solid State Sciences. vol. 20 (Springer, 1978).

[32] J.A.D. Wattis, Approximations to solitary waves on lattices, III: Lattice with second neighbour interactions, in preparation.

[33] J.A.D. Wattis, Approximations to solitary waves on lattices, II: Quasi-continuum approximations for fast and slow waves, to be published in J. Phys A (1993).

[34] H-L. Wu and V.M. Kenkre, Generalized master equations from the nonlinear Schrödinger equation and propagation in an infinite chain, Phys. Rev. B 39 (1989) 2664–2669.

[35] V.E. Zakharov and A.B. Shabat, Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media, Sov. Phys. JETP [in English translation] 34 (1972) 62–69.

PHYSICA D

# On the prediction of the number of solitons excited by an arbitrary potential: an observation from inverse scattering

P.J. Hansen[a] and K.E. Lonngren[a,b]

[a]*Department of Physics and Astronomy, The University of Iowa, Iowa City, IA 52242, USA*
[b]*Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242, USA*

A heuristic estimate for the soliton production rate by a pulse is verified for the Korteweg–de Vries equation using inverse scattering. An observation from this result, which is shown to hold for some other nonlinear equations and for the case of the "forced" nonlinear Schrödinger equation, is that production is determined by quantities that are invariant under rescaling of the original nonlinear equations. We speculate that this result may be useful to the development of an inverse scattering theory for "forced" nonlinear systems.

## 1. Introduction

A heuristic method for determining the number of Korteweg–de Vries (KdV) solitons produced in an experiment has recently been reported [1]. That method was verified numerically by testing it on the Toda lattice, which becomes a KdV in a continuum limit.

The method arises in noting that for KdV solitons, the product $W\sqrt{A}$ = constant, where $A$ is the soliton amplitude and $W$ its width. Suppose one now applies to the medium an excitation potential amplitude $B$, width $L$. First compute the integer $N$ given by

$$N = (L\sqrt{B}/W\sqrt{A}),\qquad (1)$$

where for the real number $R$, $(R)$ denotes the integer part of $R$. For example, if $R$ is written as the decimal form $R = N.abc\ldots$, $(R) = N$. Then eq. (1) gives the number of whole boxes of area $W\sqrt{A}$ fitting in the box of area $L\sqrt{B}$. Finally, remembering that the solitons are nonoverlapping, the soliton production rule is

$$\text{number of solitons excited} = N + 1.\qquad (2)$$

If a fraction of a box of area $W\sqrt{A}$ appears in computing (1), i.e. the portion $O.abc\ldots$ of $R$ above, this fraction is taken to produce "radiation rather than another soliton". This intuitive way of thinking of solitons arising from a pulse lacks completely any theoretical basis, but the accuracy of (2) has nonetheless been verified numerically [1].

General solutions to soliton equations may be obtained by the method of the inverse scattering transform [4,5]. The nonlinear problem presented by the soliton equation is mapped to a linear scattering problem, for which the solution of the nonlinear equation serves as the scattering potential. The scattering problem is solved by standard linear techniques, and the subsequent inverse mapping back to the nonlinear problem provides its solution. This consists of a number of solitons plus radiation, the solitons corresponding to the bound states of the scattering problem, the radiation to the continuum (scattering) states.

We wish now to establish a connection between the production rule (2) and inverse scattering. It is our intention in this note to suggest that the success of (1) and (2), which after all

amount to no more than a simple minded scaling argument, may actually rest on a deeper, mathematically rigorous property of soliton equations. We have been unable to discover this conjectured rule as yet.

Toward our goal, we present in section 2 a number of scattering problems for soliton equations indicating a connection between the bound state spectrum (obtained from the associated linear scattering problem) and the scaling properties of the original nonlinear soliton equation. In particular, eqs. (1) and (2) are verified for the case of the KdV equation. A possible application of the present observations to the problem of "forced integrable systems" [6] is discussed in our conclusions in section 3.

## 2. Some examples

We consider first the KdV equation

$$\partial_t q(x, t) + 6q\partial_x q + \partial_x^3 q = 0 , \tag{3a}$$

with the initial condition

$$q(x, 0) = \sigma V_0 , \quad -a \le x \le a$$
$$= 0 , \quad |x| > a . \tag{3b}$$

The sign $\sigma = \pm 1$ is present for generality. Associated with (3) is the linear scattering problem [4]

$$\partial_x v_1 + i\zeta v_1 = q(x, 0)v_2 ,$$
$$\partial_x v_2 - i\zeta v_2 = -v_1 , \tag{4}$$

for the eigenvalues $\zeta$. In general, $\zeta$ is complex,

$$\zeta = \xi + i\eta . \tag{5}$$

Solitons produced by (3b) are determined by the bound state spectrum of (4) in the upper half $\zeta$-plane, i.e., $\eta > 0$ [4]. Such states vanish as

$|x| \to \infty$. With $v = (v_1, v_2)$ and $A$, $B$ arbitrary amplitudes, we obtain from (4) the solutions

$$\psi = (0, B e^{i\zeta x}) , \quad x > a ;$$
$$\varphi = (1, -i/2\zeta)A e^{-i\zeta x} , \quad x < -a , \tag{6}$$

which have the appropriate bound state behavior as $|x| \to \infty$. We also take $\xi = 0$ for the bound case. This is done because we are aware beforehand that $v_1$ may be eliminated from (4), leaving the quantum mechanical Schrödinger equation for $v_2$ with eigenvalues $\zeta^2$. This Hermitean operator permits only real $\zeta^2$, so that $\zeta$ corresponding to bound states is purely imaginary.

For $|x| \le a$, insertion of

$$v = (\Gamma_1, \Gamma_2) e^{i\gamma x} \tag{7}$$

in (4) yields the condition

$$\Gamma_2/\Gamma_1 = i/(\gamma - \zeta) = i(\gamma + \zeta)/\sigma V_0 . \tag{8}$$

With

$$\gamma = \kappa + i\rho , \tag{9}$$

we obtain the two solutions

$$\kappa_\pm = \pm\{\tfrac{1}{2}[\xi^2 - \eta^2 + \sigma V_0$$
$$+ ([\xi^2 - \eta^2 + \sigma V_0]^2 + 4(\xi\eta)^2)^{1/2}]\}^{1/2} ,$$
$$\rho_\pm = \xi\eta/\kappa_\pm ,$$
$$\gamma_\pm = \kappa_\pm + i\rho_\pm . \tag{10}$$

As in (8), $\Gamma_{1\pm} = -i(\gamma_\pm - \zeta)\Gamma_{2\pm}$ so that

$$v_1 = -i(\gamma_+ - \zeta)\Gamma_{2+}\rho^{i\gamma_+ x} - i(\gamma_- - \zeta)\Gamma_{2-}\rho^{i\gamma_- x} ,$$
$$|x| \le a ,$$
$$v_2 = \Gamma_{2+}\rho^{i\gamma_+ x} + \Gamma_{2-}\rho^{i\gamma_- x} . \tag{11}$$

Continuity of the solution determines the matching conditions

$$\psi(x = +a) = v(x = +a) \,,$$

$$\psi(x = -a) = v(x = -a) \,. \tag{12}$$

$$\int_{-\infty}^{+\infty} dx \, \sqrt{q_s(x, t)} = \pi\sqrt{2} \,.$$

With a little algebra this gives

$$(\gamma_+ + \zeta)\rho^{-i\gamma_+ a} = \pm(\gamma_+ - \zeta)\rho^{i\gamma_+ a} \,. \tag{13}$$

Finally, recalling $\xi = 0$ for bound states and letting

$$X = \kappa_+ a \,, \quad Y = \eta a \,, \tag{14}$$

equating real and imaginary parts of (13) gives

$$+: \quad X \tan X = Y \,, \tag{15a}$$

$$-: \quad X \cos X = -Y \,. \tag{15b}$$

Also, in either case, from (10),

$$X^2 + Y^2 = a^2(\sigma V_0) \geq 0 \,. \tag{16}$$

Using (8) we find that for (15a), $\Gamma_{2+} = \Gamma_{2-}$, an even solution, while for (15b) $\Gamma_{2+} = -\Gamma_{2-}$, an odd solution.

As previously stated, upon eliminating $v_1$ from (4), one derives the quantum mechanical Schrödinger equation for the potential $-q(x, 0)$. For a different symmetric potential, the conditions (15), (16) have been obtained and solved graphically [7]. The results is that

number of bound states $= N + 1$

if

$$N\pi < 2a\sqrt{\sigma V_0} \leq (N + 1)\pi \,. \tag{17}$$

With the correspondence $W\sqrt{A} \to \pi$ and $L\sqrt{B} = 2a\sqrt{|V_0|}$, this is just the result (1), (2). We note here that the one soliton solution $q_s(x, t)$ of (3a)

$$q_s(x, t) = (2/W^2)\,\text{sech}^2([x - 4t/W^2]/W)$$

obeys

If we substitute $Y$ from (15) into (16) and consider the limit $X \to 0$, we find with (15a) that $Y \sim X^2 > 0$ and $X \simeq a\sqrt{|V_0|}$. Thus this solution, the even case, exists for arbitrarily small $V_0$. This is the origin of the additional term of unity added to $N$ in (2). There is no threshold condition. For the odd case, (15b), this limit gives for $X \to 0$ (but $X \neq 0$), $1 \simeq a\sqrt{|V_0|}$. There is a threshold for the odd case.

Considering next the invariance of (3a) under the scaling transformation

$$x' = \epsilon^a x \,, \quad t' = \epsilon^b t \,, \quad q' = \epsilon^c q \tag{18a}$$

we find

$$b = 3a \,, \quad c = -2a \,, \tag{18b}$$

so that

$$x\sqrt{q} = x'\sqrt{q'} \,. \tag{18c}$$

This invariant combination reflecting shape preservation under (18a) by the nonlinear differential equation is interesting if we make the association of $x$ with $2a$ and $q$ with $\sigma V_0 = |V_0|$. To the extent that this purely dimensional association actually means something, the invariant quantity in (18c) is seen by (17) to be the determiner of soliton production. The soliton spectrum was found from the linear scattering problem using the initial condition for the nonlinear equation. We will return to this shortly.

We consider next the nonlinear Schrödinger equation (NLSE)

$$i\partial_t q(x, t) + \partial_x^2 q \pm 2q|q|^2 = 0 \,, \tag{19}$$

for which the associated scattering problem is

$$\partial_x v_1 + i\zeta v_1 = qv_2 \,,$$

$$\partial_x v_2 - i\zeta v_2 = \mp q^* v_1 \,. \tag{20}$$

Again, with $\sigma = \pm 1$, let

$$q(x, 0) = \sigma V_0 , \quad -a \le x \le a ,$$
$$= 0 , \quad |x| > a . \tag{21}$$

For bound state solutions with $\zeta$ in the upper half plane, solutions are

$$\psi = (0, d\rho^{i\zeta x}) , \quad x > a ;$$
$$\varphi = (\tilde{a}\rho^{-i\zeta x}, 0) , \quad x < -a . \tag{22}$$

For $|x| \le a$, use of the form (7) gives

$$\Gamma_1 / \Gamma_2 = \sigma V_0 / i(\gamma + \zeta) = i(\gamma - \zeta)/(\mp \sigma V_0^*) , \tag{23}$$

while use of (5), (9) yield

$$\kappa = \pm \{ \tfrac{1}{2} [ \pm |V_0|^2 + \xi^2 - \eta^2 $$
$$+ ([\pm |V_0|^2 + \xi^2 - \eta^2]^2 + 4(\xi\eta)^2)^{1/2}]\}^{1/2} ,$$

$$\rho_{\pm} = \xi\eta / \kappa_{\pm}$$

$$\gamma_{\pm} = \kappa_{\pm} + i\rho_{\pm} . \tag{24}$$

Again we use the conditions (12). Then $\gamma_- = -\gamma_+$ results in

$$(-\gamma_+ + \zeta) e^{2i\gamma_+ a} = (\gamma_+ + \zeta) e^{-2i\gamma_+ a} \tag{25}$$

analogous to (13). Defining

$$X = 2a\kappa_+ , \quad P = 2a\xi ,$$

$$r = 2a\rho_+ , \quad Y = 2a\eta , \tag{26}$$

we obtain upon equating real and imaginary parts of (25)

$$\cos X [e^r (X + p) - e^{-r} (-X + p)]$$
$$= - \sin X [e^r(r + Y) + e^{-r}(-r + Y)] ,$$

$$\cos X [e^r (r + Y) - e^{-r} (-r + Y)]$$
$$= \sin X [e^r (X + p) + e^{-r} (-X + p)] , \tag{27a}$$

and from (24),

$$X^2 = \tfrac{1}{2} \{ \pm 4a^2 |V_0|^2 + p^2 - Y^2$$
$$+ [(\pm 4a^2 |V_0|^2 + p^2 - Y^2)^2 + 4(pY)^2]^{1/2} \} . \tag{27b}$$

In terms of $\xi$, a single NLSE soliton may be written as [6]

$$q_s(x, t) = 2\eta \frac{\exp[-4i(\xi^2 - \eta^2)t - 2i\xi x]}{\cos h[2\eta(x - x_0) + 8\eta\xi t]} ,$$

the speed of which is seen to be $dx/dt = -4\xi$. Then the condition for production of solitons at rest is found by letting $\xi = 0$. From (26), $p = 0 = r$ and (27) is

$$X \cot X = -Y ,$$

$$X^2 + Y^2 = \pm (2a|V_0|)^2 \ge 0 , \tag{28}$$

so that only the + sign may be considered. This is the odd case, (15b) and (16), of the KdV example. As noted, there is a production threshold for $|V_0|$. Eliminating $Y$ and letting $X \to 0$, we find $Y \simeq -1$, $X^2 \simeq (2a|V_0|)^2 - 1 \ge 0$.

The quantity governing production (i.e., the bound state spectrum) in (28) is $2a|V_0|$. According to the earlier dimensional speculation, this corresponds to $xq$. Substituting the scaling (18a) into (19) yields

$$b = 2a , \quad c = -a ,$$

$$x'q' = xq . \tag{29}$$

Interestingly, the modified Korteweg – de Vries (MKdV) equation

$$\partial_t q(x, t) \pm 6q^2 qx + q_{xxx} = 0 \tag{30}$$

yields, upon employing (18a),

$$b = 3a , \quad c = -a \Rightarrow x'q' = xq , \tag{31}$$

i.e., the $x$, $q$-invariance as the NLSE. An analysis along previous lines of the associated linear scattering problem

$$\partial_x v_1 + i\zeta v_1 = qv_2,$$

$$\partial_x v_2 - i\zeta v_2 = \mp qv_1 \qquad (32)$$

leads to the condition (25) already obtained for the NLSE. Of course, subsequent time evolution is different in the two cases. In particular, $b$ is different in (29) and (31).

We also see from (29) that $tq^2 = t'q'^2$ holds. As the inverse scattering transform contains a pair of equations describing the time evolution of the scattering functions [4], the previous discussion suggests that a determination of the eigenvalue problem using the temporal rather than the spatial scattering problem, would show the eigenvalue spectrum to be determined by $\tau|V_0|^2$, with $\tau$ the time of a pulse $|V_0|$ applied to the system at some fixed position $x$. Indeed, a consideration of the forced NLSE, driven by a time dependent pulse at $x = 0$, shows this to be the case [6]. (To see this in ref. [6], one must note that in eq. (3) of that paper, $a = \tau^2$, and then consider eqs. (16) and (17), along with the intervening discussion.)

## 3. Discussion

For the KdV, the single soliton solution shows that the product of width times the square root of the amplitude is constant. We have seen that the heuristic estimate of soliton production made by considering a pulse of certain width and amplitude as an integral number of solitons according to the box counting procedure given, with any excess relegated to "radiation", holds rigorously in the KdV example. We also note that the powers of width and amplitude determining the soliton spectrum are actually those obtained upon demanding invariance under a rescaling of the KdV, and that a similar result occurs for the

NLSE and MKdV. This not only held for the initial value problem (the spatial evolution equations of the inverse scattering transform), but also for the "forced" problem, as seen by considering the time evolution equations of the inverse scattering transform the NLSE.

We established eq. (17) with use of the integral of the single KdV soliton. In the spirit of the WKB we generalize the results to speculate that

$$\int dx \, q_N^P(x, 0) = CN,$$

with $N$ = integer number of solitons produced, $C$ the constant determined by $N = 1$, and $P$ the exponent obtained from the scale invariance of the governing equation. A similar formula can be written with time integration for the "forced" problem. As such a relation relates the soliton production to a certain integral of the initial/boundary conditions, our speculation is that the scaling behavior of the nonlinear differential equation could perhaps provide a useful approach to solving the problem of developing an inverse scattering transform for forced nonlinear equations [8].

## Acknowledgements

## References

[1] K.E. Lonngren, P. Hansen and T. Kuusela, Physica Scr. 46 (1992) 63.

[2] T. Kuusela and J. Hietarinta, Rev. Sci. Instrum. 62 (1991) 2266.

[3] T. Lonngren, in: Solitons in Action, eds. K. Lonngren and A.C. Scott (Academic Press, New York, 1978) p. 127.

[4] M. Ablowitz, D. Kaup, A. Newell and H. Segur, Stud. Appl. Math. 53 (1974) 249.

[5] G.L. Lamb, Elements of Soliton Theory (Wiley-Interscience, New York, 1980).

[6] D.J. Kaup and P.J. Hansen, Physica D 18 (1986) 77.

[7] L.I Schiff, Quantum Mechanics, 3rd. Ed. (McGraw-Hill, New York, 1968).

[8] D.J. Kaup and P.J. Hansen, Physica D 25 (1987) 369.

PHYSICA D

# Data-parallel Langevin dynamics simulations of Josephson junction arrays

P.S. Lomdahl and N. Grønbech-Jensen
*Theoretical Division and Advanced Computing Laboratory, Los Alamos National Laboratory,
Los Alamos, NM 87545, USA*

We discuss the advantages of parallel computing with respect to both computing speed and coding techniques. As an example of an application, we show results of dynamical simulations of two dimensional Josephson lattices in external magnetic fields. Using data parallel computing, we are able to study the dynamics of very large systems and identify the interplay between spatial and temporal scales in complex systems.

## 1. Introduction

Typically nonlinear problems with origin in condensed matter physics and materials science are ideally suited for implementation on massively parallel computers. This is of course true for many other branches of science, but for condensed matter problems the reasons are especially clear. First, nonlinear problems in higher dimensions are most often formulated on a lattice, the lattice for example represents the underlying discreteness of a substrate of atoms or discrete $k$-values in Fourier space. The fundamental unit of parallelism thus has a direct analog in the physical system we are interested in. Secondly, we are often interested in understanding the influence of temperature on the system, i.e. the system is in contact with a heat bath. There are of course several was to introduce temperature in a problem, but often it is most convenient to operate in the canonical ensemble, i.e. having the temperature constant and the energy fluctuate. This situation is also closer to many experimental setups. Third, interactions are often local, i.e. we can represent the important physics with a nearest or next-nearest neighbor interaction. Even when this is not the case parallel implementations of the problem is often possible

which nevertheless maintain a local communication topology on the machine. Fourth, we are often interested in the behavior of large systems, as close to experimental situations as possible. This naturally leads to formulations with a large number of lattice sites, especially for 2D and 3D problems. Finally, we are interested in the dynamic response of the system, we want to follow its time evolution. All of these factors lead to a formulation of the problem in terms of a large number of coupled Langevin ordinary differential equations with near neighbor interaction.

Fine grained massively parallel multicomputers like the CM-2 and CM-5 from Thinking Machines Corporation are ideally suited for these types of problems. We will illustrate this point in the following, first with a simple example of a data-parallel solution to Laplace's equation, secondly with an example of data-parallel techniques applied to problems from the dynamics of Josephson junction arrays. One might argue that massively parallel computers are very expensive and only accessible to a few fortunate researchers and therefor not of general interest. We do not think that argument holds. Key computer industry analysts now predict that by the end of this decade the majority of computers, from the PC and workstation to the high-end

number-chruncher will be multiprocessors and parallel processors. Over the next five years, it is critical that computational physicists understand this revolutionary technology if we want to make progress on "Grand Challenge" problems in nonlinear science and elsewhere.

## 2. Data-parallel programming

In this section we will briefly illustrate how programming a massively parallel computer is surprisingly simple when the data-parallel approach is followed. In fact, once a few basic concepts are understood a typical code is significantly *simpler* that a corresponding code for a serial computer.

The Connection Machine 2 (CM-2) is a massively parallel distributed memory computer containing between 4096 and 65536 bit-serial processors. Each 32 processors share a floating-point unit, so for most scientific calculations a 65536 processor CM-2 can also be viewed as a 2048 floating-point processor. The CM-2 is a SIMD (single instruction–multiple data) computer, i. e. all the processors execute the same instruction in lock step. All processors are synchronized at all times and problems such as dead-locks, race-conditions, etc. can not occur. Each processor has control over its own memory, - up to 128 Kbytes, for a total of 8 Gbytes on a 64K processor machine. For communications purposes the processors can dynamically be arranged as a hypercube shape in up to 31 dimensions. For most applications in nonlinear condensed matter physics the dimension of the hypercube is low, - corresponding to 2D and 3D problems. The CM-5 is a newer and more general massively parallel multicomputer, which in addition to the SIMD data-parallel programming model, also supports the more general MIMD (multiple instructions–multiple data) programming concept. The interested reader is referred to [1] for more details.

The data-parallel computing model maps very cleanly on to the CM-2 SIMD structure. Data parallel computation is parallelism through the simultaneous execution of the *same* instruction on a large set of data. This should be contrasted with control-parallelism , where speed-up is achieved through the simultaneous execution of *different* instructions. This kind of parallelism is much harder to manage and best results are usually achieved at the hardware level. The data-parallel concept is most easily illustrated with an example. Given three $N \times N$ ($N = 256$ e. g.) matrices $A, B$, and $C$; a matrix element from each matrix is associated with a processor - i. e. each processor has allocated memory for three matrix elements only. The statement $C = A + B$ is a single statement in a program and will be *executed* as such in the data-parallel computing model. One each processor the individual matrix element calculation is performed simultaneously. This high-level abstract model is important because it makes programs less machine dependent and thus more portable. The data-parallel model is supported in the languages C* and CM-Fortran on the CM-2 and CM-5. The data-parallel languages also support the notion of *virtual* processors. Let us say that $N$ in the above example was 512, in this case $N^2 = 262144$ which is four times larger than the maximum number of processors. With virtual processors, each real processors is now performing the work of four fictitious processors. Of course, at this level the work for each of the four virtual processors must proceed in serial on each of the real processors. Nevertheless, this model of computing is extremely useful and conceptually simple. Note for example, that a program written for say a 65K processor CM-2 will run unchanged on a 16K CM-2, but take four times as long. The current hardware of the CM-2 restricts the layout of the processors to have axis that are powers of two. On the CM-5 these restrictions do not apply.

As a slightly more interesting example, consider a finite-difference approximation for calcu-

```
shape           [512][512]grid;
float:grid      a, b;
boolean:grid    interior_element;

    ...         /* Initialize the variables to something appropriate */

where (interior_element) {
    b = [. + 1][.]a + [. - 1][.]a + [.][. + 1]a + [.][. - 1]a
        - 4 * [.][.]a;
} else {

    ...         /* Do something for boundary elements */

}
```

Fig. 1. C* code implementing a finite difference approximation for calculating the Laplacian in 2D.

lating the Laplacian in 2D. We would discretize space such that $a(x,y) = a(i\Delta x, j\Delta y) = a_{ij}$ and an approximation of the Laplacian correct to 2nd order in $\Delta x$ and $\Delta y$ would be: $(a_{i+1j} + a_{i-1j} + a_{ij+1} + a_{ij-1} - 4a_{ij})/\Delta x^2 \Delta y^2$. The essential part of a C* program implementing this calculation would look like fig. 1.

The first line defines a processor layout with the new parallel C type shape. Here we have defined a 512 × 512 2D shape called grid. The next line defines two floating-point parallel variables, a and b, each with the grid shape. The third line defines a similar logical variable, interior_element. The rest of the code show a where-else control structure, the body of which calculates the actual Laplacian. The left-index notation [.][.+1]a, is another C* extension that allows addressing of neighboring processors. The "dot" notation should be read as "this processor's coordinate". This code is executed on all processors in parallel and depending on the value of interior_element, a processor will either sit idle or participate in the calculation of the interior Laplacian. The code to initialize interior_element has been omitted. The else branch of the control structure is then subsequently executed, here only the boundary processors are active and all others sit idle. On the CM-5 the two parts of the control structure could execute in parallel.

We have chosen to show an actual piece of code, because we believe that it shows most convincingly that the basic concepts of data-parallel programming are indeed extremely simple. In comparison with a conventional serial program, one notes the complete absence of any loops over the lattice sites. The Laplacian is calculated with *one* statement in the data-parallel code, which represents very cleanly how we think about the problem. Having illustrated the basic principles behind data-parallel computing we now turn to a real application from the dynamics of 2D Josephson junction arrays. Interested readers who wants to know more about data-parallel programming are referred to [2].

## 3. Josephson junction arrays

As an example of an illustrative system modeled by data-parallel computing we show long-time Langevin dynamics simulations of large $(N_x \times N_y, ; N_x = N_y = 128)$ 2-dimensional arrays of Josephson junctions in a uniformly frustrating external magnetic field. The results demonstrate: (i) Relaxation from an initially random flux configuration as a "universal" fit to a "glassy" stretched-exponential type of relaxation for the intermediate temperatures, and an "activated dynamic" behavior for $T \sim T_c$; (ii) A *glassy* (multi-time, multi-length scale) voltage response to an applied current. Intrinsic dynam-

ical symmetry breaking induced by boundaries as nucleation sites for flux lattice defects gives rise to *transverse* and noisy voltage response.

The Lagrangian for the JJA takes the form (see refs. [3–5] and references therein)

$$\mathcal{L} = -E_0 \sum_{i,j} [2 - \cos(\theta_{ij} - \theta_{i-1j} - A_{i-1j,ij})$$
$$- \cos(\theta_{ij} - \theta_{ij-1} - A_{ij,ij-1})]$$
$$+ \tfrac{1}{2} C_0 \sum_{i,j} \dot{\theta}_{ij}^2 - E_1 \sum_{i,j} \theta_{ij}(\delta_{1j} - \delta_{N_xj}). \quad (1)$$

Here, $\theta_{ij}$ is the phase of the superconducting island with the discrete coordinates ($i = 1, 2, \ldots, N_x$, $j = 1, 2, \ldots, N_y$) of the lattice, and $A_{ij,kl} \equiv (2e/\hbar c) \int_{ij}^{kl} A \cdot d l$ is the integral of the vector potential from island $(i,j)$ to a neighboring island $(k,l)$. The energy associated with each Josephson link is denoted by $E_0$, and the capacitance between each superconducting island and the ground plane is $C_0$. An external potential $E_1$ is imposed at the edges, $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$. The vector potential $A$ is given in the Landau gauge,

$$A_{ij,kl} = 2\pi f \frac{x_{ij} + x_{kl}}{2} (y_{kl} - y_{ij}), \quad (2)$$

$$x_{ij} = x_{ij}^0 + \delta x_{ij}, \quad y_{ij} = y_{ij}^0 + \delta y_{ij}.$$

Here, $\delta x_{ij}$ and $\delta y_{ij}$ represent a spatial disorder around the ordered positions $(x_{ij}^0, y_{ij}^0)$. The model of the Josephson links sketched above lacks two potential contributions. One is the capacitive coupling between the superconducting islands. This interaction introduces implicit terms in to the equations of motion and complicates the computational procedure slightly, but there are no formal problems in the modeling as can be learned from e.g. ref. [6]. We have, however, decided to study the system assuming that a ground plane is present, since this case makes the capacitance between the islands insignificant when compared to the capacitance to the ground plane. The other contribution to the dynamics is the long range inductive interaction between the currents in the network. This

depends on the magnetic properties around the system and the inductive interaction is therefore not easy to define theoretically. Secondly, this type of interaction poses potential problems for a numerical scheme since the dynamics at a given point depend on all other variables in the system. Some attempts to model the mutual inductance have recently been published [7], using interaction between nearest neighbor plaquettes only. However, omitting the inductive interaction does not seem to prevent the model from presenting results that are close to experiments (see e.g. ref. [8]).

The $A_{ij,kl}$ summed around a plaquette obeys the following relation: $A_{ij,kl} + A_{kl,kl-1} + A_{k-1l-1,ij-1} + A_{ij-1,ij} = 2\pi f$, where the frustration $f = Ha^2/\Phi_0$ is a constant giving the average number of flux quanta $\Phi_0 = hc/2e$ of the external magnetic field $H$ through the area $a^2$ of each plaquette of the array. We also introduce the fractional charge $q_{ij}$, obtained as the gauge invariant phase sum around the $ij$'th plaquette:

$$q_{ij} = \frac{1}{2\pi} [ (\theta_{ij} - \theta_{ij-1} - A_{ij-1,ij}) \bmod \pi$$
$$+ (\theta_{ij-1} - \theta_{i-1j-1} - A_{i-1j-1,ij-1}) \bmod \pi$$
$$+ (\theta_{i-1j-1} - \theta_{i-1j} - A_{i-1j,i-1j-1}) \bmod \pi$$
$$+ (\theta_{i-1j} - \theta_{ij} - A_{ij,i-1j}) \bmod \pi]. \quad (3)$$

The dynamical equations derived from eq. (1) are (in normalized units)

$$\ddot{\theta}_{ij} = \sin(\theta_{ij+1} - \theta_{ij} - A_{ij,ij+1})$$
$$+ \sin(\theta_{ij-1} - \theta_{ij} - A_{ij,ij-1})$$
$$+ \sin(\theta_{i+1j} - \theta_{ij} - A_{ij,i+1j})$$
$$+ \sin(\theta_{i-1j} - \theta_{ij} - A_{ij,i-1j})$$
$$- \eta(\dot{\theta}_{ij+1} + \dot{\theta}_{ij-1} + \dot{\theta}_{i+1j} + \dot{\theta}_{i-1j} - 4\dot{\theta}_{ij})$$
$$+ \lambda_{ij}(t) + J_1(\delta_{1j} - \delta_{N_xj}). \quad (4)$$

Time is normalized to $\tau = (C_0\hbar/2eI_0)^{1/2}$, where $I_0 = (2e/\hbar)E_0$ is the critical current of a link, and the normalized external bias current at the edge points is $J_1 = (2e/\hbar I_0)E_1$. The normalized dissipation is given by $\eta = (1/R)(\hbar/2eC_0I_0)^{1/2}$ with $R$ the normal resistance of the junctions, and the thermal noise

associated with the dissipation is introduced in the classical Langevin sense: $\langle \lambda_{ij}(t) \rangle = 0$, $\langle \lambda_{ij}(t) \lambda_{kl}(t) \rangle = 2\eta T \delta(t - t')$, where the temperature $T$ is normalized to $E_0/k_B$, $k_B$ being the Boltzmann constant. We note here that the form of eq. (4) lend itself to a very natural implementation in data-parallel $C^*$, like the example in the previous section.

(i) Relaxation studies with periodic boundary conditions:

For this part of the study we have chosen the damping parameter to be $\eta = 1$, and of course $J_1 = 0$, since there are no open boundaries. The periodic boundary conditions are implemented through the physical observables - not through the phase, $\theta_{ij}$. For any choice of gauge for the external magnetic field there are many ways of choosing the boundary conditions for $\theta_{ij}$. We have made the choice of introducing the phases:

$$\theta_{0j} = \theta_{N_xj},$$
$$\theta_{N_x+1j} = \theta_{1j} + A_{N_xj,N_x+1j} - A_{0j,1j}$$
$$\theta_{i0} = \theta_{iN_y},$$
$$\theta_{iN_y+1} = \theta_{i1}, \tag{5}$$

which, when applied to the system defined above, will give the right periodicity of the physical system. The ground state for the maximal frustration parameter $f = \frac{1}{2}$ the fractional charge $q$ exhibits a checkerboard pattern, $q = \pm(-1)^{i+j}$. Therefore the following gauge invariant correlation function is defined as a measure of the magnetization of the system (for $f = \frac{1}{2}$):

$$C(t) = \frac{1}{N_x N_y}$$
$$\times \sum_{i,j} q_{ij}(q_{i+1j} + q_{i-1j} + q_{ij+1} + q_{ij-1}). \tag{6}$$

This quantity is $C = -1$ for the $f = \frac{1}{2}$ ground state and $C = 0$ for the random flux configuration. The initial conditions in $\theta_{ij}$ are random and we have followed the dynamics in normalized times up to $10^4$.

We have shown the evolution of $C(t)$ for various temperatures $T$ in fig. 2. It is important to note that the transition temperature $T_c$ for $f = \frac{1}{2}$ is $T_c \simeq 0.45$: for $T > T_c$ long-range flux order and superconductivity are lost. In fig. 2a we see how the large temperature prevents the system from reaching its ground state ($T = 0.5$) and how a lower temperature ($T = 0.3$) results in a final state very close to the ground state, $C = -1$. However, we also observe how a very small temperature ($T = 0.01$) may result in a final state which is far from the ground state. This behavior of the final state as a function of the temperature can be understood as trapping into metastable (after the initial relaxation) flux configuration at very low temperatures because of the uniform frustration. As the temperature is increased the thermal tunneling over the frustration barriers is allowed and $C$ approaches closer to the ground state value. For further increasing temperatures the thermal energy approaches (or exceeds) the frustration pinning energy and thermal randomization occurs ($C \to 0$ for $T \to \infty$).

In fig. 2b we have shown the quantity $\ln \ln C(t)$ as a function of $\ln t$. For the intermediate temperatures ($0.3T_c \lesssim T \lesssim 0.7T_c$) we find an excellent fit to a 'glassy' stretched exponential type of relaxation [10]. Specifically, we find the dependence, $C(t) \sim \exp[-(t/\tau)^\beta]$, with $\beta \approx 0.45$, in good agreement with studies of other multi time-scale systems [10].

In fig. 2c we have shown $\ln \ln |C(t)|$ as a function of $\ln \ln t$ for a few temperatures around the critical value $T_c = 0.45$. Here critical effects dominate and 'activated dynamics' has been proposed in the form [10], $C(t) \sim \exp[-(\ln(t/\tau))^\delta]$. As is seen this is in fair agreement for $T \approx T_c$, where we find the exponent, $\delta \approx 0.9$.

Figure 3 displays the states for $t = 40000$ for three different temperatures, (a) $T = 0.01$, (b) $T = 0.25$, and (c) $T = 0.45$. We have shown (left) the staggered order parameter, $\tilde{q}_{ij} = (-1)^{i+j}$, since this quantity will be uniform as a function of $i$ and $j$ if the system is in a ground state $f = \frac{1}{2}$. We have also shown (right) the spatial Fourier transforms, $g(k)$, of
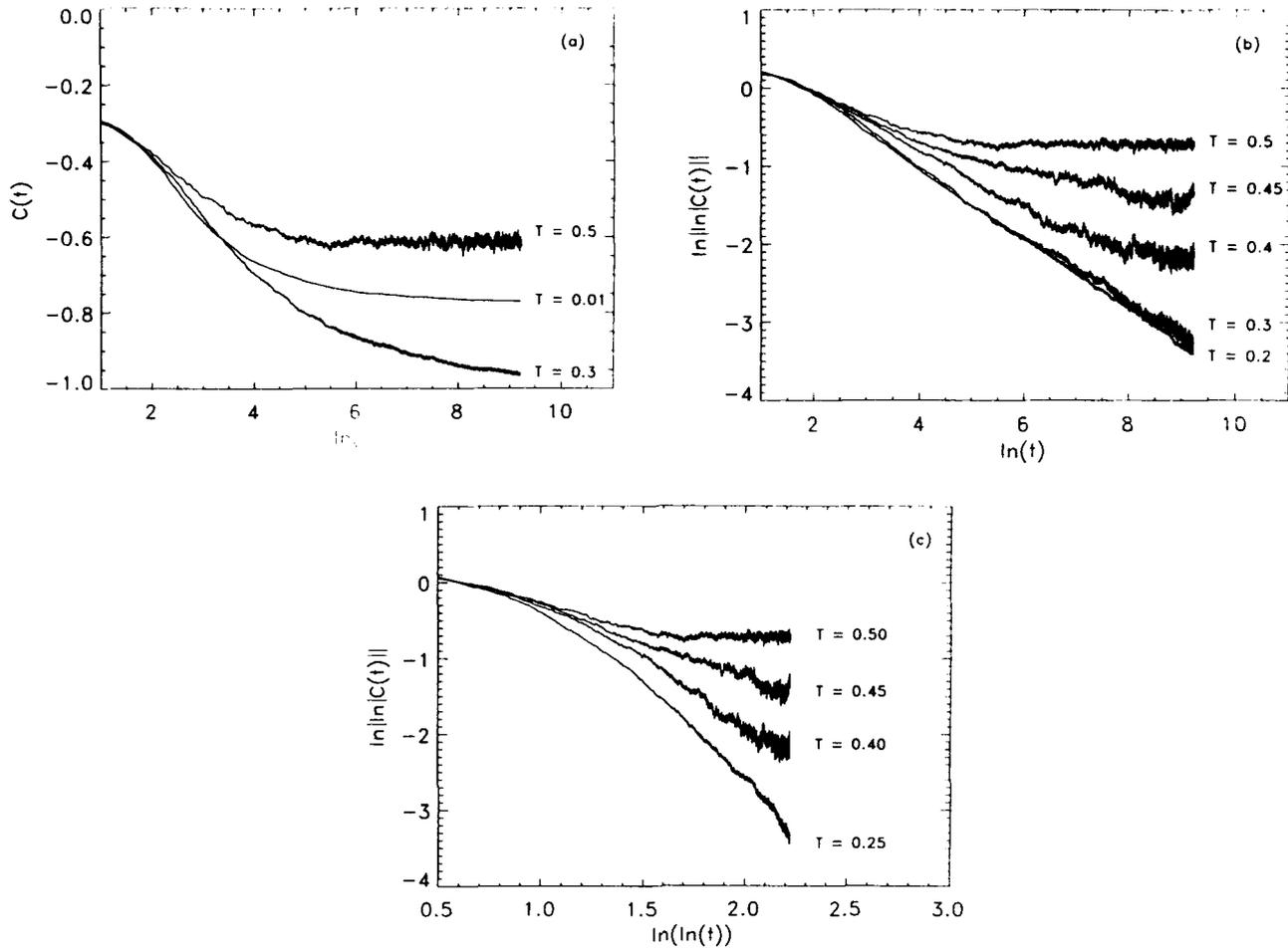
Fig. 2. The global correlation function $C(t)$ during relaxation from random initial condition for $f = \frac{1}{2}$ and zero disorder, $\delta x_{ij} = \delta y_{ij} = 0$.

the shown $\tilde{q}_{ij}$ distributions. This function will be sensitive to the distribution of domain sizes rather than the detailed structure of domain walls. Comparing the structures of the different temperatures we see that the intermediate temperature creates a system which is dominated by only a few $k$ scales, when compared to the system in very small (a) or in large temperature (c). The spatial structures of the states also show some qualitative differences between the very cool and the warm system. Looking at the spectral plots we find that the spectral power centers around the horizontal and vertical axes, whereas for the warm system the distribution is more independent of the angular orientation.

This is due to the thermally nucleated vortex antivortex pairs visible in fig. 3c (left).

Figure 4 shows how a thermally nucleated vortex–antivortex pair may be annihilated by a domain wall. For three different times we show the same part of the system. The first shows the domain wall and the nucleation in its close vicinity. The second shows the wall getting contact with the nucleation, and the third shows the total annihilation. The normalized time span from fig. 4a to 3c is $\approx 2$.

(ii) Externally driven dynamics with open boundary conditions: In this case the boundary conditions are described by the phases,

$$\theta_{0j} = \theta_{1j} + A_{0j,1j},$$

$$\theta_{N_x+1j} = \theta_{N_xj} + A_{N_xj,N_x+1j},$$

$$\theta_{i0} = \theta_{i1} + A_{i0,i1},$$

$$\theta_{iN_y+1} = \theta_{iN_y} + A_{iN_y,iN_y+1}. \tag{7}$$

These boundary conditions will result in zero current flow through the edges - except the current supplied by the external source, $J_1$. We have studied the system as defined above for the dissipation parameter $\eta = 5$ and the external magnetic field $f = \frac{1}{2}$. In fig. 5 we show the current–voltage $(IV)$ characteristics of the system. The current is supplied at the vertical edges of the system and the resulting dc voltages, $V_x$ and $V_y$, are measured and displayed in fig. 5a and fig. 5b, respectively. The voltages are defined in the following way,

$$V_x = \frac{1}{N_y} \sum_{j=1}^{N_y} (\dot{\theta}_{N_xj} - \dot{\theta}_{1j}),$$

$$V_y = \frac{1}{N_x} \sum_{i=1}^{N_x} (\dot{\theta}_{iN_y} - \dot{\theta}_{i1}). \tag{8}$$

At the macroscopic level of the $J_1 - V_x$ curve we see no essential differences from the periodic boundary conditions [3]. Examination of the dynamics reveals the presence of a window of noisy response for $J_c(f = \frac{1}{2}) \lesssim J_1 \lesssim 0.55$. This multi time-scale response is understood in terms of irregular domain wall separations, shown in the insets ($J_1 = 0.4$) in fig. 5a, where the $\tilde{q}_{ij}$ distribution is displayed. Figure 5b shows that a large transversal voltage is uniquely associated with the 'chaotic' regime of the dynamics. Indeed, we see that the amplitude of $V_y$ is largest in the most noisy regime (lowest $J_1$) and vanishes in the periodic regime. The origin of the transversal voltage generation can be understood from the insets in fig. 5. In fig. 5b the solid curves show the time evolution of the voltage,

$$V_y^{(l)} = \frac{2}{N_x} \sum_{i=1}^{N_x/2} (\dot{\theta}_{iN_y} - \dot{\theta}_{i1}), \tag{9}$$

and the dotted curves show the time evolution of the voltage,



Fig. 3. The $\tilde{q}_{ij}$ configuration (left) for the relaxed states ($t = 10^4$) at different temperatures and the spectral power (right) $|g(k)|$. $\tilde{q}_{ij} > 0$ – black; $\tilde{q}_{ij} < 0$ – white. The color code for the spectral power is $\log_{10}(|g(k)|^2) > 10^3$ – black; $10^5 < \log_{10}(|g(k)|^2) < 10^3$ – grey; $\log_{10}(|g(k)|^2) < 10^5$ – white. (a) $T = 0.01$; (b) $T = 0.25$; (c) $T = 0.45$.
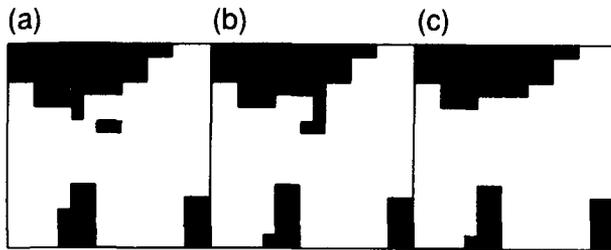


Fig. 4. Annihilation of a thermally nucleated ($T = 0.25$) vortex–antivortex pair on a domain boundary. The plots show $16 \times 16$ of the total $128 \times 128$ system for three different times ($t_a < t_b < t_c$, $t_a - t_c \approx 2$). The $\tilde{q}_{ij}$ configuration is shown as in fig. 2 (left).

Fig. 5. The voltages across the JJA as a function of the applied bias current. (a) The longitudinal dc voltage. The insets show the $\tilde{q}_{ij}$ configuration (color code as in fig. 2) for two characteristically different cases. The bias current is forced through the system along the horizontal axis. (b) The amplitude of the transversal voltage. Insets: time evolution of the transversal voltages, $V_y$ (dashed), $V_y^l$ (solid), and $V_y^r$ (dotted).

$$V_y^{(r)} = \frac{2}{N_x} \sum_{i=N_x/2+1}^{N_x} (\dot{\theta}_{iN_y} - \dot{\theta}_{i1}), \qquad (10)$$

and the dashed curves show the sum, $V_y$. In the periodic regime we find large oscillations for $V_y^{(l)}$ and $V_y^{(r)}$ (see fig. 5b for $J_1 = 0.7$), but since they are of equal size and opposite sign, the total transversal voltage is zero. For the chaotic regime, on the other hand, the dynamically induced spatially disorder (see fig. 5a for $J_1 = 0.4$) makes the two contributions unequal in size, and hence the sum, $V_y$, starts to oscillate as shown in fig. 5b (inset for $J_1 = 0.4$). Furthermore, as suggested in the inset of fig. 5a, and as we have observed through careful examination of the time evolution, the asymmetry between the left and right half planes develops by local flux defects nucleating at the boundaries and propagating into the interior. It is here important to note that the dynamically induced transversal voltage response to the longitudinal dc current is purely oscillating and has no dc value. Detailed studies [5] of the power spectra of the longitudinal and transversal voltages show that the frequencies of $V_x$ and $V_y$ are closely related. Not surprisingly, the Josephson frequency is present in the signals, but also a large fre-

quency contribution arises from the injection of domain walls at the driven edges. Further, the breaking of the symmetry in the chaotic regime results in an exact half harmonic generation, since each half of the system now contributes separately. Let us conclude this section by noting that the symmetry breaking and transverse voltage generation has been observed for a disordered system as well [5] and the results of the dynamically induced disorder and the positional disorder are qualitatively very similar.

## 4. Conclusion

We have used the data-parallel programming model to study the dynamics of large frustrated arrays of Josephson junctions on the CM-2. This approach allow for a flexible and intuitive implementation of the basic dynamical equations. Our study reveals an interesting relation between spatial and temporal scales, including relaxation from random flux configurations which exhibit a stretched-exponential behavior. The dynamics of the noisy response is controlled by "defects" structures - vortices and domain walls, which are defined with respect to the underlying ground-

state flux structure. The Josephson junction arrays are ideal laboratory scale systems for studying the complexity of spatio-temporal nonlinear phenomena. Using data-parallel computing on the new massively parallel computers we are now in a position to model these systems in a one-to-one correspondence.

We thank the Los Alamos Advanced Computing Laboratory for generous support and for making their facilities available to us. This work was performed under the auspices of the US Department of Energy.

## References

[1] C.E. Leiserson et al., in: Proc. Symposium on Parallel and Distributed Algorithms '92, June 1992, San Diego.

[2] P.J. Hatcher and M.J. Quinn, Data-Parallel Programming of MIMD Computers, The MIT Press (1991).

[3] F. Falo, A.R. Bishop, P.S. Lomdahl, Phys. Rev. B41 (1990) 10983.

[4] N. Grønbech-Jensen, A.R. Bishop, F. Falo and P.S. Lomdahl, Phys. Rev. B45 (1992) 10139.

[5] N. Grønbech-Jensen, A.R. Bishop, F. Falo and P.S. Lomdahl, Phys. Rev. B46 (1992) 11149.

[6] Ravi Bhagavatula, C. Ebner and C. Jayaprakash, Phys. Rev. B45 (1992) 4774.

[7] D. Dominguez and Jorge V. José, Phys. Rev. Lett. 69 (1992) 514.

[8] S.P. Benz, M.S. Rzchowski, M. Tinkham and C.J. Lobb, Phys. Rev. Lett. 64 (1990) 693; J.U. Free, S.P. Benz, M.S. Rzchowski, M. Tinkham and C.J. Lobb, Phys. Rev. B41 (1990) 7267.

[9] I.M. Lifshitz and V.V. Slyozov, J. Phys. Chem. Solids 19 (1961) 35.

[10] See e. g. C. Dekker, A.F.M. Arts and H.W. de Wijn, Phys. Rev. B40 (1989) 11243.

**PHYSICA** Ⓓ

# Non-linear properties of Josephson junctions

## N.F. Pedersen

*Physics Laboratory I, The Technical University of Denmark, DK-2800 Lyngby, Denmark*

The non-linear properties of Josephson junctions are reviewed. Those properties include the possibility to have soliton propagation in a Josephson junction transmission line, i.e., a very ordered state. The non-linear properties also include the possibility for chaos in many different Josephson junction systems, i.e., a very non-ordered state. Several examples from experiments are given.

## 1. Introduction

The Josephson junction is a unique non-linear system for several reasons. Firstly one may find most of the generic non-linear behaviour in various Josephson junction systems. For example chaos transitions of several different kinds may be found, including the famous period doubling route to chaos. Another non-linear phenomenon such as soliton propagation may also be found in a spatially extended Josephson junction, which is described by the socalled perturbed sine-Gordon equation [1]. Also other typical non-linear phenomena such as phase-locking, subharmonic generation, phase transitions etc. may be found in Josephson junction systems.

Secondly a most important aspect of the non-linear behaviour of Josephson junctions is that most of the above mentioned phenomena are accessible for experimental measurements. This has given a unique possibility of cooperation between researchers working with computer simulations on one side and researchers doing experimental work on the other. The benefit of such crossfertilisation has produced results that through the mutual interaction by far exceed the sum of the individual contributions. I hope to convincingly demonstrate this unique feature of Josephson junction physics in the examples presented on the following pages. The paper is organized in the following way. In section 2 the autonomous Josephson is described. In section 3 we discuss soliton propagation with one subsection describing the theory and numerical simulations and the other describing experiments. Chaos is discussed in section 4 with a similar subdivision between simulations and experiments. Conclusions are found in section 5.

## 2. The autonomous Josephson junction

The most widely investigated Josephson junction system is the current-driven Josephson junction, a description of which may be found in recent books on the subject [2,3]. The Josephson junction consists of two superconducting electrodes separated by a tunneling barrier. The equations with a bias current $I_{dc}$ may be written

$$\frac{C dV}{dt} + \frac{V}{R} + I_0 \sin \varphi = I_{dc} , \tag{1a}$$

$$\frac{d\varphi}{dt} = \frac{2eV}{\hbar} . \tag{1b}$$
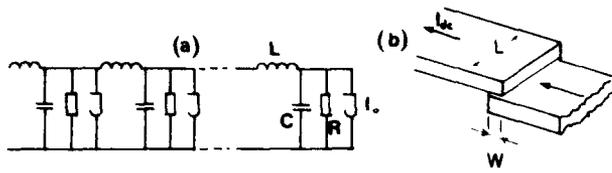
Fig. 1. Equivalent diagram (a) and schematic drawing of a long Josephson junction.

Equation (1b) is the famous Josephson frequency-to-voltage relation and eq. (1a) is Kirchoff's law applied to the Josephson junction equivalent circuit, a section as shown in fig. 1a. Here $\varphi$ is the pair phase difference across the junction, and $I_0$ is the maximum pair current. In addition, a shunt resistance $R$ carries a normal electron current ($V/R$), a capacitance $C$ carries a capacitive current $C(dV/dt)$, $\hbar$ is Planck's constant and $e$ is the electron charge. With time normalized to the reciprocal plasma frequency $\omega_0^{-1} = (\hbar C/2eI_0)^{1/2}$ and current normalized to the critical current $I_0$, these equations may be combined into a single dimensionless equation (the pendulum equation),

$$\varphi_{tt} + \alpha\varphi_t + \sin\varphi = \eta \, . \tag{2}$$

The plasma frequency, $\omega_0$, is the natural oscillation frequency for the Josephson junction, corresponding to the oscillation frequency in a pendulum. The McCumber parameter $\beta_C$ is given by $\beta_C = 2eR^2I_0C/\hbar$, and the damping parameter $\alpha$ is given by $(1/\sqrt{\beta_C})$. $\eta$ is the normalized (to $I_0$) dc bias current.

## 3. Soliton propagation

### 3.1. Long Josephson junctions: spatial dependence

As an extension of the simple Josephson junction let us now consider what happens when a spatial variation of the pair phase is allowed. We may get a soliton. Indeed, the long Josephson junction is one of the physical systems where

soliton propagation is accessible for direct experimental measurements. For the purpose of this presentation it is sufficient to note that the physical manifestation of the soliton is a fluxon, i.e., a quantum of magnetic flux $\Phi_0 = h/2e = 2.064 \times 10^{-15}$ Vs. Moving fluxons in the long Josephson junction manifest themselves as the so-called zero field steps (ZFS) in the dc current-voltage characteristic. Figure 1a shows the equivalent diagram and fig. 1b the geometry of a long junction of the overlap type. The corresponding wave equation may be written as a rather simple extension of eq. (2):

$$-\varphi_{XX} + \varphi_{tt} + \sin\varphi = \eta - \alpha\varphi_t \, , \tag{3}$$

with (normalized) voltage $V = \varphi_t$ (as in eq. (1b)) and (normalized) current $i = -\varphi_X$. The additional normalizations used are as follows. Length is measured in units of the Josephson penetration depth $\lambda_J = \sqrt{(\hbar/2de\mu_0 J)}$, i.e., $l = L/\lambda_J$. $J$ is the current density, and $d$ is the magnetic thickness of the junction. The junction length, $L$, is assumed large, and the width, $W$, is assumed small compared to the Josephson penetration depth. Finally, velocities become normalized to the velocity of light in the barrier, $\bar{c}$. For typical experimental junctions (of "old" superconductors), $\bar{c}$ is a few percent of its value in vacuum.

### 3.1.1. Perturbation calculation for the infinite Josephson line

The methods in this section are based mainly on the work of McLaughlin and Scott [1]. With the right hand side equal to zero, eq. (3) is the sine-Gordon equation. The loss and bias terms on the right hand side are considered as a perturbation to the sine-Gordon equation. The unperturbed sine-Gordon equation has the well known analytical single soliton solution [1,4],

$$\varphi = 4\tan^{-1} e^\theta \, , \tag{4}$$

where $\theta = (x - ut)\gamma(u)$, and $\gamma(u) = 1/\sqrt{(1 - u^2)}$ is the Lorentz factor. The solution gives rise to a

$2\pi$ phase shift over a length of a few $\lambda_J$, and its derivative $\varphi_t$ represents a voltage pulse. Note that the form of the solution – a traveling wave in the parameter $\theta = \gamma(u)(x - ut)$ – is a consequence of the Lorentz invariance of the sine-Gordon equation. In eq. (4) the velocity $u$ is a free parameter. As shown in refs. [1,4] the perturbation terms are included by assuming a solution of the same form as that in eq. (4), but with $u$ to be determined by a power balance equation. Requiring either the Hamiltonian or the momentum to be independent of time, one finds the velocity to be determined by [1,4]

$$u = 1/\sqrt{(1 + (4\alpha/\pi\eta)^2)} \,. \tag{5}$$

A more complicated solution to the sine-Gordon equation that may be perturbed under the influence of bias and losses is the soliton – antisoliton solution that may be written [1]

$$\varphi = 4\tan^{-1}(\sinh T/u \cosh X) \,, \tag{6}$$

where $T = u\gamma(u)\,t$ and $X = \gamma(u)\,x$. It is possible in a manner similar to that for the single soliton case to perform a power balance calculation by requiring the time rate of change of the energy $H$ to be zero, i.e., calculating the integral [4–6]

$$\frac{dH}{dt} = \int (\eta\varphi_t - \alpha\varphi_t^2)\,dx \,, \tag{7}$$

with $\varphi$ inserted from eq. (6). For this case a qualitatively new phenomenon occurs. For high incident energies the soliton and antisoliton will pass through each other with a phaseshift $\delta$ (spatial advance) given by

$$\delta = -2(1 - u^2)\ln u \,. \tag{8}$$

For bias below a certain threshold, $\eta_{th}$, to be calculated below, the soliton and antisoliton will annihilate each other, create a breather mode, and eventually die out as small amplitude damped plasma oscillations. In evaluating eq. (7)
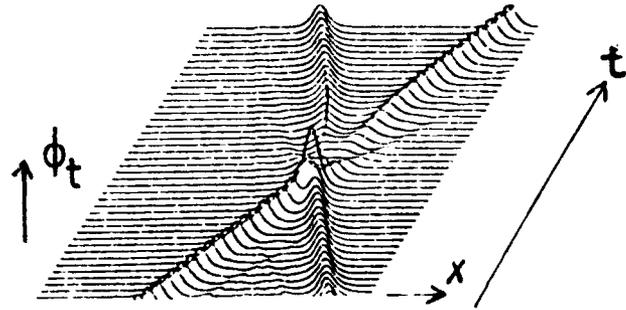


Fig. 2. Fluxon–antifluxon collision. $\alpha = 0.2$, $\eta = 0.22$, $l = 40$.

one finds after rather lengthy calculations [6] that the collision gives rise to an energy loss $\Delta H$, given by [6]

$$\Delta H \cong 4\pi^2\alpha \,. \tag{9}$$

Part of this energy is dissipated in propagating, but decaying, oscillations of the line. The annihilation threshold [6] may be found by requiring that the total energy of the soliton and antisoliton before the collision, $H = 16\gamma(u)$, is equal to the energy loss plus the rest energy of a stationary soliton and antisoliton, i.e.,

$$16\gamma(u) \approx 4\pi^2\alpha + 16\gamma(0) \,. \tag{10}$$

Equation (10) together with eq. (5) leads to [6]

$$\eta_{th} \approx (2\alpha)^{3/2} \,. \tag{11}$$

Figure 2 shows a numerical calculation of a soliton – antisoliton collision where both of the above phenomena – the energy loss and the phase shift – are easily observed.

### 3.1.2. The overlap and annular Josephson transmission line

In the overlap junction (fig. 1b) the long dimension of the junction is perpendicular to the current direction and the bias current is uniformly distributed over the junction length. Hence $\eta = I_0/JWL$ may be assumed in eq. (3). Due to the moving fluxon, a phase shift of $2\pi$ takes

place in a time interval $l/u$, where $l$ is the (normalized) length of the junction. This, in turn, gives rise to a dc voltage $v$, given by

$$v = (2\pi/l)u .\tag{12}$$

The overlap junction has boundary conditions requiring that no currents flow out at the ends, i.e.,

$$\varphi_x(0, t) = \varphi_x(l, t) = 0 .\tag{13}$$

It may be shown that this boundary condition is mathematically equivalent to a soliton – antisoliton collision, which was treated in the previous section.

The annular geometry on the other hand looks mostly like a long overlap Josephson junction that is folded back into itself; it has the simple periodic boundary conditions

$$\varphi_x(0, t) = \varphi_x(l, t) + 2p\pi ,\tag{14}$$

which for its simplicity is favoured by many authors. Here $p$ gives the number of full phase rotations along the line. For topological reasons $p$ is a conserved number as long as the superconducting ring is not opened up by taking the junction through the transition temperature [7–10], or by local hearing with an e-beam [11,12].

The simplest case to consider is $p = 1$, i.e. a single fluxon on the circular line. This case is shown in fig. 3a for the equivalent closed pendulum array [9]. In that case there is no supercurrent, since as soon as a uniform bias current is applied, the fluxon starts moving with a velocity $u$ determined by eq. (5) and a voltage given by eq. (12) develops. Hence for the annular junction with one trapped fluxon no supercurrent exists and the dc voltage is a direct measurement of the velocity of the single soliton. The IV curve is shown qualitatively in fig. 3b. In addition to the single soliton, further solitons may be created only by introducing soliton – antisoliton pairs (for topological reasons), in which case the
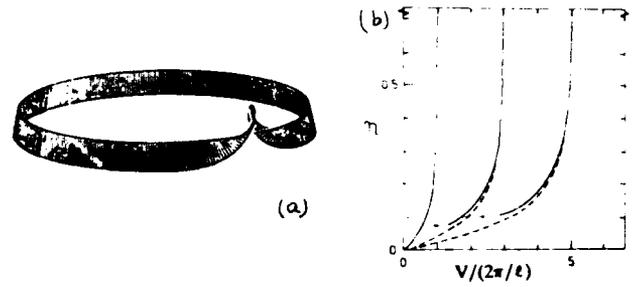


Fig. 3. (a) Equivalent pendulum array for the annular Josephson with one fluxon (from ref. [9]). (b) Schematic IV curves: Annular junction $p = 1$.

effects of collisions must be taken into account. Disregarding the collisions for simplicity, the voltages of these different ZFS are given by multiplying the one fluxon voltage with the total number $n$ of fluxons and antifluxons, i.e., voltage steps are to be expected at voltages $v_1$, $v_3 = 3v_1$, $v_5 = 5v_1$, etc. (shown as the dashed curves in fig. 3b). For the higher order branches, $v_3, v_5, \ldots$ the qualitative effect of the collisions is to lower the average velocity and voltage somewhat compared to $nv_1$, where $n = 3, 5, 7, \ldots$. Also a lower bias threshold $\eta_{th}$, where a fluxon and an antifluxon annihilate each other (cf. eq. (11)), is to be expected. Figure 3b shows qualitatively the higher order steps (full curves) based on these arguments. Reference [8] shows for $p = 2$ a numerically simulated IV curve which contains all the essential features of fig. 3b.

## 3.2. Soliton experiments: dc IV-curves

In the IV curve of an overlap junction the moving soliton gives rise to zero field steps (ZFS). The mechanism for the first ZFS, $n = 1$, is that a fluxon moves along the junction and is reflected at the boundary as an antifluxon. Since the reflection at $x = l$ is equivalent to a collision with a virtual antifluxon at $x = l$, the problem may be treated in the framework of eq. (8) for the phase shift and eq. (9) for the energy loss. If the junction length $l$ is very large, the details at the boundaries play only a minor role, and the
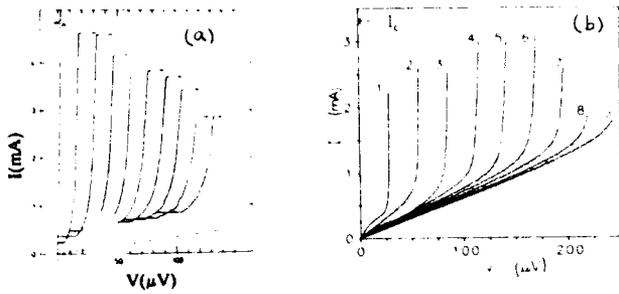
Fig. 4. (a) Experimental zero field steps for an overlap junction. (After ref. [5]). (b) Experimental *n*-fluxon zero field steps on an annular junction. *n* = 1–9 (after refs. [11,12]).

voltage of the first step is given by eq. (12). For example, Pedersen and Welner [5] were able to neglect completely the effects of collisions in a comparison between experimental ZFS on a very long overlap junction ($l = 45$) and perturbation theory. Figure 4a shows an example of one of their experimental curves. If the junction length is smaller (for example of the order 5–10), the energy loss and the phase shift will give rise to corrections.

Experimental measurements of solitons on the annular junction [9,10] have also been reported. In the experiment the *p* value could be changed only by taking the junction through the transition temperature. $p = 1$ appeared qualitatively as discussed above. $p = 0$ (zero fluxons trapped) showed the full supercurrent and fluxon – anti-fluxon steps at voltages $v_2 \approx 2v_1$, $v_4 \approx 4v_1$ ... etc.

In a more recent experiment [11,12] the introduction of fluxons could be made in a much more controlled way. Here an electron beam was used to open one side of the annular junction and introduce a fluxon. Each time the electron beam was crossing the junction a new fluxon was introduced. Passing the beam in the opposite direction removed a fluxon. Figure 4b shows an IV curve from this experiment [11,12], in which successively one up to nine fluxons were introduced. The annular junction experiment demonstrates in a very clean way the existence of a topological sine-Gordon soliton with a phase

change of $2\pi$. The experiment and interpretation are elegantly connected with fundamental theory which requires the phase to change only in multiples of $2\pi$ around a superconducting ring.

## 4. Chaos

### 4.1. Chaos in Josephson junctions

Another non-linear signature appearing quite often in Josephson junctions is chaos together with its accompanying bifurcations. In fact, quite often the Josephson junction is used as a model system for chaos in numerical simulations. A particular feature of chaos in Josephson junctions is that both the effect of thermal noise and the effect of deterministic noise (chaos) are very important in relation to the experiments. The interplay between those two sources of noise is very complicated, and at worst makes it difficult to interpret experiments. This has led to new theoretical and numerical work on the non-linear interaction between thermal and deterministic noise.

### 4.1.1. Deterministic chaos in the Josephson junction

The most widely investigated chaotic Josephson junction system is the rf-driven Josephson junction for which the equation may be obtained by adding a term $I_{rf}\sin\omega t$ to eq. (1a), i.e., in normalized units ($\Omega = \omega/\omega_0$)

$$\varphi_{tt} + \alpha\varphi_t + \sin\varphi = \eta_0 + \eta_1\sin\Omega t. \qquad (15)$$

Since analytical solutions do not exist, one has to do numerical simulations [13–18] in the parameter space of $\alpha$, $\eta_0$, $\eta_1$, and $\Omega$. A particularly thorough investigation of the parameter space was done in ref. [18]. Typically, the system has been investigated numerically in the $\eta_1$–$\Omega$ plane for a fixed damping parameter $\alpha$, a plot which more or less has become a standard for such systems. Such a parameter plane shows a very

complicated mapping of different dynamical be-
haviour [14]. For $\omega > \omega_0$, i.e., $\Omega > 1$, the
threshold rises because the capacitor shorts out
the applied rf current. For $\omega < 1/RC$, i.e.,
$\Omega < 1/\sqrt{\beta_C}$, the system is able to follow adiabati-
cally the rf current, and chaos occurs only if
$\eta_1 > 1$. For $\omega \sim \omega_0$, the threshold for chaos is
lowest.

Besides numerical simulation another method
of a somewhat computational nature is to use an
electronic analog simulating the Josephson equa-
tion [19]. Such systems have the advantage of
being very fast, and Poincaré sections and bifur-
cation diagrams may be readily displayed. The
disadvantage is the limited precision and res-
olution, and the drift of analog electronic cir-
cuits.



Fig. 5. Numerical calculation of phase locking steps of an IV
curve. $\Omega = 0.16$, $\alpha = 0.5$, $\eta_1 = 1.05$. After ref. [22].

### 4.1.2. Thermally affected chaos in the Josephson junction

In ref. [20] a Josephson junction system with
parameters such that two solutions existed, was
investigated. The authors found that the basin
boundaries between the two solutions were frac-
tal. A small amount of thermal noise may then
take the system back and fourth between the
solutions. The authors found that this mecha-
nism gave rise to approximately $1/f$ noise for
some parameter regions.

In an extensive numerical simulations [21] in-
cluding a thermal noise term in eq. (2), Kautz
was able to obtain the very high noise tempera-
tures ($\sim 10^6$ K) that have been observed ex-
perimentally. For a situation with overlapping
rf-induced steps, the origin of the very high noise
temperatures was hopping between phase locked
and metastable chaotic states induced by thermal
fluctuations. This may even lead to the result
that the low frequency noise power increases as
the temperature is reduced. In the absence of
thermal noise, numerical calculations of chaotic
regions in the IV curve typically contain a wealth
of complicated structure displaying bifurcations,
chaos, periodic solutions, etc. This may be seen
in fig. 5, which shows a numerical calculation of

an rf-induced step with loss of phase lock. In
experiments that have been performed, such in-
teresting and complicated structure have been
typically washed out because of thermal noise,
and only a smooth curve which does not in a
simple and convincing way demonstrate chaos, is
obtained [22]. Thermal smoothing due to a tem-
perature of less than 100 mK is sufficient to
remove most of the traces of complicated dy-
namical behaviour. By comparing experiments
with a calculation that includes thermal noise,
however, the existence of chaos may be shown
indirectly.

### 4.2. Experiments on real Josephson junction

Common to all the experimental results is that
they are not nearly as spectacular as the numeri-
cal simulations. The main reason is that thermal
noise, which is most often not taken into account
in simulations, has a major effect on the out-
come of the experiments. This is because the
energy levels in the thermal oscillations may very
well be of the same order of magnitude as the
intrinsic energy levels in the Josephson junc-
tions, and complicated non-linear interactions
occur.

## 4.2.1. dc observation of chaos

Before the term chaos was connected to Josephson junctions, researchers sometime noted very irregular and erratic IV curves in samples subject to strong applied rf signals. In many cases such junctions were discarded because of assumed defects during fabrication. It is known that such irregular IV curves may be a signature of chaos. Examples of such irregular behaviour, in particular the loss of phaselock on an rf-induced step, may be found in refs. [22–24].

By comparing such experimentally obtained, irregular IV curves to numerically obtained ones, one has in principle the simplest experiment on chaos. An example of such a dc experiment is illustrated in fig. 6, which shows an experimentally obtained IV curve with loss of phaselock on the rf-induced step. Also shown in the figure is the spectrum of half-harmonic generation as measured with a sensitive microwave spectrometer. Note that these experimental curves, which contain two period-doubling bifurcations and a chaotic region on an rf-induced step, can be considered as a standard example of the period-doubling route to chaos. There experimental results are very similar to the numerical results shown in fig. 1 of ref. [17] and to analog results.
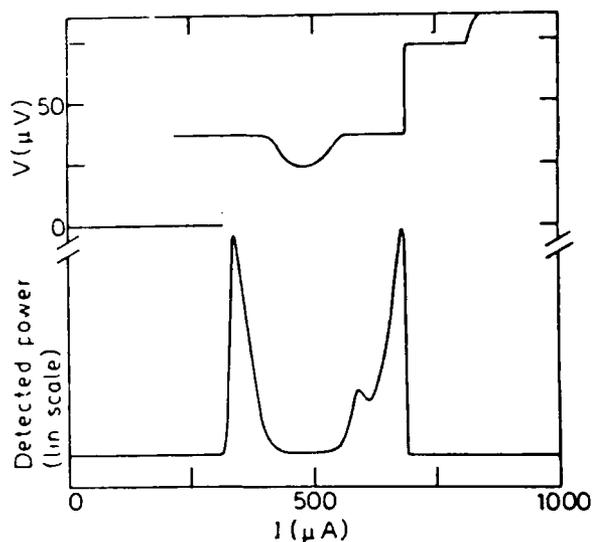
## 4.2.2. Chaos and parametric amplification

For Josephson junction parametric amplifiers, low noise temperatures were found in some cases, however, more often experiments showed considerable excess noise [25]. For experimentalists the observed noise rise has been a major puzzle. A large number of theoretical and numerical papers [26] have dealt with the problem. The conclusion is that the very large noise temperatures cannot be explained by traditional noise sources such as Johnson noise, shot noise, or quantum noise. Hubermann et al. [13] first suggested that chaos was the origin of the excess noise. This suggestion was further substantiated in ref. [14].

More recently a slightly different type of Josephson junction parametric amplifier was investigated [27] by another group, which drew the conclusion that noise in this amplifier cannot arise from deterministic chaos alone. The observed noise increase required the presence of thermal noise. They suggested that the noise increase was due to thermally induced hopping between a bias point that would be stable in the absence of thermal noise and an unstable point. This observation demonstrates the importance of thermal noise in modeling chaos in Josephson junction systems. A noise temperature of as much as $10^6$ K may be obtained.

## 5. Conclusion

This paper has discussed mainly the non-linear properties of Josephson junctions, which have so much promise for both applications and continued research on fundamental problems. The problems we have dealt with have all been defined on the basis of the "old" superconductors. Future work involving the new high $T_c$ superconductors will most likely deal not only with the same type of problems as discussed above for



Fig. 6. IV curve a and half harmonic microwave power. Experiments after ref. [24].

different parameters, but also with completely new non-linear phenomena, for example, due to anisotropy and complicated flux dynamics. It may be safely predicted that a lot of interesting non-linear physics lies ahead.

# References

[1] D.W. McLaughlin and A.C. Scott, Phys. Rev. A 18 (1978) 1652.

[2] A. Barone and G. Paterno, Physics and Applications of the Josephson Effect (Wiley – Interscience, New York, 1982).

[3] K.K. Likharev, Dynamics of Josephson junctions and Circuits (Gordon and Breach, New York, 1986).

[4] N.F. Pedersen, in: Solitons MPCMS, eds. A.A. Maraduddin and V.H. Agranovich (North Holland, 1986) p. 469.

[5] N.F. Pedersen and D. Welner, Phys. Rev. B 29 (1984) 2551.

[6] N.F. Pedersen, M.R. Samuelsen and D. Welner, Phys. Rev. B 30 (1984) 4057.

[7] P.M. Marcus and Y. Imry, Sol. Stat. Commun. 33 (1980) 345;
S.E. Burkov and A.E. Lifsic, Wave Motion 5 (1983) 197.

[8] A. Davidson and N.F. Pedersen, Appl. Phys. Lett. 44 (1984) 465.

[9] A. Davidson, B. Dueholm and N.F. Pedersen, J. Appl. Phys. 60 (1986) 1447.

[10] A. Davidson, B. Dueholm, B. Kryger and N.F. Pedersen, Phys. Rev. Lett. 55 (1985) 2059.

[11] A.V. Ustinov, T. Doderer, B. Mayer, R.P. Huebener and V.A. Oboznov, Europhys. Lett. (1992).

[12] A.V. Ustinov, T. Doderer, R.P. Huebener, N.F. Pedersen, B. Mayer and V.A. Oboznov, submitted to Phys. Rev. Lett. (1992).

[13] B.A. Huberman, J.P. Crutchfield and N.H. Packard, Appl. Phys. Lett. 37 (1980) 750.

[14] N.F. Pedersen and A. Davidson, Appl. Phys. Lett. 39 (1981) 830.

[15] D. D'Humieres, M.R. Beasley, B.A. Huberman and A. Libchaber, Phys. Rev. A 26 (1982) 3483.

[16] R.L. Kautz, J. Appl. Phys. 52 (1981) 3528;
M. Octavio, Phys. Rev. B 29 (1984) 1231;
K. Sakai and Y. Yamaguchi, Phys. Rev. B 30 (1984) 1219.

[17] R.L. Kautz, J. Appl. Phys. 52 (1981) 6241.

[18] R.L. Kautz and R. Monaco, J. Appl. Phys. 57 (1985) 875.

[19] M. Cirillo and N.F. Pedersen, Phys. Lett. A 90 (1982) 150;
H. Seifert, Phys. Lett. A 98 (1983) 213;
D.R. He, W.J. Yeh and Y.H. Kao, Phys. Rev. B 31 (1985) 1359.

[20] M. Iansiti, Q. Hu, R.M. Westervelt and M. Tinkham, Phys. Rev. Lett. 55 (1985) 746.

[21] R.L. Kautz, J. Appl. Phys. 58 (1985) 424.

[22] D.C. Cronemeyer, C.C. Chi, A. Davidson and N.F. Pedersen, Phys. Rev. B 31 (1985) 2667.

[23] K. Okuyama, H.J. Hartfuss and K.H. Gundlach, J. Low Temp. Phys. 44 (1981) 283.

[24] N.F. Pedersen, O.H. Sørensen, B. Dueholm and J. Mygind, J. Low Temp. Phys. 38, 1 (1980).

[25] R.Y. Chiao, M.J. Feldmann, D.W. Peterson, B.A. Tucker and M.T. Levinsen, AIP Conf. Proc. 44 (1978) 259.

[26] For references on this see e.g. N.F. Pedersen, in: SQUID 80, eds. H. Hahlbohm and H. Lübbig (de Gruyler, Berlin, 1990) p. 739.

[27] R.F. Miracky and J. Clarke, Appl. Phys. Lett. 43 (1983) 508.

**PHYSICA** Ⅾ

# Subharmonic self-locking of a Josephson soliton oscillator coupled to a resonator

J. Holm [a,b], Paola Barbara [a,c], A. Davidson [d], G. Filatrella [c], J. Mygind [a] and N.F. Pedersen [a]

[a] Physics Laboratory 1, B309 Technical University of Denmark, DK-2800 Lyngby, Denmark
[b] Danish Institute of Fundamental Metrology, B307 Lundtoftevej 100, DK-2800 Lyngby, Denmark
[c] Department of Physics, University of Salerno, I-84100 Salerno, Italy
[d] IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

Josephson soliton oscillators integrated in a resonator consisting of two closely spaced superconducting microstrips have been investigated experimentally. The dc IV-characteristic shows zero-field steps containing a number of resonator-induced substeps, occurring because of locking of the soliton oscillator to the resonator. The position of the substeps follows the structure of a "devil's staircase", and their appearance depends strongly on the externally applied dc magnetic field.

## 1. Introduction

The study of the dynamical behaviour of long Josephson junctions is of great fundamental and practical interest. Besides showing a rich variation of nonlinear properties and chaotic regimes, they have potential applications as electronic devices, such as local oscillators for SIS-mixers in the realization of cryogenic integrated millimeter wave receivers for use in radio-astronomy and space communication [1].

The one-dimensional Josephson soliton oscillator is a Josephson tunnel junction which is much longer than the Josephson penetration depth, $\lambda_J$, in one direction and comparable to or less than $\lambda_J$ in the other direction. When the junction is properly dc biased, a localized quantum of magnetic field – a soliton – can oscillate back and forth in the junction, hereby emitting radiation in the millimeter wave frequency range [2].

The junctions discussed here are integrated in a resonant structure consisting of two closely spaced coplanar superconducting microstrips as
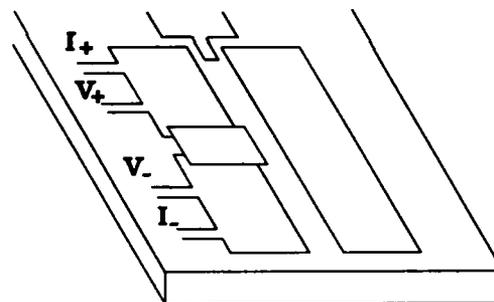


Fig. 1. Schematic drawing of the sample geometry. The coplanar resonator has its electromagnetic field confined in the gap between the two closely spaced Nb-strips. The two Josephson junctions are formed where the top and bottom films of the left strip overlap. The microstrip antenna shown at the top of the figure is used for the microwave measurements.

shown in fig. 1. Earlier experiments [3] have shown locking between one soliton oscillator and the resonator manifesting itself as substeps on the zero-field steps in the IV-characteristic. Here we report on a more detailed study of these substeps in order to clearify the nature of the interaction between the soliton oscillator and the resonator.
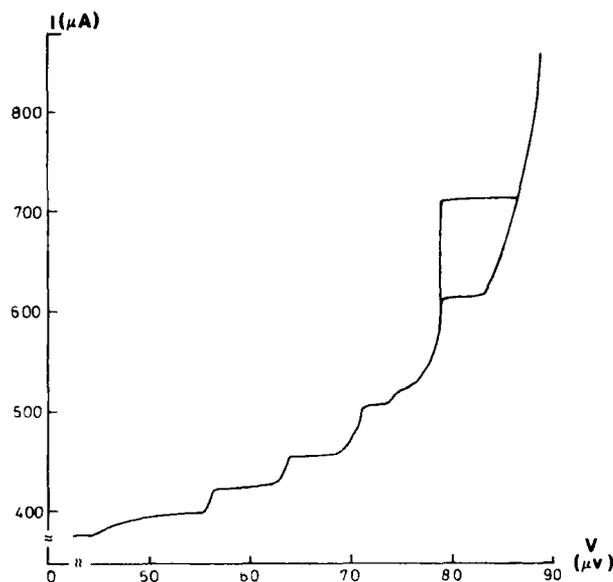
Fig. 2. Lower part of the ZFS1 for a four-junction sample (U1-4JJ). $T = 4.2$ K, $B = 0$.

## 2. Experiments

The samples were fabricated in Nb/AlOx/Nb trilayer technology with a critical current density of $10^3$ A/cm$^2$ as described elsewhere [4]. The dimensions of the junctions are 200 $\mu$m by 10 $\mu$m and they are closely spaced in pairs placed in the middle of one of the Nb-strips (see fig. 1). The critical current was 10 mA. The Josephson penetration depth is $\lambda_J \simeq 10$ $\mu$m giving a maximum plasma frequency of $f_p \simeq 130$ GHz. The maximum voltage of the first zero-field step (ZFS1) is approximately 100 $\mu$V corresponding to a soliton oscillator frequency of 24 GHz. The variation in the measured parameters was less than 5%.

Fig. 2 shows the lower part of ZFS1 for one of the junctions in a four-junction sample. A hysteretic step at 79 $\mu$V, corresponding to a soliton frequency of 19 GHz, is present, but a series of minor steps can also be seen at lower voltages. All the steps are interpreted as coupling between the soliton oscillator and the external resonator [5]. By applying a dc magnetic field in the plane of the junction and thereby changing the coupling,

we noted that the substeps were approximately one order of magnitude more sensitive to the magnetic field than the critical current.

## 3. Discussion

The dominant mode in the present geometry is the coplanar resonator mode, where the millimeter wave electric and magnetic fields are mainly confined to the narrow space between the strips.

In order to model the geometry used in ref. [1], Grønbech-Jensen et al. developed a simple analytical approach [6], where the overlap junctions are coupled to a lumped element linear tank circuit. The model predicts locking between the soliton and the tank circuit when the junction is operated near the fundamental resonance frequency of the tank or its subharmonics [7]. This has been confirmed numerically. The height of the subharmonic steps was seen to vary approximately inversely proportional to the subharmonic number [7].

In contrast to this model, the distributed resonator in our setup will exhibit resonant behaviour at all multipla $n$ of one half-wavelength giving rise to locking whenever the soliton frequency equals a fractional harmonic $n/m$ of the half-wavelength resonance frequency $\nu_c$ of the coplanar mode. Indeed, the position of the steps in fig. 2 corresponds very well to $n/m = 2/5$ and $n/m = 4/9$, assuming that the hysteretic step corresponds to $n/m = 1/2$.

The height of the substeps decreased with increasing temperature. At a temperature where the substeps were completely suppressed, the junction was irradiated with microwaves via the microstrip antenna shown at the top of fig. 1, and rf-induced phase-locked steps were observed. Not only locking to the fundamental frequency was seen, but also locking to fractional harmonics, following the structure of a "devil's staircase" with locking at $n/m = \{1/3, 2/5, 3/7, 4/9, 5/11, \ldots, 1/2\}$ was observed. Comparing to fig. 2, the 3/7 and 5/11 harmonics
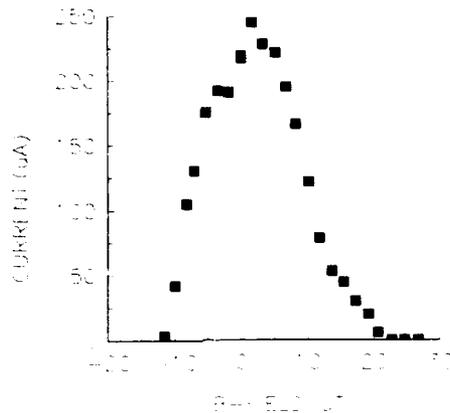
Fig. 3. Curve showing the height of the hysteretic step in a two-junction sample (U3-2JJ). The first minimum of the critical current vs dc magnetic field occurs at a much larger field ($B = 320\,\mu T$). $T = 4.2$ K.

are missing. However by applying a small external dc magnetic field also these steps actually appeared in the $IV$-curve.

Another geometry with only one pair of junctions showed the same behaviour. A large hysteretic step occured at 75 $\mu V$, together with a number of other substeps. By comparing the different junctions, it was noted that the structure scaled with the position of the large substep. The symmetry of the "devil's staircase" around $1/2$ was confirmed by the following sequence of observed substeps: $n/m = \{1/2, 6/11, 5/9, 4/7, 3/5\}$. For even values of $n$, it was necessary to apply a small dc magnetic field to observe the substeps. The height of the substep was smaller for larger $n$ in agreement with theory. Fig. 3 shows the height of the hysteretic step for this sample as function of applied magnetic field.

## Acknowledgements

## References

[1] S. Pagano, R. Monaco and G. Costabile, Microwave oscillator using arrays of long Josephson junctions, IEEE Trans. Magn. MAG-25 (1989) 1080.

[2] D.W. MacLaughlin and A.C. Scott, Perturbation analysis of fluxon dynamics, Phys. Rev. A 18 (1978) 1652.

[3] A. Davidson, N. Grønbech-Jensen and N.F. Pedersen, Josephson soliton oscillator arrays for SIS-mixers, IEEE Trans. Magn. MAG-27 (1991) 3347.

[4] M.B. Ketchen, D. Pearson, A.W. Kleinsasser, C.-K. Hu, M. Smyth, J. Logan, K. Stawiasz, E. Baran, M. Jaso, T. Ross, K. Petrillo, M. Manny, S. Basavaiah, S. Brodsky, S.B. Kaplan, W.J. Gallagher and M. Bhushan, Sub-$\mu$m, planarized, Nb-AlOx-Nb Josephson process for 125 mm wafers developed in partnership with Si technology", Appl. Phys. Lett. 59 (1991) 2609.

[5] P. Barbara, A. Davidson, G. Filatrella, J. Holm, J. Mygind and N.F. Pedersen, Coupling of a Josephson soliton oscillator to coplanar and microstrip cavities, Phys. Lett. A 165 (1992) 241.

[6] R. Monaco, N. Grønbech-Jensen and R.D. Parmentier, Self-locking of fluxon oscillators in series arrays of niobium Josephson tunnel junction coupled to a linear resonator, Phys. Lett. A 151 (1990) 195.

[7] G. Filatrella, N. Grønbech-Jensen, R. Monaco, S. Pagano R.D. Parmentier, N.F. Pedersen, G. Rotoli, M. Salerno and M.R. Samuelsen, in: Nonlinear Superconductive Electronics and Josephson Devices, G. Costabile, S. Pagano, N.F. Pedersen and M. Russo, ed. (Plenum Press, New York, 1991).

PHYSICA D

# Soliton bunching in annular Josephson junction

M.P. Sørensen [a], B.A. Malomed [b], A.V. Ustinov [c,d] and N.F. Pedersen [d]

[a] *Laboratory of Applied Mathematical Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark*
[b] *Department of Applied Mathematics, School of Mathematical Sciences, Tel Aviv University, Ramat Aviv 69978, Israel*
[c] *Institute of Solid State Physics, Russian Academy of Sciences,*
 *Chernogolovka, Moscow district, 142432, Russian Federation*
[d] *Physics Laboratory I, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Numerical simulations have been performed for an annular Josephson junction with a normalized length equal to 6, using the perturbed sine-Gordon equation. The calculated current–voltage characteristics reveal a small jump on the steps 3 and 4 in accordance with experimental observations. The computer simulations show that the jumps are caused by a transition from a non-bunched to a bunched fluxon oscillation state as the bias current is increased.

The annular Josephson junction allows for the study of undisturbed motion of sine-Gordon solitons (magnetic flux quanta) in the absence of boundaries and soliton–soliton collisions. Recently the dynamics of a soliton chain trapped in an annular Josephson junction has been investigated experimentally [1]. The experiment showed the evidence for a crossover in the soliton chain velocity $v$ dependence versus the driving force $\gamma$ ($I$–$V$ characteristics), which has been suggested to be due to soliton bunching [2] at high velocities. The surprising fact observed in the experiment was that the crossover has been found only for the configurations with $n = 3$ and 4 solitons trapped in the annular Josephson junction of the normalized length $l \approx 6$. For the cases of $n = 2$ and $n > 4$ no specific features have been observed in the $I$–$V$ curves.

In this contribution we report on our recent numerical simulations, which where performed in order to check if the behavior which was observed in the experiment [2] can be explained by the soliton bunching effect.

The system discussed here is described by the perturbed sine-Gordon equation [3],

$$\varphi_{xx} - \varphi_{tt} = \sin\varphi + \alpha\varphi_t - \beta\varphi_{xxt} + \gamma, \qquad (1)$$

where $\varphi(x,t)$ is the space and time dependent superconducting phase difference. The spatial coordinate $x$ is normalized to the Josephson penetration depth $\lambda_j$, the time $t$ to the inverse plasma frequency $\omega_0^{-1}$, $\alpha$ and $\beta$ are the dissipation coefficients, and $\gamma$ is the bias current. For the annular junction, the boundary conditions for eq. (1) are periodic,

$$\varphi(l,t) = \varphi(0,t) + 2\pi n,$$

$$\varphi_x(l,t) = \varphi_x(0,t). \qquad (2)$$

As known from previous studies [4], the presence of the surface-loss term $\beta$ in eq. (1) leads to a spatially oscillating trailing tail behind a moving soliton, which appear at high soliton velocity $v_s$. This may give rise to bunched (bound) states of solitons in the chain, thereby breaking the symmetry in the system.

For solving eq. (1) with the boundary conditions (2) we were using a pseudo-spectral Fourier method [5]. The parameters have been chosen close to the experimental values [1]: $l = 6.0$, $\alpha = 0.02$, $\beta = 0.02$, $n = 1, 2, ..., 6$. In order to achieve a stationary dynamic state, for each value of the bias current $\gamma$ the integration was running to let the solitons make from 60
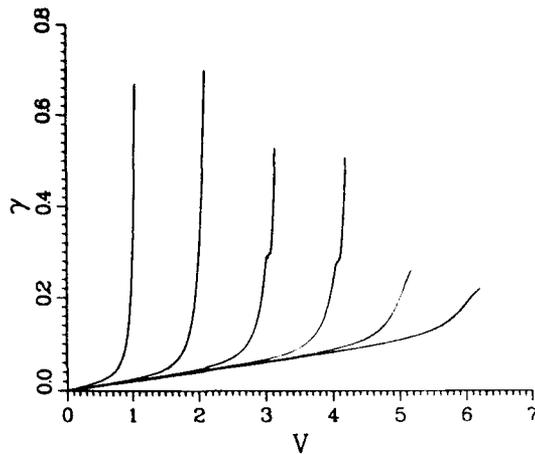
Fig. 1. The calculated *I–V* curves of the annular Josephson junction with different numbers $n = 1,...,6$ of trapped ions.
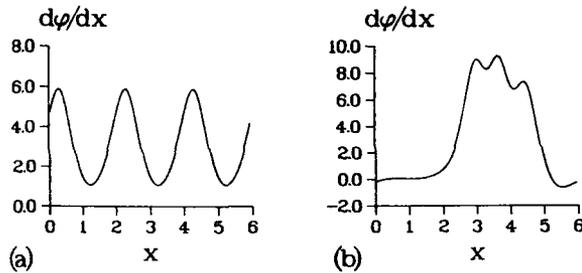


Fig. 2. The spatial distribution of the normalized magnetic field in the junction on step 3 at $\gamma=0.26$ (a) and at $\gamma=0.34$ (b).



Fig. 3. The spatial distribution of the normalized magnetic field in the junction on step 4 at $\gamma=0.26$ (a) and at $\gamma=0.34$ (b).

up to 1200 revolutions in the system. After that the average voltage $v=\langle\phi_t\rangle$ was calculated using both temporal and spatial averaging.

The normalized *I–V* characteristics $(\gamma-v)$ have been calculated for a different number of solitons $n$ in the system. The results are shown in fig. 1. Apparently, this collection of curves looks quite similar to fig. 2a of experiment [1]. In a nice agreement with the experiment, the crossover manifested by small steps on the curves for $n = 3$ and 4 solitons is very pronounced (compare with fig. 2b of experiment [1]).

The spatial magnetic field distributions $\varphi_x(x, t)$ for step 3 at $\gamma=0.26$ and at $\gamma=0.34$ are shown in fig. 2. It is apparent that below

the small step in the $\gamma-v$ curve, the three solitons are equidistantly spaced (the system has a translational symmetry in the $x$-direction with a period of $l/n$) and accordingly the solitons are not bunched. However, above the small step we observe bunching of the solitons. A similar crossover from non-bunched to bunched solitons are detected for the step 4 as the bias current $\gamma$ is increased (see fig. 3). These pictures prove that the crossover we observe in the experiments is due to the bunching of solitons in the junction. The bunching effect "helps" the chain of solitons to overcome the dissipative losses. Their average velocity becomes higher than that of a single soliton at the same bias current.

*Discussion.* At the crossover from the equidistantly spaced solitons to the bunched soliton dynamic mode, a jump to higher voltages occurs. The experiments [1] show a relative increase in the voltage which is less than 1%, whereas in the numerical simulations this relative voltage jump is about 2%. This discrepancy in the numerical simulations may be attributed to the influence of the neglected spatial extension in the $y$-direction of the annular junction where the inner circumference is smaller than the outer one.

In the numerical simulations the step hights follow the pattern seen in the experiments. The second step is higher than the first and the third one. Similarly, the form of the steps 5 and 6 in the numerical simulations resembles very closely

the form of the steps 8 and 9 in the experiment shown in ref. [1]. The fact that three more steps are seen in the experiments indicates that the experimental junction is 9 times the magnetic penetration depth and not 6 as anticipated in the numerical simulations.

We are grateful to Alwyn C. Scott and Niels Grønbech-Jensen for stimulating discussions.

## References

[1] A.V. Ustinov, T. Doderer, I.V. Vernik, N.F. Pedersen, R.P. Huebener and V.A. Oboznov, Physica D 68 (1993) 41, in these proceedings.
[2] A.C. Scott and W.J. Johnson (1968), unpublished.
[3] D.W. McLaughlin and A.C. Scott, Phys. Rev. A 18 (1978) 1652.
[4] A. Davidson, N.F. Pedersen and S. Pagano, Appl. Phys. Lett. 48 (1986) 1306.
[5] F. If, M.P. Sørensen and P.L. Christiansen, Phys. Lett. A 100 (1984) 68.

PHYSICA D

# Experiments with solitons in annular Josephson junctions

A.V. Ustinov[a,b,c], T. Doderer[b], I.V. Vernik[a], N.F. Pedersen[c],
R.P. Huebener[b] and V.A. Oboznov[a]

[a]Institute of Solid State Physics, Russian Academy of Sciences,
142432 Chernogolovka, Moscow district, Russian Federation
[b]Lehrstuhl Experimentalphysik II, University of Tübingen, W-7400 Tübingen, Germany
[c]Physics Laboratory I, Technical University of Denmark, DK-2800 Lyngby, Denmark

It is known that the long annular Josephson junction allows to avoid the influence of boundaries and soliton(fluxon)–soliton collisions on the dynamics of soliton chains. With new experimental methods and junction design a fully controlable way of introducing individual solitons into such a system is realized. For the homogeneous junction crossover at high velocities of the fluxon chain is seen and explained in terms of the fluxon bunching. In the periodically modulated junction a new resonant mode called a supersoliton is observed.

## 1. Introduction

Josephson junction fluxons represent a well-known example of nonlinear electro-magnetic waves [1]. The system is described by the perturbed sine–Gordon equation [2]. In case of conventional topology of a long Josephson junction, a moving soliton (magnetic flux quantum, or fluxon) cannot avoid collisions with other solitons and the junction boundaries. This strongly complicates the analysis and interpretation of the experimental data.

This paper emphasizes new results obtained for the ideal experimental system – the annular Josephson junction [3]. In the annular junction one may avoid fluxon collisions and study in a controllable way the motion of individual fluxons. With new experimental methods and junction design we have realized for the first time *a fully controllable way of introducing individual solitons* into such a system. This has allowed us to study multisoliton dynamics without collisions with boundaries and to make a comparison with existing soliton chain perturbation theory. Two types of junctions, homogeneous and spatially modulated rings, have been investigated.

## 2. Experimental methods of fluxon trapping

Annular junctions with the inner radius of the ring $100\,\mu m$ and with the width of the tunnel barrier ring $20\,\mu m$ were used. The critical current density was about $40\,A/cm^2$, which corresponds to the characteristic size of a fluxon $\lambda_j \approx 70\,\mu m$. In order to achieve a high homogeneity of the tunnel barrier it was formed in the area between two insulating SiO rings covering the edges of the bottom electrode. On each substrate together with *homogeneous* junctions we have fabricated several *modulated* annular LJJ. In order to provide a periodic modulation of the critical current density along the LJJ, the regular lattice of inhomogeneities was made using SiO strips placed across the junction.

We have realized experimentally two different techniques for trapping the magnetic flux in the junctions. First, using low temperature scanning electron microscopy (LTSEM) we were able to trap one by one several magnetic flux quanta in the tunnel barrier area of the Josephson junction. Second, by applying the external monopole-like magnetic field (MMF) using a special coil we trapped the flux in the junction at the
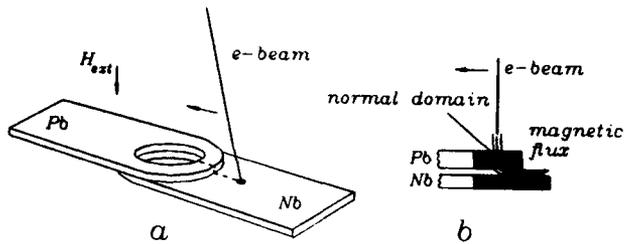
Fig. 1. (a) The magnetic flux trapping procedure in an annular junction using LTSEM. (b) The cross-section of the junction along the dashed line shown in (a) at the moment of crossing by the electron beam (the insulating SiO window is not shown).

temperature near the critical temperature of the superconducting electrodes.

In the LTSEM method, the focused electron beam was used for trapping magnetic flux quanta in the annular junctions. The crossing of the top electrode by the electron beam focus resulted in trapping the magnetic flux in the tunnel barrier (fig. 1a). This was due to the local heating of the sample in the focus of the beam. A normal domain carrying the magnetic flux was moved across the Pb film together with the electron

beam focus as shown in fig. 1b. The details of this technique are presented elsewhere [4].

In the MMF method we have used a current-carrying ring above the junction in order to produce a radial magnetic field component along the junction circumference. The magnetic flux happens to be trapped in the junction at the temperature close to the critical temperature of the superconducting electrodes. The details are described elsewhere [5].

## 3. Results and discussion

Fig. 2a shows the experimental $I-V$ characteristics for the trapping of solitons in the annular Josephson junction obtained by a sequential crossing of the junction with the electron beam. Typically, each crossing resulted in trapping a one flux quantum in the ring. The critical current of the junction without trapped flux was about 3.3 mA. The resonant soliton steps corre-
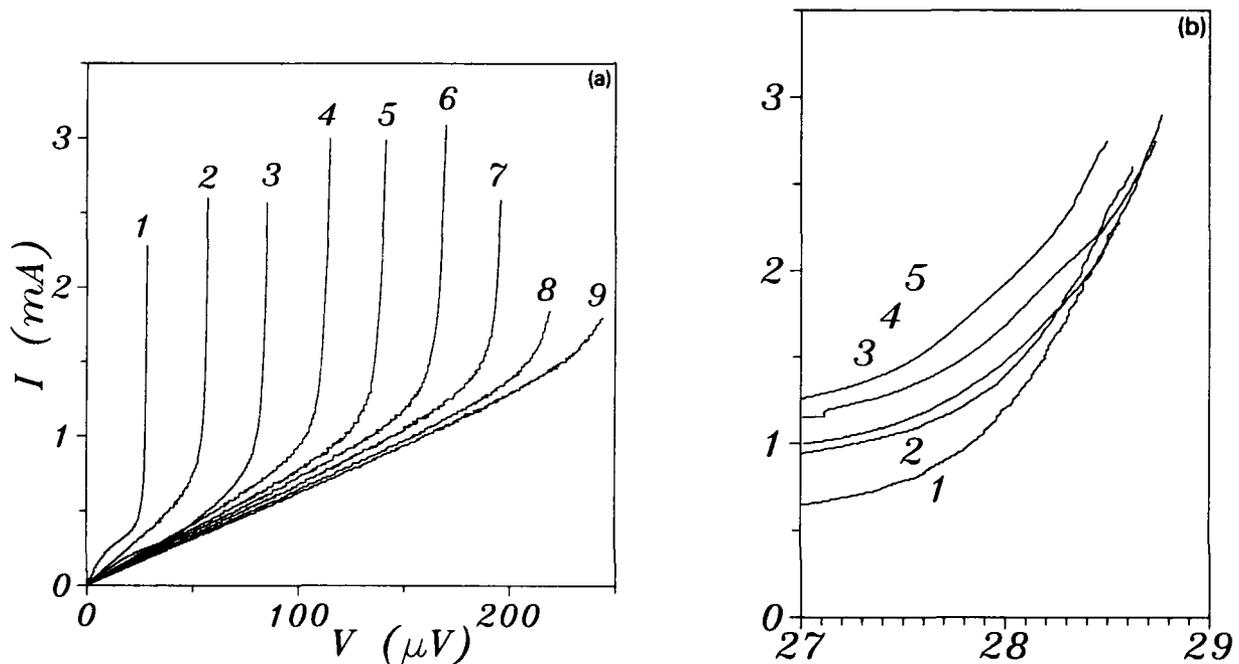


Fig. 2. (a) The $I-V$ curves corresponding to the sequential trapping of solitons in the homogeneous annular Josephson junction using the LTSEM method. The numbers correspond to the number of solitons trapped. (b) The detailed comparison of the high-current parts of the scaled soliton steps in a magnified voltage scale (the voltage of the $n$th step is divided by $n$).

sponding to different numbers of unipolar solitons $n$ are observed. The voltages of the steps are quantized as expected, according to the formula: $V_n = n\Phi_0\bar{c}/L$, where $\Phi_0$ the magnetic flux quantum, $\bar{c}$ the maximum velocity of the electromagnetic waves in the junction, and $L$ is the junction circumference. At the step voltage $V_n$ the solitons move as relativistic particles with the velocity close to $\bar{c}$. If solitons are trapped in an ideally homogeneous annular junction, theoretically the non-dissipative critical current $I_c$ should be equal to zero. This is perfectly fulfilled for all the curves in fig. 2a and is a good proof for the high homogeneity of our junctions. The multi-soliton states for the perturbed sine-Gordon system with periodic boundary conditions were analyzed by Marcus and Imry [6]. The tendency to reach an asymptotic linear slope with the increasing number $n$ of solitons predicted from theory [6] is clearly seen in the experiment.

Fig. 2b shows the current plotted versus the voltage per fluxon $V_n/n$ and displays the top parts of the steps in a magnified voltage scale. Since $V_n/n$ is proportional to the average fluxon velocity $v$, we can compare the dependence of $v$ on the external driving force $I$ for different fluxon densities. With increasing step number of crossover in the step voltage is clearly seen. At certain fluxon velocities the steps with $n = 3$ and $n = 4$ display a kind of shift to higher voltages, which is not expected from theory [6]. As it was known already from early computer simulations (A.C. Scott and W.J. Johnson, unpublished, 1968), the dissipation due to the surface ($\beta$) losses [7,8] at high velocity generates an effective attraction between unipolar solitons and results in bunching in a localized soliton train moving in the junction. This bunching effect breaks the symmetry in the annular system and "helps" the chain of solitons to overcome the dissipative losses. Their average velocity then becomes higher than that of a single soliton at the same bias current. Recent numerical simulations [9] have shown a nice agreement with our experi-

ment, also displaying the crossover at high velocity for the cases of 3 and 4 solitons.

In periodically modulated rings we observe a new resonant mode, called the *supersoliton mode*, which has been found before in numerical simulations [10] and described theoretically [11]. This mode is justified by the additional voltage steps in the current–voltage characteristics of the junction, which appear when the spatial periods of solitons and the junction modulation are incommensurate. Fig. 3 shows the $I$–$V$ characteristics of the modulated annular junction with $N = 5$ inhomogeneities (the modulation period is equal to $\frac{1}{5}$ of the junction circumference) and $n = 6$ solitons trapped in it using the MMF method. Each inhomogeneity represents a region in the junction where the critical current density is suppressed, so it attracts a soliton. It is energetically advantageous to place the 6-th soliton as a defect (supersoliton) on already pinned regular lattice of 5 other solitons. Under the influence of bias current this defect (supersoliton) moves along the junction, while all the other solitons remain being pinned. In the $I$–$V$ curve this mechanism leads to the pronounced first step at $V \approx 32\mu V$ in fig. 3. The detailed explanation for the features observed in the $I$–$V$ curves may be found elsewhere [5].
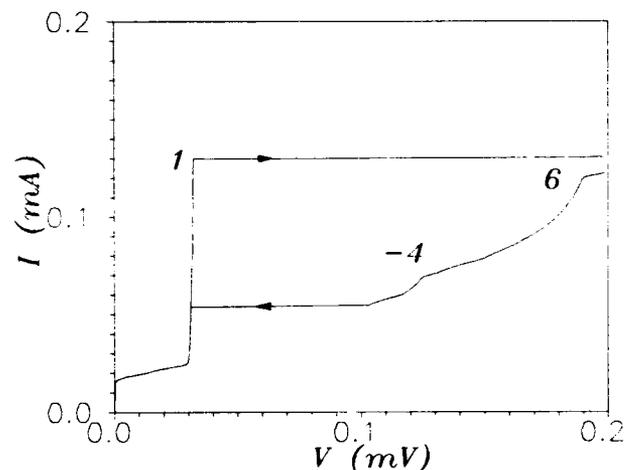


Fig. 3. The $I$–$V$ curve of the junction with $N = 5$ inhomogeneities and $n = 6$ trapped solitons using the MMF method.

## Acknowledgement

## References

[1] R.D. Parmentier, Solitons in physical systems: Josephson junction fluxons, Presented at this meeting.

[2] D.W. McLaughlin and A.C. Scott, Phys. Rev. A 18 (1978) 1652.

[3] A. Davidson, B. Dueholm, B. Kryger and N.F. Pedersen, Phys. Rev. Lett. 55 (1985) 2059.

[4] A.V. Ustinov, T. Doderer, B. Mayer, R.P. Huebener and V.A. Oboznov, Europhys. Lett. 19 (1992) 63.

[5] I.V. Vernik, V.A. Oboznov and A.V. Ustinov, Phys. Lett. A 168 (1992) 319.

[6] P.M. Marcus and Y. Imry, Solid State Commun. 33 (1980) 345.

[7] A.C. Scott, Solid State Electronics 7 (1964) 137.

[8] A. Davidson, N.F. Pedersen and S. Pagano, Appl. Phys. Lett. 48 (1986) 1306.

[9] M.P. Sørensen, B.A. Malomed, A.V. Ustinov and N.F. Pedersen, Soliton bunching in annular Josephson junction, Physica D 68 (1993) 38, these proceedings.

[10] A.V. Ustinov, Phys. Lett. A 136 (1989) 155.

[11] B.A. Malomed, Phys. Rev. B 41 (1990) 2616.

**PHYSICA** $\square$

# H-localized mode in chains of hydrogen-bonded amide groups

Mariette Barthes[a], Hassan Kellouai[a], Gabriel Page[a], Jacques Moret[a],
Susanna W. Johnson[b] and Juergen Eckert[b]

[a]GDPC – cc 26 – Université Montpellier II, 34095 Montpellier cedex 5, France
[b]Lansce – Los Alamos National Laboratory, Los Alamos, NM 87545, USA

New infrared measurements of the anomalous amide modes in acetanilide and its derivatives are presented. Preliminary results of structural data obtained by neutron diffraction at low temperature are also described.

Besides the well-known anomalous amide-1 mode ($1650\,\text{cm}^{-1}$), it is shown that the NH out-of-plane bend ($770\,\text{cm}^{-1}$) and the "H-bond strain" (at about $105\,\text{cm}^{-1}$) exhibit an anomalous increase of intensity proportional to the law $\exp(-T^2/\Theta^2)$, suggesting that the amide proton bears a significant electronic distribution as formerly observed for $H^-$ localized modes. Structural data, moreover, show that the thermal ellips of the amide proton has an increasing anisotropy at 15 K.

Considering these new results, the theoretical model of a self-trapped "polaronic" state seems to be the most consistent with the whole set of observed anomalies in this family of crystals.

## 1. Introduction

The orthorhombic acetanilide crystal ($C_6H_5NHCOCH_3$ or ACN) and its deuterated derivatives contain one-dimensional networks of hydrogen-bonded ...H–N–C–O··· amide groups, like polypeptides and alpha-helix proteins [1]. The family of ACN compounds display some anomalous infrared and Raman modes [2], the origin of which is still a subject of considerable controversy. Tentative explanations of these extra-intensities observed at low temperature have been given in terms of Davydov-like soliton theory [3], vibron soliton [4], "polaronic" localized mode [5,6], nonlinearly coupled oscillators [7,8]. Alternative interpretations involve Fermi resonance [9] and more recently a model based on two slightly nondegenerate configurations of the amide proton [10].

In this paper we present recent data focused on the unconventional amide modes and propose a new comprehensive view of the anomalies in ACN family in terms of a hydrogen localized mode similar to the U centers formerly observed in ionic crystals [11].

## 2. New infrared results

### 2.1. The amide A, B, region [2,12]

Absorption measurements of selectively deuterated ACN and of the methyl fluorinated ACN (abbreviation 3F) [13] allow to estimate the relative strength $W$ of the hydrogen-bond in each derivative.

$$\nu_{3F} = 3303\,\text{cm}^{-1}, \quad \nu_{ACN} = 3281\,\text{cm}^{-1},$$

$$\nu_{D5} = 3275\,\text{cm}^{-1}, \quad \nu_{D3} = 3273\,\text{cm}^{-1},$$

$$\nu_{D8} = 3272\,\text{cm}^{-1}, \quad \nu_{ND} = 2414\,\text{cm}^{-1},$$

it follows that [14]

$$W_{3F} < W_{ACN} < W_{D5} < W_{D3} < W_{D8} < W_{ND} < W_{D9}.$$

(ND means $C_6H_5NDCOCH_3$; D3 means $C_6H_5NHCOCD_3$; D5 means $C_6D_5NHCOCH_3$; and so on.)

The energy shift of $\nu(NH)$ in ACN as a function of the temperature exhibits a characteristic

sigmoid shape, representative of the existence of competing interactions, while $\nu(ND)$ obeys a normal monotonous temperature dependence.

## 2.2. The amide-I region

Our new data are summarized in fig. 1. The amide-I mode exhibits a classical thermal behaviour in D3 and D8: the frequency decreases with decreasing temperature and no extra-peak is observed. On the contrary the anomalous band is present in D5 and in 3F, at 1647.5 cm$^{-1}$ and 1693 cm$^{-1}$, at 20 K, respectively. The energy shift between the fully delocalized exciton (stan-

dard amide-I frequency) and the supposed self-trapped state or "soliton" (anomalous mode frequency) is $\Delta E = 16.5$ cm$^{-1}$ in pure ACN, $\Delta E = 15$ cm$^{-1}$ in D5, and $\Delta E = 11$ cm$^{-1}$ in 3F.

## 2.3. The amide-V modes

The anomalous thermal behaviour of the intensity of this mode (or $\gamma(NH)$) [15–18] consists of an increase of the integrated intensity of about 20% or more, when the temperature decreases from 300 K to 20 K. Fig. 2 shows the anomalous behavior in ACN, at 772 cm$^{-1}$ at about 20 K.
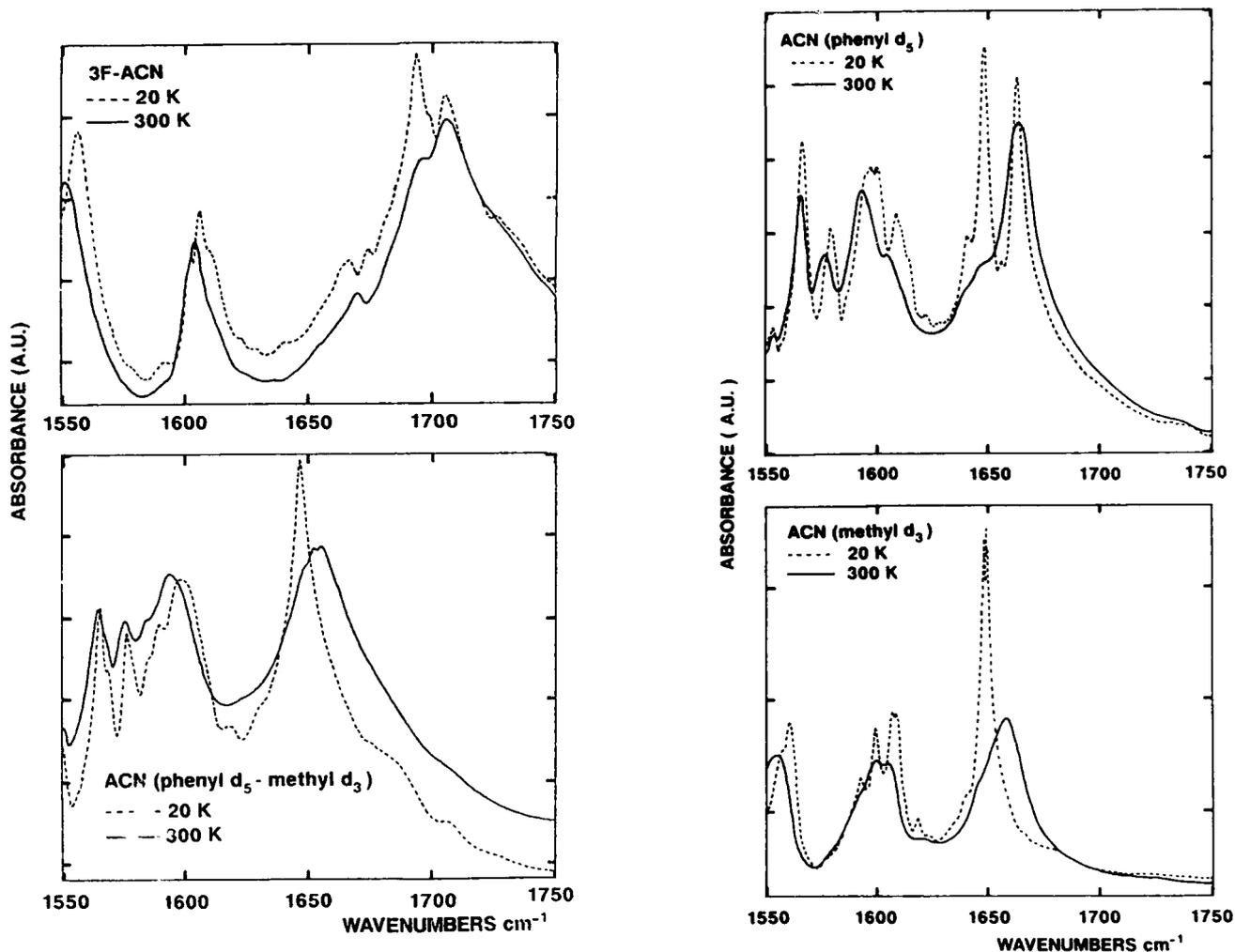


Fig. 1. Amide-I modes in four derivatives of ACN: the D5 and 3F compounds exhibit an anomalous mode like pure ACN, with a very small energy shift in the range 300–20 K. On the contrary, the D3 and D8 compounds display a classical behaviour without anomaly.
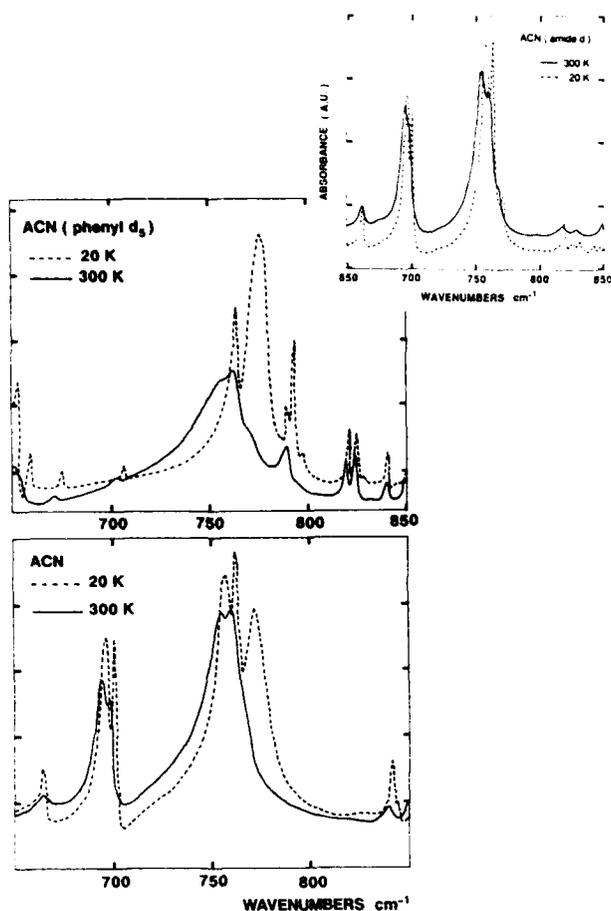
Fig. 2. Amide-V modes (or NH out-of-plane bend) in ACN and D5, showing an increase of the intensity at 20 K. The structure at about 760 cm$^{-1}$ in the ND compound is assigned to the C–H out-of-plane bends in the aromatic ring.

The other modes around 760 cm$^{-1}$ are assigned to the C–H out-of plane bends in the phenyl ring. Their intensity is quasi-independent of the temperature. The anomaly is also observed in D3, D5, D8, and 3F compounds.

The temperature dependence of the anomalous intensity [17] is compatible with the predictions of the "polaronic" model [5,6], or of the "vibron soliton" model [4].

## 2.4. The anomalous modes at about 1500 cm$^{-1}$

They were first identified [17,14] in D8 at about 1490 cm$^{-1}$. Fig. 3 indicates that the same anomaly exists in ACN as a strong increase (more than 30%) of the integrated intensity of
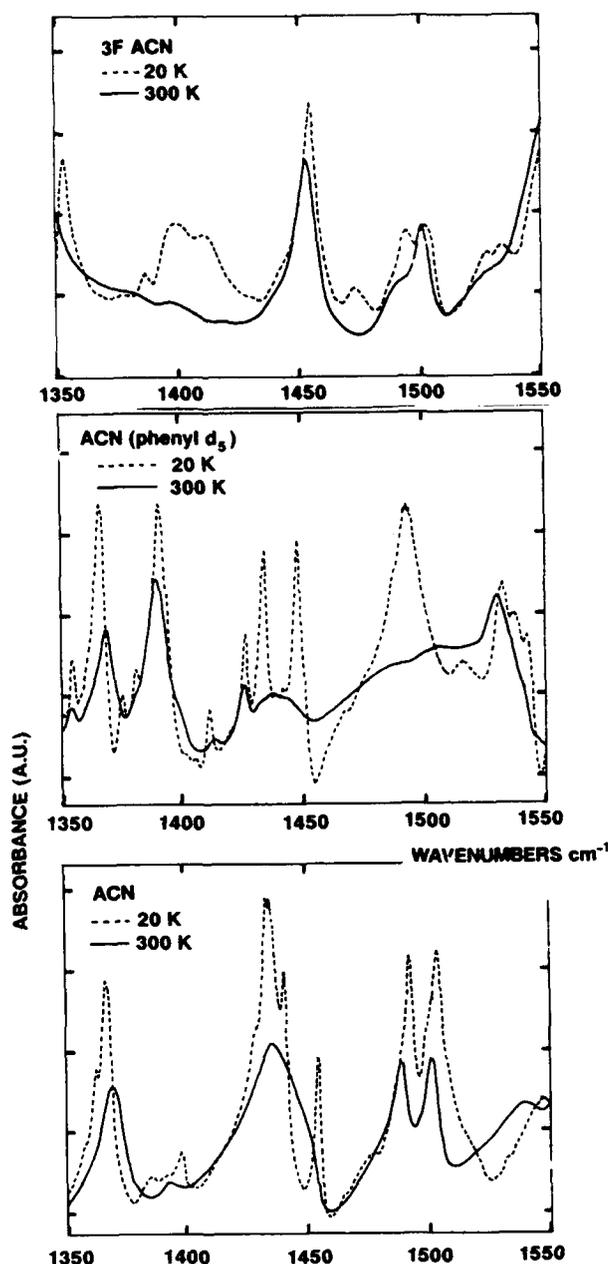


Fig. 3. Anomalous increases of intensities at 20 K in ACN, in D5 and in 3F.

the doublet (1492 and 1504 cm$^{-1}$). Same strong anomalies are observed in D3, D5, and 3F, at 1498, 1492 cm$^{-1}$, and 1400–1411 cm$^{-1}$ respectively. No anomaly is observed in this frequency range on the ND compound.

These anomalous modes are tentatively assigned to the overtone of the anomalous $\gamma$(NH)

mode. They are also accidentally degenerated with the C–C stretch of the phenyl ring, except in 3F. The C–C stretching modes are observed in the ND compound, with a classical temperature behaviour.

An increase of intensity is also observed at about 1430 cm$^{-1}$ in ACN and D5 only, for a mode assigned to a methyl motion.

## 2.5. The amide-II and amide-III modes

The $\delta$(NH) or in-plane bend of the amide proton, at about 1560 cm$^{-1}$, and the N–C stretching modes (presumably at 1260 and 1320 cm$^{-1}$) do not show any anomalous behavior.

## 2.6. Far-infrared modes

Far infrared spectra are dominated by the features at 100–110 cm$^{-1}$ in ACN. This mode displays a strong increase of intensity: about 50% in the whole temperature range (fig. 4). The temperature dependence of its integrated intensity is compatible with a law $\propto \exp(-T^2/\Theta^2)$, its width is consistent with a $T^2$ power law description. All these three specific properties are strikingly resembling that of the fundamental vibrational mode (or zero-phonon mode) of the H$^-$ ion localized oscillator, extensively investigated in alkali halides [11]. Furthermore, the energy of this mode is related to the elastic constants of ACN in the same way as the H$^-$ localized mode in alkali halides and may be scaled on the same plot (cf. fig. 7-2, in ref. [11]).

Its frequency shift with successive deuterations is not related with the molecular mass, but with the H-bond strength (for example the maximum is found at 107 cm$^{-1}$ in ACN, 107 cm$^{-1}$ in D5, 109 cm$^{-1}$ in D3 at about 20 K). Thus it is not a simple classical external mode, but it involves the H-bond. It is worth to remark that in D3 and D8 the mode appears to be split and is accidentally degenerated with the deuterated methyl (CD3) torsion [15].

The motion associated to this major intensity
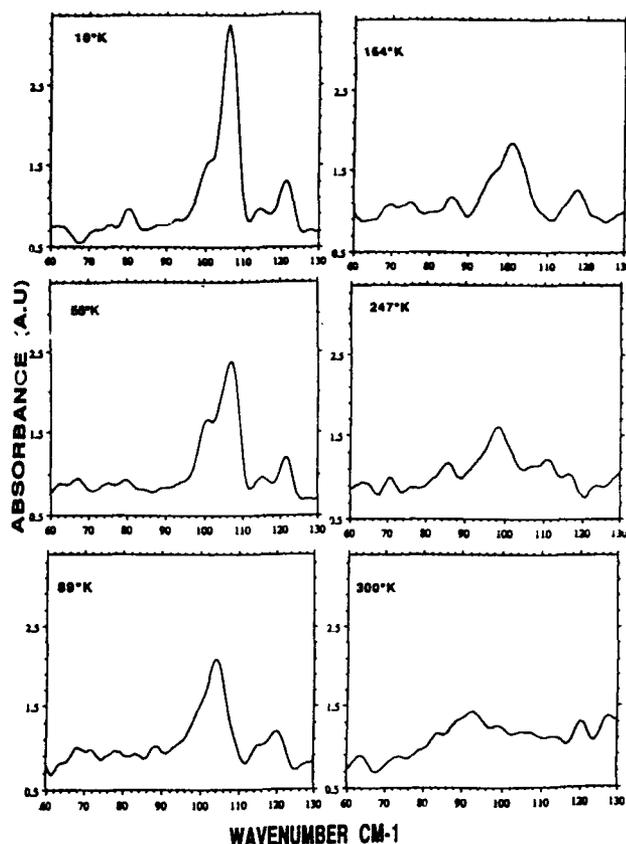


Fig. 4. Evolution of the intensity of the "H-bond strain" in ACN with respect to the temperature.

at 100–110 cm$^{-1}$ may be tentatively assigned to the "H-bond strain" by analogy with the same features in other amide compounds [19]. It consists on a distorsion normal to the (b) H-bond direction, or a hindered rotation of the whole molecule around the $a$ axis involving a significant change of the electric dipole moment, so giving rise to a strong infrared intensity. Much attention has been given to the H-bond strain mode because it is suspected to be the driving mechanism for the opening of the double helix in transcription and replication processes in DNA [19].

## 3. Low temperature neutron diffraction

The low temperature crystallographic structure
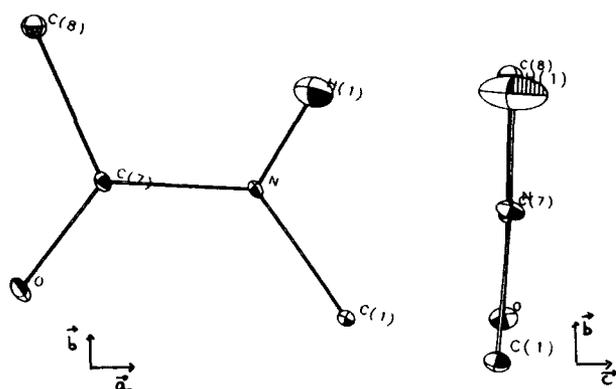
Fig. 5. ORTEP drawings of the amide group in the ACN molecule at 15 K, in the $(a, b)$ molecular plane, and perpendicular. Note the large anisotropy of the $H(1)$ ellipsoïd in the $c$ direction.

of ACN has been measured by neutron diffraction with the aim of determining the amide proton localisation. Preliminary results indicate that the thermal ellipsoïd of the proton is elongated in the $c$ direction, that is perpendicular to the molecular plane. The axis ratios of the ellipsoïd are:

$$\frac{c}{a} = 1.05 \quad \text{and} \quad \frac{c}{b} = 1.69 \quad \text{at room temperature,}$$

$$\frac{c}{a} = 1.54 \text{ and } \frac{c}{b} = 2.05 \text{ at 15 K.}$$

reflecting an increase of the localization of the proton along the $c$ axis as the temperature decreases (fig. 5). However, the difference Fourier maps show only one maximum density region, and not two as could be expected in the model of slightly non-degenerate configurations of the proton.

## 4. Discussion

The hypothesis of a negative charge distribution borne by the amide hydrogen atom was formerly suggested to explain some special NMR chemical shifts in H-bonded systems [20]. The existence of some positive charge density localized on the nitrogen atom was also conjectured

by Pauling [21]. Moreover, the $H^-$ ion existence as an impurity of a color center in alkali halides is well established.

The H-bond is the result of a delicate balance between electrostatic attractive interactions, dipolar forces, and repulsion from electronic orbitals of neighbouring ions. In a classical system like ND, the attractive interactions dominate and the bond strength monotonically increases as the temperature decreases.

In ACN on the contrary, competition occurs and explains the S-shaped curve of the energy shift of the $\nu(NH)$ mode.

Assuming that some negative charge density is bound to the proton, it is possible to explain that repulsive interactions increases as the O...H distance decreases with decreasing temperature, thus explaining the strong localization of the hydrogen ion in the $c$ direction, perpendicular to the bond. It is now easy to understand why both the H-bond strain mode and the $\gamma(NH)$ mode increases in intensity and energy with decreasing temperature.

The anomalous peak at $1650 \text{ cm}^{-1}$ may be a self-trapped state resulting from a nonlinear coupling of the $C = O$ stretch with transverse motions like (in-phase and out-of-phase) librations around the $a$ axis, instead of the longitudinal modes considered in the past. Nevertheless, the experimental values of $\Delta E$ are not in contradiction with the well known prediction [2]:

$$\Delta E = \chi^2/2W - 2J .$$

Following our hypothesis, the disappearance of the anomalous mode in the methyl deuterated compounds (D3 and D8) is no more mysterious. Remembering the accidental degeneracy of the deuterated methyl libration with the "H-bond strain" mode, it is clear that the former counteracts with the latter, so probably restraining its amplitude and prevents its coupling with the $C = O$ stretch. Fermi resonance mechanisms involving either the $CH_3$ "torsional mode", or the "H-bond strain" may be ruled out because the anomalous amide-1 mode still exists in 3F at

about $1693 \, cm^{-1}$, in contradiction with such explanation.

As a conclusion, we have shown that a very simple hypothesis of a $H^-$ center in the amide group suffices to explain the whole set of anomalous features in acetanilide, which, submitted to competing opposite interactions, is a subtle frustrated system. On the other hand we hope that the present data give useful informations on the general knowledge of the hydrogen-bond, involved in most biological processes.

# References

[1] C. Brown, Acta Cryst. 21 (1966) 442.

[2] G. Careri, U. Buontempo, F. Galluzi, A.C. Scott, E. Gratton and E. Shyamsunder, Phys. Rev. B 30 (1984) 4689.

[3] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, Phys. Rev. B 30 (1984) 4703.

[4] S. Takeno, Prog. Theor. Phys. 75 (1986) 1.

[5] D.M. Alexander and J.A. Krumhansl, Phys. Rev. B 33 (1986) 7172.

[6] A.C. Scott, I.J. Bigio and C.T. Johnston, Phys. B 39 (1989) 12883.

[7] A. Campa, A. Giansanti and A. Tenembaum, Phys. Rev. B 36 (1987) 4394.

[8] J. Léon, Phys. Lett. A 152 (1991) 178.

[9] C.T. Johnston and B.J. Swanson, Chem. Phys. Lett. 114 (1985) 547.

[10] W. Fann, L. Rothberg, M. Roberson, S. Benson, J. Madey, S. Etemad and R. Austin, Phys. Rev. Lett. 64 (1990) 607.

[11] J. Klein, in: Physics of Color Centers, ed. W.B. Fowler (Academic Press, New York, 1968) ch. 7.

[12] G. Blanchet and C. Fincher, Phys. Rev. Lett. 54 (1985) 1310.

[13] J. Lapasset and J. Moret, PS 05.04.18, Congrés International de Cristallographie (Bordeaux, France, 1990).

[14] J.L. Sauvajol, G. de Nunzio, R. Almairac, M. Barthes and J. Moret, Solid State Commun. 77 (1991) 199.

[15] M. Barthes, R. Almairac, J.L. Sauvajol, J. Moret, R. Currat and J. Dianoux, Phys. Rev. B 43 (1991) 5223.

[16] M. Barthes, J. Eckert, S. Johnson, B. Swanson, C. Unkefer and J. Moret, J. Phys. (France) 2 (1992) 1929.

[17] M. Barthes, in: Davydov's Soliton Revisited, Self-Trapping of Vibrational Energy in Protein, eds. P.L. Christiansen and A.C. Scott (Plenum, New York, 1990).

[18] C.T. Johnston and B.I. Swanson, unpublished.

[19] N. Østergard, PhD Thesis, The Technical University of Denmark (1991).

[20] W. Hamilton and J. Ibers – Benjamin, eds., Hydrogen Bonding in Solids (New York, 1968) see p. 17.

[21] N. Østergard, PhD Thesis, The Technical University of Denmark (1991) pp. 32, 33.

PHYSICA D

# Quantum chemical calculation and analysis of third-order nonlinear polarizabilities for linear conjugated π-electron molecules

Brian M. Pierce

*Hughes Aircraft Company, Bldg. R02, MS V518, P.O. Box 92426, Los Angeles, CA 90009-2426, USA*

The nonresonant π-electronic component of $\gamma(-3\omega; \omega, \omega, \omega)$, $\gamma_\pi$, is calculated and analyzed for two important types of linear conjugated π-electron molecules: linear cyanine cations and linear polyenes. The specific linear cyanine cations are all-*trans*, linear symmetric cyanine and streptocyanine cations with no double-single bond-length alternation, and cyanine cations with asymmetric geometries resulting from the artificial imposition of double-single bond-length alternation. The linear polyenes of interest are all-*trans* linear polyenes with double-single bond-length alternation, and those with geometries resulting from the artificial imposition of no double-single bond-length alternation. Bond-length alternation is found to dramatically affect $\gamma_\pi$ for the linear cyanines: the $\gamma_\pi$'s for the symmetric cyanines are calculated to be *negative*; the $\gamma_\pi$'s for the assymmetric cyanines *change* from negative to positive with increasing chain length. The $\gamma_\pi$'s for the linear polyenes are always *positive* regardless of the extent of bond-length alternation; the $|\gamma_\pi|$'s for the linear polyenes increase with decreasing bond-length alternation. The $|\gamma_\pi|$'s for the symmetric linear cyanines increase more rapidly with the number of π-electrons than the $|\gamma_\pi|$'s for the linear polyenes: $|\gamma_\pi(\text{symmetric cyanines})| \propto N_{\pi\text{-}e}^8$ and $|\gamma_\pi(\text{linear polyenes})| \propto N_{\pi\text{-}e}^4$, where $N_{\pi\text{-}e} = 4, 6, 8, 10, 12$. The linear cyanine cations comprise a very promising class of nonlinear optical, π-electron molecules that merit further experimental study.

## 1. Introduction

Nonlinear optical phenomena are increasingly being utilized in a variety of photonic applications. The second-order nonlinear optical process of second-harmonic generation is used to extend the frequency ranges of lasers [1–3]. Electro-optic modulators and switches (e.g., Pockel's cell) operate via the second-order nonlinear optical effect of an electric-field-induced change in refractive index [3,4]. The third-order nonlinear optical phenomenon of a light-induced change in refractive index is fundamental to soliton light-pulse propagation in optical media [5], optical switching and computing [6,7], as well as phase conjugate adaptive optics [8]. The efficiencies of these nonlinear optical processes are dependent upon the material employed to couple the given combination of electrical and/or optical signals.

Organic molecules and polymers with large nonlinear electronic polarizabilities are of great interest world-wide as nonlinear optical (NLO) materials [9]. The organic molecules with the largest nonlinear electronic polarizabilities tend to fall in one of three classes of molecules with conjugated π-electron bonding networks: (1) polyenes and related compounds with pronounced π-electron bond-order alternation, e.g., the asymmetric merocyanines, $H_2N–CH=(CH–CH)_n=O$, (2) symmetric polymethines with attenuated π-electron bond-order alternation, e.g., the symmetric cyanine cations, $H_2N\dot{=}CH\dot{=}(CH\dot{=}CH)_n^+\dot{=}NH_2$, and (3) porphyrins and phthalocyanines. The conjugated π-electron bonding networks in these molecules not only endow them with large nonlinear π-electronic polarizabilities, but with large linear π-electron polarizabilities, and high one-photon absorptivities in the visible and near-IR spectral regions [9].

The attenuated π-bond-order alternation for

the symmetric polymethines results from a π-electron ground state that can be described as a resonance hybrid of two equivalent structures, e.g., $^+H_2N=CH-(CH=CH)_nNH_2 \leftrightarrow H_2N-CH=(CH-CH)_n=NH_2^+$ for the symmetric cyanine cations. The pronounced bond-order alternation in the polyene-like molecules results from a π-electron ground state that is best described by a single structure, e.g., the $^+H_2N=CH-(CH=CH)_n-O^-$ structure for the merocyanines in a vacuum is not equivalent to the more stable $H_2N-CH=(CH-CH)_n=O$ structure.

In free-electron theory, the potential energy of a π-electron in a symmetric polymethine can be approximated by a constant value because of the π-electron density delocalization associated with the resonance hybrid; the potential energy of a π-electron in a polyene-like molecule can be approximated by a periodic function with the troughs and crests of the function coinciding with the double and single bonds [10]. An important consequence of these different potential energy functions is that the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) goes to zero with increasing chain length for symmetric polymethines, but converges to a finite value with increasing length for polyenes. In other words, symmetric cyanines become metallic in the limit of infinite length, while the polyenes resemble semiconductors in this limit [11,12].

This free-electron/Hueckel theoretical picture is in agreement with experimental and theoretical studies of the *linear* optical properties of symmetric polymethines and polyene-like molecules. For example, there is the experimental observation that $\lambda_{max}$ for the first one-photon allowed $^1\pi\pi^* \leftarrow S_0$ transition increases linearly with chain length for symmetric all-*trans*, polymethine cations with two terminal dimethyl amino groups $[(CH_3)_2N\dddot-CH\dddot-(CH\dddot-CH)_n^+\dddot-N(CH_3)_2$ or streptocyanines] [13], but reaches a limiting value in the case of all-*trans*-α, ω-dialkyl-polyenes [14], and the dimethylamino-

merocyanines [13], $(CH_3)_2N-CH=(CH-CH)_n=O$. Also, the linear, static electric-dipole polarizabilities of symmetric cyanine cations have been measured [15] and calculated [12,16,17] to be larger than those for all-*trans*-linear polyenes of comparable methylation, length, and number of π-electrons.

One would also expect the *nonlinear* optical properties of symmetric polymethines to be different from those of polyene-like molecules. Third-harmonic generation (THG) measurements of the nonresonant electronic component of the third-order, electric-dipole polarizability $[\gamma_e(-3\omega; \omega, \omega, \omega)$, with $\omega = 1.91$ μm] for a variety of symmetric cyanine dyes in the condensed phase yielded $\gamma_e(-3\omega; \omega, \omega, \omega)$'s with both *negative and positive signs*, while the measurements for a selection of polyenes yielded only positive $\gamma_e(-3\omega; \omega, \omega, \omega)$'s [15]; the magnitudes of the $\gamma_e(-3\omega; \omega, \omega, \omega)$'s measured for streptocyanines met or exceeded those for $\alpha$, $\alpha$, $\omega$, $\omega$-tetramethyl-polyenes of comparable size and number of π-electrons [15]. The THG measurements at $\omega = 1.85-2.15$ μm for the polycrystalline powder of a symmetric cyanine dye with terminal quinoline rings determined $|\gamma_e(-3\omega; \omega, \omega, \omega)|$ to be greater than that of the well-studied poly-diacetylene system, poly-(2, 4-hexadiyn-1, 6-diol-bis-(p-toluene sulfonate)), or PDA-PTS [18]. A recent study of the third-order nonlinear optical properties for a set of zwitterionic polymethine dyes, the squaryliums, emphasizes the importance of the polymethines as a class of nonlinear optical molecules distinct from the polyenes [19].

The principal subject of this paper is the calculation and analysis of the nonresonant, π-electronic components of third-order, electric dipole polarizabilities for a set of linear streptocyanines and unmethylated streptocyanines (hereafter known simply as cyanines). In contrast to earlier theoretical studies of these polarizabilities for linear cyanines [16,17], we (1) use a sum-over-states expression derived from time-dependent perturbation theory [20] to calculate the π-electronic component of the *dynamic* third-order

polarizability for third-harmonic generation, $\gamma_\pi(-3\omega; \omega, \omega, \omega)$, as well as the *static* third-order polarizability, $\gamma_\pi(0; 0, 0, 0)$; and (2) treat the π-electronic states using the all-valence-electron, semiempirical INDO (intermediate neglect of differential overlap) molecular orbital procedure combined with single- and double-excitation configuration interaction (SDCI) of the singlet π-electron configurations, instead of using original [16] or refined [17] versions of Kuhn's free-electron method [10]. Although our theoretical method is more complicated than those based on free-electron theory, it has been shown to provide a *consistent* treatment of the π-electronic components of the third-order polarizabilities and the linear and nonlinear spectroscopic properties of the $^1\pi\pi^*$ states of linear polyenes and benzene [21]. Critical to our consistent calculation of these properties is the treatment of π-electron correlation at least at the level of SDCI [21].

We calculated $\gamma_\pi$'s for all-*trans*, linear symmetric cyanine and streptocyanine cations with no double-single bond-length alternation, and for cyanine cations with asymmetric geometries resulting from the artificial imposition of double-single bond-length alternation. These $\gamma_\pi$'s are compared with those previously calculated for all-*trans* linear polyenes with double-single bond-length alternation [21], and for a new set calculated with geometries resulting from the artificial imposition of no double-single bond-length alternation. The structures of the five sets of molecules are shown in fig. 1.

Our study of the π-electron, third-order polarizabilities for the molecules shown in fig. 1 consists of two parts: (1) a comparison of the calculated and measured $\gamma(-3\omega; \omega, \omega, \omega)$'s for the linear streptocyanines to extend the validation of our theoretical method beyond that of the linear polyenes and benzene [21], and (2) a comparison of the calculated $\gamma(-0; 0, 0, 0)$'s that emphasizes analyzing the effect of terminal methyl group substitution on the $\gamma_\pi$'s for the symmetric cyanine cations, the effect of bond-
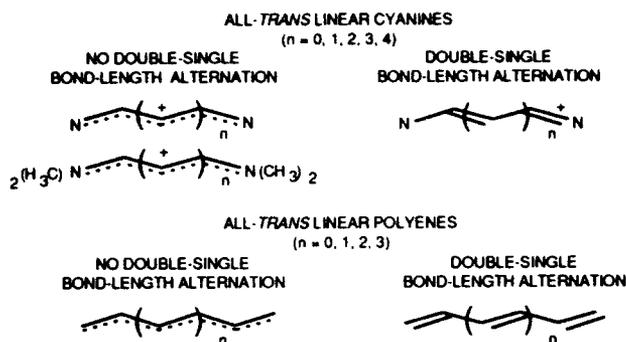


Fig. 1. Structures of the linear cyanines and polyenes with and without bond-length alternation.

length alternation on the signs and magnitudes of the $\gamma_\pi$'s for the linear cyanines and polyenes, and the dependence of $|\gamma_\pi|$ on the number of π-electrons in both sets of molecules. A preliminary account of this study is given in ref. [22].

## 2. Theoretical method

This section consists of three parts. The first part establishes the general relationship between the macroscopic third-order electric susceptibility and the microscopic, molecular third-order polarizability. One of the most important third-order electric susceptibilities from a technological point of view is the light-induced change in refractive index. This nonlinear refractive index is related to the molecular third-order polarizability in the second part of the section. The third and final part describes the quantum chemical calculation of the nonresonant π-electronic component of the third-order molecular polarizability.

### 2.1. Relation between macroscopic susceptibility and molecular polarizability

The mathematical expression for the polarization of a non-conducting, non-magnetic medium by an applied external electric field(s) is given by the Taylor series expansion [1,3]

$$P_A(\omega_{\text{pol}}) = \sum_B \chi^{(1)}_{AB} E_B(\omega_1)$$

$$+ \frac{1}{2!} \sum_B \sum_C \chi^{(2)}_{ABC} E_B(\omega_1) E_C(\omega_2)$$

$$+ \frac{1}{3!} \sum_B \sum_C \sum_D \chi^{(3)}_{ABCD} E_B(\omega_1) E_C(\omega_2)$$

$$\times E_D(\omega_3) + \cdots ,\qquad (1)$$

where $P_A$ is the component of the polarization vector along the $A$ Cartesian coordinate axis in the laboratory reference frame; $\omega_{\text{pol}}$ is the angular frequency of the polarization field, which contains contributions from the linear and nonlinear polarization of the material, e.g., $\omega_1$, $\omega_1 + \omega_2$, and $\omega_1 + \omega_2 + \omega_3$; $\chi^{(1)}_{AB}$, $\chi^{(2)}_{ABC}$, and $\chi^{(3)}_{ABCD}$ are the real components of the electric susceptibility tensors of rank 2, 3, and 4, respectively; and $E_\alpha(\omega_k)$ is the $\alpha = A$, $B$, or $C$ component of the applied electric field vector oscillating at angular frequency $\omega_k$. All terms in eq. (1) are in esu units. The 1/2! and 1/3! factors in eq. (1) are often incorporated into the definition of $\chi^{(2)}_{ABC}$, and $\chi^{(3)}_{ABCD}$. Note that $\chi^{(2)}_{ABC}$ is zero for centrosymmetric media.

Most organic nonlinear optical materials (e.g., organic molecular crystals) are composed of molecular units whose polarizabilities sum to define the macroscopic electric susceptibilities. The electric susceptibility tensors for these media can then be expressed as a function of the number density of molecular units [$N$], a term containing the local electric field tensor and the direction cosines that relate the molecule-fixed coordinate axes to the laboratory coordinate axes [$L_n(\theta)$], and the molecular polarizability tensor of rank $n + 1$ [$\alpha^{(n)}$]:

$$\chi^{(n)} = f(N, \alpha^{(n)}, L_n(\theta)) \quad (n = 1, 2, 3).\qquad (2)$$

The molecular polarizability tensors, $\alpha^{(n)}$ ($n = 1, 2, 3$), are more commonly represented as $\alpha$ (molecular polarizability), $\beta$ (second-order molecular polarizability), and $\gamma$ (third-order molecular polarizability), respectively. For weakly-coupled molecules randomly oriented in a given medium, $\chi^{(n)}$ can be expressed simply as [23,24]

$$\chi^{(n)} = N\langle\alpha^{(n)}\rangle L_n \quad (n = 1, 2, 3),\qquad (3)$$

where $L_n$ is a scalar local-field factor defined by the refractive indices of the medium at the frequencies of the applied electric fields and the polarization field, and $\langle\alpha^{(n)}\rangle$ is the orientationally-averaged molecular polarizability tensor. The product, $\langle\alpha^{(1)}\rangle L_1$, defines the averaged first-order dipole moment induced at the given molecular unit by an applied electric field. In a similar fashion, the products, $\langle\alpha^{(2)}\rangle L_2$ and $\langle\alpha^{(3)}\rangle L_3$, define the averaged second- and third-order dipole moments induced at this unit.

## 2.2. Light-induced change in refractive index

The third-order nonlinear optical phenomenon of a light-induced change in refractive index is fundamental to soliton light-pulse propagation in optical fibers [5], optical switching and computing [6,7], as well as phase conjugate adaptive optics [8]. The dependence of the real refractive index at wavelength $\lambda$ on the incident light intensity can be expressed as

$$n(\lambda, I) = n_0(\lambda) + \Delta n(\lambda, I),\qquad (4)$$

where $n_0(\lambda)$ is the linear refractive index at $\lambda$, and $\Delta n(\lambda, I)$ is the change in refractive index as a function of $\lambda$ and the incident light intensity $I$. The leading terms in the series expansion for $\Delta n(\lambda, I)$ are $n_2(\lambda) I$ and $n_{\text{th}}(\lambda) I$, so that

$$n(\lambda, I) \cong n_0(\lambda) + [n_2(\lambda) + n_{\text{th}}(\lambda)]I.\qquad (5)$$

The nonlinear refractive index coefficient, $n_2(\lambda)$, defines the change in refractive index arising from the electric-field induced polarization of the material, while the $n_{\text{th}}(\lambda)$ coefficient describes the change in the real part of the refractive index resulting from the light-induced heating of the material, i.e., the thermo-optic

effect [6]. The thermo-optic effect is to be avoided in all device applications.

The $n_2(\lambda)$ coefficient is related to the third-order nonlinear electric susceptibility tensor element, $\chi^{(3)}_{ABCD}(\lambda)$, by the following equation [25]:

$$n_2(\lambda) = \tfrac{4}{3}\pi^2\chi^{(3)}_{ABCD}(\lambda)/(cn_0(\lambda)^2), \qquad (6)$$

where $c$ is the speed of light. The units of all parameters in eq. (6) are in esu. The expression for $n_2(\lambda)$ in eq. (6) is appropriate only for the Taylor series expansion given in eq. (1). This third-order nonlinear optical response is also known as the optical Kerr effect. According to eq. (3), eq. (6) can be rewritten to directly relate $n_2(\lambda)$ to the third-order polarizability, $\lambda$, for weakly-coupled molecules randomly oriented in a given medium:

$$n_2(\lambda) = \tfrac{4}{3}\pi^2 N\langle\gamma(\lambda)\rangle L_3(\lambda)/(cn_0(\lambda)^2). \qquad (7)$$

## 2.3. Calculation of third-order molecular polarizabilities

Only two types of third-order molecular polarizabilities are calculated here, the π-electronic component of the *dynamic* third-order polarizability for third-harmonic generation, $\gamma_\pi(-3\omega;\omega,\omega,\omega)$, and the *static* third-order polarizability, $\gamma_\pi(0;0,0,0)$. The calculation of $\gamma_\pi(-3\omega;\omega,\omega,\omega)$ and $\gamma_\pi(0;0,0,0)$ for a given molecular unit proceeds in three steps: (1) determination of a molecular geometry, (2) calculation of the transition energies, permanent state dipole moments, and transition dipole moments associated with the ground and excited π-electron states for this geometry, and (3) evaluation of the sum-over-states expression for the $\gamma_\pi$'s using these transition energies and dipole moments.

The π-electron bonding networks in all the molecules in fig. 1 were kept planar. The bond-lengths and bond-angles used for these molecules are given in ref. [22]. It is noted that the π-electron polarizabilities and spectroscopic properties of the molecules are not very sensitive to the lengths of the N–H and C–H bonds.

The INDO–SDCI method employed in the calculation of the π-electron transition energies, permanent state dipole moments, and transition dipole moments for the various molecules is the same as that described in ref. [21], with one exception. This exception is that the transition energies were calculated using the formula [26], $\Delta E_{i\leftarrow 0} = E_i(5/N)E_0$, where $E_i$ is the energy of excited state $i$, $E_0$ is the ground electronic state energy, and $N$ is the number of conjugated π-molecular orbitals. This formula scales the ground state energy to compensate for the imbalance in the degree to which the ground state and the $^1\pi\pi^*$ states are stabilized by single- and double-excitation interaction [26].

Our evaluation of the sum-over-states expressions for $\gamma_\pi(-3\omega;\omega,\omega,\omega)$ and $\gamma_\pi(0;0,0,0)$, and the isotropic averages of these polarizabilities, $\langle\gamma_\pi(-3\omega;\omega,\omega,\omega)\rangle$ and $\langle\gamma_\pi(0;0,0,0)\rangle$, are as given in ref. [21]. Unless indicated otherwise, $\gamma_\pi(-3\omega;\omega,\omega,\omega)$ and $\gamma_\pi(0;0,0,0)\rangle$ will henceforth be taken to be synonymous with $\langle\gamma_\pi(-3\omega;\omega,\omega,\omega)\rangle$ and $\langle\gamma_\pi(0;0,0,0)$. The $x$-axis of the molecule-fixed coordinate systems for the linear molecules in fig. 1 is taken to be coincident with a line connecting the two terminal atoms of the π-electron bonding network. As a result, the $\gamma_{xxxx}$ component is expected to dominant the $\gamma_\pi$ tensor.

## 3. Results and discussion

### 3.1. Comparison of calculated and measured $\gamma_\pi(-3\omega;\omega,\omega,\omega)$'s for linear streptocyanines

Our theoretical method has been successful in calculating nonresonant $\gamma_\pi(-2\omega;\omega,\omega,0)$'s for ethylene, linear polyenes, and benzene that are in very good agreement with the $\gamma$'s measured for these molecules in the gas phase [21]. A similar comparison of calculated and measured

γ's is given in table 1 for three linear strep-tocyanine cations of different lengths. The calculated values of $\gamma_\pi$ and $\gamma$ in table 1 are rounded-up to the most significant digit. This table contains the $\gamma(-3\omega; \omega, \omega, \omega)$'s measured at $\omega = 1.908 \, \mu$m for the three molecules in the liquid phase [15], and the $\gamma(-3\omega; \omega, \omega, \omega)$'s calculated for the streptocyanines with *both* symmetric and asymmetric geometries. The $\gamma(-3\omega; \omega, \omega, \omega)$'s for the two geometries were calculated since it is unclear which geometry was stabilized by the Cl⁻ counterion and the solvent used in the experimental samples. In this regard, Fabian and Mehlhorn [27] comment that the asymmetric geometries seen experimentally [28] in crystals of streptocyanines may be due to a non-symmetric arrangement of the counterion to the chain in the crystal. Note that the calculated $\gamma(-3\omega; \omega, \omega, \omega)$'s in table 1 are separated into $\sigma$- and $\pi$-electron components. The $\gamma_\sigma$ term was estimated using the bond-additivity approximation described in ref. [21], while the $\gamma_\pi$ term was that calculated for the unmethylated cyanines.

As shown in table 1, the magnitudes of the measured γ's and those calculated for either the symmetric or asymmetric streptocyanines are in rough agreement. The positive signs of the γ's measured for the two shorter streptocyanines can be explained in terms of the stabilization of the symmetric geometries by the Cl⁻ counterion and solvent, while the negative sign of the γ mea-

sured for the longest streptocyanine can be explained in terms of the stabilization of the asymmetric geometry. This interpretation of the measured data needs to be investigated further. Recent measurements of non-resonant third-harmonic generation in liquid solutions containing cyanine molecules confirm the above theoretical prediction that the sign of γ for these molecules can be negative, and that the sign and magnitude of γ should be very sensitive to the dielectric properties of the solvent [29].

## 3.2. Comparison of calculated $\gamma_\pi(0; 0, 0, 0)$'s

The $|\gamma_\pi(0; 0, 0, 0)|$'s calculated for the five sets of molecules in fig. 1 are plotted on a logarithmic scale in fig. 2 as functions of the number of π-electrons $(N_{\pi-e})$ in each molecule. The symmetric, linear streptocyanines with no bond-length alternation are designated LC1 in fig. 2; LC2 is the label for the symmetric, linear cyanines with no bond-length alternation; LC3 represents the asymmetric linear cyanines with bond-length alternation; LP1 designates the linear polyenes with bond-length alternation; and LP2 represents the linear polyenes with no bond-length alternation. A dashed line in fig. 2 indicates a negative $\gamma_\pi(0; 0, 0, 0)$, and a solid line a positive $\gamma_\pi(0; 0, 0, 0)$.

In regard to the *sign* of $\gamma_\pi(0; 0, 0, 0)$ for the different molecules, the symmetric linear cyanines with and without terminal methyl

Table 1
Calculated and measured γ's for selected linear strep-tocyanines.

| Molecule | Theory ($10^{-36}$ esu) | | | Expt ($10^{-36}$ esu) |
|---|---|---|---|---|
| | $\gamma_\sigma$ | $\gamma_\pi$ | $\gamma$ | |
| 2(Me)N≈≈≈N(Me)₂ | 2.8 | −6 | −3 | |
| | | | | 0.5–3 ± 0.4 |
| 2(Me)N≈≈≈N(Me)₂ | 2.8 | −1 | 2 | |
| 2(Me)N≈≈≈N(Me)₂ | 3.3 | −80 | −80 | |
| | | | | 52 ± 1 |
| 2(Me)N≈≈≈N(Me)₂ | 3.3 | 40 | 40 | |
| 2(Me)N≈≈≈N(Me)₂ | 3.7 | −800 | −800 | |
| | | | | −510 ± 50 |
| 2(Me)N≈≈≈N(Me)₂ | 3.7 | 1000 | 1000 | |



Fig. 2. Calculated $|\gamma_\pi(0; 0, 0, 0)|$'s for the five sets of molecules shown in fig. 1.

π-electron molecules that merit further experimental study.

## Acknowledgements

## References

[1] N. Bloembergen, Nonlinear Optics (Benjamin, New York, 1965).

[2] F. Zernike and J.E. Midwinter, Applied Nonlinear Optics (Wiley, New York, 1973).

[3] Y.R. Shen, The Principles of Nonlinear Optics (Academic Press, New York, 1984).

[4] A. Yariv and P. Yeh, Optical Waves in Crystals. Propagation and Control of Laser Radiation (Wiley, New York, 1984).

[5] L. Mollenauer and R.H. Stolen, Opt. Lett. 9 (1984) 13, see also contributions to this Proceedings.

[6] G.I. Stegeman and E.M. Wright, Optical and Quantum Electronics 22 (1990) 95.

[7] H.M. Gibbs, Optical Bistability: Controlling Light with Light (Academic Press, New York, 1985).

[8] R. Fisher, ed., Optical Phase Conjugation (Academic Press, New York, 1983).

[9] (a) S.R. Marder, J.E. Sohn and G.D. Stucky, eds. Materials for Nonlinear Optics: Chemical Perspectives, American Chemical Society Symp. Series, Vol. 455 (American Chemical Society, Washington, DC, 1991);
(b) G. Khanarian, ed., Nonlinear Optical Properties of Organic Materials III, SPIE Proc., Vol. 1137 (SPIE, Bellingham, WA, 1991);
(c) A.J. Heeger, J. Orenstein, D.R. Ulrich, eds., Nonlinear Optical Properties of Polymers, Materials Research Society Symp. Proc., Vol. 109 (Materials Research Society, Pittsburgh, PA, 1988);

(d) D.S. Chemla and J. Zyss, Nonlinear Optical Properties of Organic Molecules and Crystals (Academic Press, New York, 1987).

[10] H. Kuhn, J. Chem. Phys. 17 (1949) 1198.

[11] J.M. André, Etude comparative des methods de Hueckel et de l'electron libre, Ms.Sc Thesis, Universitaire Catholique de Louvain, Leuven, Belgium (1965).

[12] V.P. Bodart, J. Delhalle and J.M. André, in: Conjugated Polymeric Materials: Opportunities in Electronics, Optoelectronics, and Molecular Electronics, eds. J.L. Bredas and R.R. Chance (Kluwer, Netherlands, 1990) pp. 509–516.

[13] S.S. Malhotra and M.C. Whiting, J. Chem. Soc. (1960) 3812.

[14] B.E. Kohler, J. Chem. Phys. 93 (1990) 5838.

[15] S.H. Stevenson, D.S. Donald and G.R. Meredith, in: Nonlinear Optical Properties of Polymers, eds. A.J. Heeger, J. Orenstein, D.R. Ulrich, Materials Research Society Symp. Proc., Vol. 109 (Materials Research Society, Pittsburgh, PA, 1988) pp. 103–198.

[16] K.C. Rustagi and J. Ducuing, Optics. Commun. 10 (1974) 258.

[17] C. Kuhn, Synthetic Metals 41–43 (1991) 3681.

[18] S. Matsumoto, K. Kubodera, T. Kurihara and T. Kaino, Optics Commun. 76 (1990) 147.

[19] C.W. Dirk, L.T. Cheng and M.G. Kuzyk, Int. J. Quantum Chem., Int. J. Quantum Chem. 43 (1992) 27.

[20] B.J. Orr and J.F. Ward, Molec. Phys. 20 (1971) 513.

[21] B.M. Pierce, J. Chem. Phys. 91 (1989) 791.

[22] B.M. Pierce, in: Nonlinear Optical Organic Materials IV ed., K.D. Singer, SPIE Proc., Vol. 1560 (1991) pp. 148–161.

[23] J.A. Armstrong, N. Bloembergen, J. Ducuing and P.S. Pershan, Phys. Rev. 127 (1962) 1918.

[24] B.F. Levine and C.G. Bethea, J. Chem. Phys. 63 (1975) 2666.

[25] J.L. Brédas, C. Adaut, P. Tackx, A. Persoons and B.M. Pierce, Chem. Rev., submitted.

[26] K. Schulten, U. Dinur and B. Honig, J. Chem. Phys. 73 (1980) 3927.

[27] J. Fabian and A. Mehlhorn, J. Molec. Struct. (Theochem) 109 (1984) 17.

[28] D. Daehne, Science 199 (1978) 1163.

[29] J. Perry, G. Bournhill and S. Marder, private communication (1992).

PHYSICA D

# Dynamics of the amide-I excitation in a molecular chain with thermalized acoustic and optical modes

Alexander V. Savin[a] and Alexander V. Zolotaryuk[b,c]

[a]Institute for Physico-Technical Problems, 119034 Moscow, Russian Federation
[b]N.N. Bogolyubov Institute for Theoretical Physics, Ukrainian Academy of Sciences 252143 Kiev, Ukraine
[c]Research Center of Crete, Foundation for Research and Technology-Hellas,
P.O. Box 1527, 711 10 Heraklion, Crete, Greece

An improved version of the Davydov soliton model to describe the physical mechanisms of the energy transfer in proteins has been introduced. The essential modifications have been done for the lattice subsystem: (1) the intrapeptide N—H bond is allowed to move together with its peptide group and therefore both the acoustic and optical modes appear to be coupled and (2) a harmonic on-site potential, which describes the influence of the environment of the chain, is introduced. The chain is considered to be in contact with a heat bath only via the acoustic mode, while the coupling of the amide-I excitation with the optical mode is shown to be crucial for the thermal stability of the Davydov soliton. The presence of a substrate on-site potential, modelling the influence of the environment, is shown to improve the soliton stability with respect to thermal oscillations since it essentially reduces the amplitude of acoustic oscillations.

## 1. Introduction

The study of the mechanisms for energy and charge transport in biological systems represents one of the important aspects of bioenergetics. In 1973, Davydov suggested a soliton mechanism [1] to explain the efficient transfer of the vibrational amide-I energy in proteins. According to his model, solitary waves (or solitons) are carriers of the intramolecular high-frequency vibrational amide-I energy which can be excited inside each peptide group (PG) H—N—C=O involving mainly longitudinal oscillations of the C=O bond (for a comprehensive review see Scott [2]). The formation mechanism of the Davydov soliton is based on the balance between the effects of dispersion caused by the resonance interaction of adjacent PGs and the interaction of the amide-I mode with longitudinal deformation (acoustic mode) of the chain (self-trapping mechanism).

The possibility of the successful application of the concept of the Davydov soliton in biology depends on its stability at physiological temperatures, i.e. at $T \approx 300$ K. Therefore its thermal stability has recently been the subject of intensive studies [3–10]. Thus, for instance, the numerical simulations [4,5] on the basis of the Langevin approach discover the thermal destabilization of the Davydov soliton at room temperatures. As follows from these calculations, the amplitude of thermal vibrations of PGs exceeds more than one order the amplitude of a localized deformation of the chain. In this case, thermal vibrations result in the break of the coupling of the amide-I excitation with the local deformation of the chain and as result, the soliton (self-trapping) state disappears.

However, the interaction of the intramolecular amide-I excitation with other low-frequency intramolecular vibrations should improve the soliton stability. The period of such optical vibrations, compared to acoustic oscillations, is much less than the time for the excitation to be transferred from one PG to its neighbor. The excitation is sensitive to these optical oscillations

in an averaged way, responding only on the motion of their equilibrium positions. Therefore, in contrast to thermal acoustic oscillations, thermal optical vibrations should not exert any visible destabilization influence on the Davydov soliton.

In this paper the dynamics of the self-trapping state of the amide-I excitation in a thermalized chain is studied provided that it interacts with both the acoustic and optical modes. The interaction of the chain with a heat bath is assumed to be described by the Langevin-type equation. The phenomenological damping and the color noise, which model the interaction of the chain with the bath, are involved only into the acoustic subsystem, while the optical subsystem is supposed to be thermalized via its coupling with acoustic oscillations. The interaction of the amino acids with its local environment chosen in the form of a harmonic on-site potential [11] is also included into our studies of temperature effects.

The best appropriate material which may be described on the basis of this model is the crystalline acetanilide (ACN) $(CH_3CONHC_6H_5)_x$ with parameter values very close to those of polypeptide protein chains [12,13]. Experimentally, in the crystalline ACN the amide-I excitation is observed as the maximum of the infrared absorption $1667\,cm^{-1}$. Besides this band in the infrared and Raman spectra for ACN, another band with the red shift about $17\,cm^{-1}$ has been found in the experiments. The suggestion has been done that this band is caused by the interaction of the amide-I with an intrapeptide displacement of the proton in the N—H bond and it may be identified with Davydov-like soliton states stabilized by this interaction. Thus, to describe the soliton transport of the amide-I energy along the chain of hydrogen-bonded PGs both the couplings of this excitation with the deformation of the hydrogen bonds and with the intrapeptide proton displacements should be considered.

## 2. A modified Davydov model

Consider a typical one-dimensional chain of $N$ hydrogen-bonded peptide groups. Suppose that along the axis $x = nl$, where $l$ is the lattice spacing, PGs are placed in the sites with $n = 1, \ldots, N$. Besides the longitudinal displacements of PGs as whole objects, the internal motion of the proton in the N—H bond is also considered. Schematically this lattice model is shown in fig. 1.

The standard quantum mechanical description of this model can be reduced in the adiabatic approximation to the classical Lagrangian

$$\mathcal{L} = \mathcal{L}\{\dot{\phi}_n, \phi_n; \dot{x}_n, x_n; \dot{y}_n, y_n\}$$
$$= \mathcal{L}_{ex} + \mathcal{L}_{lat} + \mathcal{L}_{int} \tag{1}$$

composed of three components. The first of these, the exciton Lagrangian, is

$$\mathcal{L}_{ex} = \sum_{n=1}^{N} [i\hbar\phi_n^*\dot{\phi}_n - E_0\phi_n^*\phi_n + J\phi_n^*(\phi_{n-1} + \phi_{n+1})], \tag{2}$$

where the dot denotes the differentiation with respect to time $t$, $E_0$ is the amide-I site energy ($E_0 = 0.205\,eV$), $-J$ is the nearest neighbor dipole–dipole coupling energy and $|\phi_n|^2$ defines the probability to find a quantum of the amide-I excitation at the $n$th PG, so that the discrete
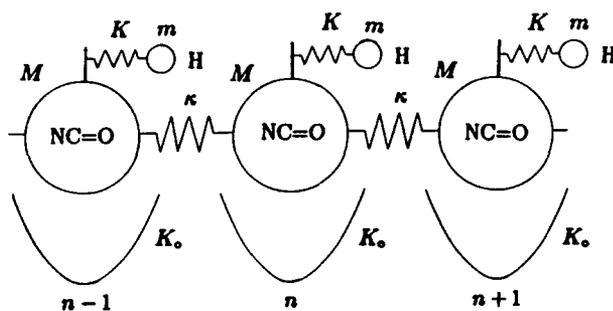


Fig. 1. Schematical representation of the chain model of hydrogen-bonded PGs.

wave function $\phi_n(t)$ is supposed to be normalized as $\sum_n |\phi_n|^2 = 1$.

The second term, the lattice Lagrangian, includes two modes: longitudinal displacements of PGs as whole objects (acoustic mode) and intrapeptide optical oscillations of the proton in the N—H bond with low frequencies as compared with the frequency of the amide-I oscillators. In comparison with the previous studies [4,5], the following two essential modifications are made for this chain model. First, the N—H bond is assumed to move together with its PG and therefore both the acoustic and optical modes come to be coupled. Second, a substrate on-site potential with single-minimum topology is introduced [11] which models the interaction of the chain with its environment as shown in fig. 1. Thus,

$$\mathcal{L}_{\text{lat}} = \sum_{n=1}^{N} \left[ \tfrac{1}{2} m \dot{x}_n^2 + \tfrac{1}{2} M \dot{y}_n^2 - V(y_{n+1} - y_n) \right.$$
$$\left. - \tfrac{1}{2} K (y_n - x_n)^2 - \tfrac{1}{2} K_s y_n^2 \right], \tag{3}$$

where $m$ and $M$ are reduced masses of the proton and PG, respectively; $x_n$ is the proton displacement in the N—H bond (with the stiffness constant $K$) of the $n$th PG and $y_n$ is the longitudinal displacement of this PG; the Morse potential $V(\rho) = \varepsilon_0(e^{-b\rho} - 1)^2$ describes the coupling between adjacent PGs and $\varepsilon_0$ is the energy of the hydrogen bond. For small deformations of this bond $V(\rho) \approx \tfrac{1}{2} \kappa \rho^2$ where $\kappa = V''(0)$ is the stiffness constant of the hydrogen bond. The phenomenological parameter $b$ for the Morse potential can be calculated according to the relation $b = \sqrt{\kappa/2\varepsilon_0}$. The last term describes the interaction of the $n$th PG with the substrate ($K_s$ is the stiffness constant).

The interaction part of the Lagrangian describes the coupling of the amide-I with both the acoustic and optical modes and it is written in the standard form

$$\mathcal{L}_{\text{int}} = \sum_{n=1}^{N} [\chi_1(x_n - y_n) + \chi_2(y_n - y_{n+1})]\phi_n^* \phi_n. \tag{4}$$

where $\chi_1$ and $\chi_2$ are the coupling constants of the amide-I with deformation of the N—H bond and of the hydrogen bond, respectively.

Then the equations of motion which correspond to the Lagrangian (1)–(4) become

$$i\hbar \dot{\phi}_n = [E_0 + \chi_1(y_n - x_n) + \chi_2(y_{n+1} - y_n)]\phi_n$$
$$- J(\phi_{n-1} + \phi_{n+1}), \tag{5}$$

$$m\ddot{x}_n = K(y_n - x_n) + \chi_1|\phi_n|^2, \tag{6}$$

$$M\ddot{y}_n = V'(y_{n+1} - y_n) - V'(y_n - y_{n-1})$$
$$- K(y_n - x_n) - K_s y_n - \chi_1|\phi_n|^2$$
$$+ \chi_2(|\phi_n|^2 - |\phi_{n-1}|^2)$$
$$- \Gamma M \dot{y}_n + \zeta_n(t), \tag{7}$$

where the prime denotes the differentiation of the function $V$ with respect of its argument. Here eq. (7) has the form of the Langevin equation describing the interaction of PGs via the amino acid residues with the thermal bath. As for the N—H optical oscillations, they are thermalized via thermal motions of PGs. The thermal bath with temperature $T$ is described by the relaxation time $t_r = 1/\Gamma$ and the normally distributed random forces $\zeta_n(t)$ with the correlation function $\langle \zeta_m(t) \zeta_n(t') \rangle = 2M\Gamma k_B T \psi(t - t')\delta_{mn}$ where the autocorrelation function $\varphi(t)$ is normalized as $\int_{-\infty}^{+\infty} \psi(t) \, dt = 1$. Using a color noise in eq. (7) we are able to model the "whole" thermalization of acoustic modes and a "partial" thermalization of optical modes. Since some analytical work can be done for the exponential color noise ($\psi(t) = \exp(-|t|/t_c)/2t_c$), it is convenient to use it for our purposes and in this case we may complete the set of eqs. (5)–(7) by the equation

$$\dot{\zeta}_n = t_c^{-1}[\zeta_n(t) - \eta_n(t')]. \tag{8}$$

where the $\eta_n$'s are the delta-correlated random forces (white noise), so that $\langle \eta_m(t) \eta_n(t') \rangle = 2M\Gamma k_B T \delta(t - t')\delta_{mn}$.

## 3. Numerical simulations of the soliton dynamics

For numerical studies of the soliton model we have considered the chain of $N = 50$ PGs with periodic boundary conditions and the following parameter values: $J = 1.55 \times 10^{-22}$ J, $\varepsilon_0 = 0.17$ eV $= 2.72 \times 10^{-20}$ J, $m = m_p = 1.67 \times 10^{-27}$ kg, $M = 113.2 m_p = 1.89 \times 10^{-25}$ kg, $\kappa = 13$ N/m, $l = 4.5$ Å. Calculations for the acoustic and optical modes with $K = 30.5$ N/m and $K_s = 4$ N/m give in the linear limit the following frequency bands: $24.3$ cm$^{-1} \leq \Omega_{ac} \leq 90.9$ cm$^{-1}$ and $719.86$ cm$^{-1} \leq \Omega_{op} \leq 719.91$ cm$^{-1}$. The Einstein temperatures can also be calculated and we have 130.7 K for the acoustic mode and 1035.8 K for the optical mode.

The damping coefficient is determined by the relation $\Gamma = 6\pi b \eta / M$ where $b = 1 \div 10$ Å is the characteristic linear size of an amino acid residue and $\eta$ is the viscosity of medium. For water $\eta = 10^{-3}$ kg/m s, therefore $\Gamma = (1-10) \times 10^{13}$ s$^{-1}$. We choose the minimal value $\Gamma = 10^{13}$ cm$^{-1}$ which corresponds to the relaxation time $t_r = 0.1$ ps.

First, we study pure soliton solutions in the unthermalized chain by using the steepest-descent method [11]. We write

$$\phi_n(t) = \varphi(nl - vt) \exp\{i[kn - (E_0 + \lambda)t/\hbar]\}, \quad (9)$$

where $\varphi$ is a smooth real function, $v$ is the soliton velocity, $k$ is the wave number, and $\lambda$ is a real parameter. Substituting (9) into eq. (5) we find the soliton velocity spectrum: $v = v_{ex} \sin k$ where $v_{ex} = 2lJ/\hbar$ is the maximum velocity of excitons. Also, we derive the discrete stationary Schrödinger equation with the spectral parameter $\lambda$ which can be calculated by using the normalization condition for the envelope $\varphi_n$. Next, inserting (9) into the Lagrangian (1)–(4) and using the approximate relations $\dot{\varphi}_n = -v(\varphi_{n+1} - \varphi_n)/l$, $\dot{x}_n = -v(x_{n+1} - x_n)/l$, and $\dot{y}_n = -v(y_{n+1} - y_n)/l$, we obtain the conditional minimum problem $\mathscr{L} \to$ min: $\Sigma_n \varphi_n^2 = 1$ in the configuration space $\{\varphi_n, x_n, y_n\}_{n=1}^N$ (for more details

see Zolotaryuk et al. [11]). Starting the descent from the point corresponding to the localization of the excitation at one PG we obtain the set $\{\varphi_n^0, x_n^0, y_n^0\}_{n=1}^N$ describing a self-trapping state with its localization width $L = (\Sigma_{n=1}^N \varphi_n^4)^{-1}$. Next, this set is used as initial conditions to simulate eqs. (5)–(7) taken without damping and forcing terms [11].

As was shown previously [11], at $\chi_1 = 0$ and $K_s > 0$ the soliton exists only for large values of $\chi_2$. For its small values there exists only the stable exciton state (note that at $K_s = 0$ the Davydov soliton is stable while the exciton is unstable for all values of $\chi_2$). The descent calculations have shown that in the chain with $\chi_1 = 0$ the static soliton solution exists only if $\chi_2 \geq 0.8 \times 10^{-10}$ N while for $\chi_1 > 0$ it exists for all the values of $\chi_2$. Besides this, with increasing $\chi_1$ (at a fixed value of $\chi_2$) and $\chi_2$ (at a fixed value of $\chi_1$), the soliton energy and its width $L$ are monotonically decreased. Also, with increasing the soliton velocity $v$, the soliton energy is increased while the width $L$ is decreased.

The motion of the amide-I excitation in the thermalized cyclic chain is governed by the system of equations (5)–(8). The dynamical simulations have been carried out for the chain of $N = 50$ PGs with the value $\chi_1 = 1 \times 10^{-10}$ N to be fixed and $\chi_2$ to be varied. The amplitude of thermal oscillations for the acoustic sublattice exceeds by one order the deformation amplitude caused by the amide-I excitation. On the background of these oscillations a local deformation of the acoustic sublattice is not sensible for the excitation and the self-trapping state has to be stablilized only by the coupling of the excitation with a local deformation of the optical sublattice. Therefore the initial conditions for the set of equations (5)–(7) which correspond to a localized self-trapping state can be chosen as follows: $\phi_n(0) = \varphi_n^0 e^{ikn}$, $x_n(0) = \tilde{x}_n + x_n^0$, $y_n(0) = \tilde{y}_n + y_n^0$, $\dot{x}_n(0) = \tilde{\dot{x}}_n - v(x_{n+1}^0 - x_n^0)/l$, $\dot{y}_n(0) = \tilde{\dot{y}}_n - v(y_{n+1}^0 - y_n^0)/l$, $\xi_n(0) = \xi_n$, $n = 1, 2, \ldots, N$, where $v$ is the soliton velocity, $k = \arcsin(v/v_{ex})$, $\{\varphi_n^0, x_n^0, y_n^0\}_{n=1}^N$ is the "pure" soliton solution of

the minimization problem, and $\{\tilde{x}_n, \tilde{y}_n, \overset{*}{x}_n, \overset{*}{y}_n, \xi_n\}_{n=1}^{N}$ is a "lattice" solution of eqs. (6)–(8) at the instant of time $t = 50$ ps, when the lattice is already thermalized, i.e., it is found in the thermal equilibrium with the heat bath.

Take at the initial instant of time the static soliton solution ($v = 0$). In the chain with $\chi_2 = 0$ the soliton is stable against to the thermal oscillations of PGs with any values of the correlation time $t_c$. For the values $t_c = 0.1$; 0.07; 0.04 ps the soliton width practically is not changed while for the values $t_c = 0.01$; 0.001 ps its insignificant oscillations are observed (fig. 2). This can be explained if we notice that the period of the optical oscillations ($\sim0.046$ ps) is less by one order than the minimum time which the excitation needs to be transferred from one unit cell of the chain to its neighbor ($l/v_{ex} = 0.34$ ps). The intramolecular excitation is sensitive to such oscillations only in the averaged way, responding only on the change of the centers of vibrations.

In the chain with $\chi_2 = 1 \times 10^{-12}$ N, thermal vibrations lead only to insignificant oscillations of the soliton width (fig. 3) while the strong interaction of the excitation with the acoustic subsystem ($\chi_2 = 4 \times 10^{-11}$ N) results in the decay of the soliton (fig. 4). This can also be explained if we note that the period of the acoustic oscillations has the same order as the time for the excitation to propagate along the chain one unit cell. In this case, the excitation cannot follow the large-amplitude deformation caused by thermal



Fig. 3. Time dependence of the soliton width $L$ ($v = 0$) in the thermalized chain with $\chi_1 = 1 \times 10^{-10}$ N and $\chi_2 = 1 \times 10^{-12}$ at $t_c = 0.1$; 0.07; 0.04; 0.01; 0.001 ps (curves 1, 2, 3, 4, 5, respectively).
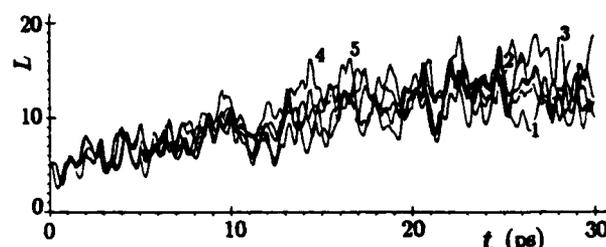


Fig. 4. Time dependence of the localization width $L$ for the static self-trapping state in the thermalized chain with $\chi_1 = 1 \times 10^{-10}$ N and $\chi_2 = 4 \times 10^{-11}$ N at $t_c = 0.1$; 0.07; 0.04; 0.1; 0.001 ps (curves 1, 2, 3, 4, 5, respectively).

oscillations of the molecules which appear randomly along the chain.

## 4. Concluding remarks

To study the thermal stability of the Davydov soliton at biological temperatures, we have considered the model with the interaction of the amide-I excitation with both the acoustic and optical modes. Since the intrapeptide N—H bonds are allowed to move together with their PGs, the acoustic and optical oscillations become to be coupled. Therefore, the optical oscillations can be thermalized via the acoustic mode for which the temperature description is introduced in the standard Langevin form.

The dynamical semiclassical simulations of the time evolution of the intrapeptide amide-I excitation have shown that thermally stable soliton (self-trapping) states can exist if they are stabil-



Fig. 2. Time dependence of the soliton width $L(v = 0)$ in the thermalized chain with $\chi_1 = 1 \times 10^{-10}$ N and $\chi_2 = 0$ at $t_c = 0.1$; 0.07; 0.04; 0.01; 0.001 ps (curves 1, 2, 3, 4, 5, respectively).

ized by the coupling of the excitation with the intrapeptide displacements (optical mode) while the coupling with deformation of hydrogen bonds (acoustic mode) is shown to have destabilizing influence. At the values $\chi_2 = (3-5) \times 10^{-11}$ N given by quantum-mechanical calculations (for references see Scott [2]) the soliton appears to be unstable with respect to thermal oscillations. Therefore, on the surface of a globular protein, where thermal oscillations of hydrogen bonds have large amplitudes, the soliton cannot survive at room temperatures. On the other hand, globular proteins are arranged with high density and therefore the stiffness of a substrate inside a globule can significantly exceed the value $K_0 = 4$ N/m. In this case the frequency of acoustic vibrations is increased while their amplitude is decreased by one order ($\sim 0.01$ Å). At these conditions the soliton becomes stable with respect to the thermal oscillations. Thus, the on-site potential, which models substrate influence [11], plays an important role for the thermal stability of the Davydov soliton.

## Acknowledgement

## References

[1] A.S. Davydov, Solitons in Molecular Systems (Reidel, Dordrecht, 1985).

[2] A.C. Scott, Davydov's Soliton, Phys. Rep. 217 (1992) 1.

[3] A.S. Davydov, Sov. Phys. JETP 51 (1980) 397 [Zh. Eksp. Teor. Fiz. 78 (1980) 789].

[4] P.S. Lomdahl and W.C. Kerr, Phys. Rev. Lett. 55 (1985) 1235.

[5] P.S. Lomdahl and W.C. Kerr, in: Physics of Many Body Systems, A.S. Davydov, ed. (Naukova Dumka, Kiev, 1987).

[6] V.N. Kadantsev, L.N. Lupichov and A.V. Savin, Phys. Stat. Solidi (b) 143 (1987) 569; 147 (1988) 155.

[7] A.F. Lawrence, J.C. McDaniel, D.B. Chang, B.M. Pierce and R.R. Birge, Phys. Rev. A 33 (1986) 1188.

[8] X. Wang, D.W. Brown and K. Lindenberg, Phys. Rev. Lett. 62 (1989) 1796.

[9] W. Förner and J. Ladik, in: Davydov's Soliton Revisited, P.L. Christiansen and A.C. Scott, eds. (Plenum, New York, 1990).

[10] L. Cruzeiro-Hansson, Phys. Rev. A 45 (1992) 4111.

[11] A.V. Zolotaryuk, St. Pnevmatikos and A.V. Savin, in: Davydov's Soliton Revisted, P.L. Christiansen and A.C. Scott, eds. (Plenum, New York, 1990).

[12] G. Careri, U. Buontempo, F. Galluzzi, A.C. Scott, E. Gratton and E. Shyamsunder, Phys. Rev. B 30 (1984) 4689.

[13] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, Phys. Rev. B 30 (1984) 4703.

**PHYSICA** $\square$

# Finite temperature simulations of the semiclassical Davydov model

## L. Cruzeiro-Hansson

*Departement of Crystallography, Birkbeck College, Malet St., London WC1E 7HX, UK*

The thermal stability of soliton solutions of the semiclassical Davydov model for energy transfer in proteins is investigated. Thermal averaging is made taking the degeneracy of the states into account. It is shown that Langelin dynamics leads to the same average states as the semiclassical Monte Carlo simulations, indicating that the former do constitute an accurate representation of the thermal bath in the semiclassical approximation.

## 1. Introduction

In 1973, Davydov [1] proposed a mechanism for energy transfer in proteins, according to which the energy released in the hydrolysis of adenosine triphosphate can propagate along an enzyme $\alpha$-helix in the form of a soliton. It is now well established that the Davydov Hamiltonian does possess soliton solutions at $T = 0$ K. An essential question is, however, whether at physiological temperatures the Davydov soliton lasts long enough to have a biological role [2–4]. Simulations with Langevin equations lead some authors [2,3] to the conclusion that, at $T = 310$ K, the Davydov soliton disappears in a few picoseconds. This conclusion is dependent on an accurate simulation of the effects of the thermal bath. Recently a test to measure the accuracy of the thermal bath representation has been proposed [6]. It aims at comparing average results from the Langevin dynamics (LD) with corresponding results from the exact Monte Carlo (MC) simulations. This article is concerned with the results of this test.

## 2. The semiclassical Davydov model

The Hamiltonian used here, designated as the semiclassical Davydov Hamiltonian $\hat{H}$, is, for a one-dimensional lattice with $N$ sites,

$$\hat{H} = \hat{H}_{\mathrm{ex}} + H_{\mathrm{ph}} + \hat{H}_{\mathrm{int}}, \tag{1}$$

where $\hat{H}_{\mathrm{ex}}$ is the exciton Hamiltonian, $H_{\mathrm{ph}}$ is the phonon Hamiltonian and $\hat{H}_{\mathrm{int}}$ is the interaction Hamiltonian. The exciton Hamiltonian $\hat{H}_{\mathrm{ex}}$ is

$$\hat{H}_{\mathrm{ex}} = \sum_{n=1,N} [\epsilon \hat{a}_n^\dagger \hat{a}_n - J(\hat{a}_n^\dagger \hat{a}_{n-1} + \hat{a}_n^\dagger \hat{a}_{n+1})], \tag{2}$$

where $\epsilon$ is the intramolecular excitation energy, $-J$ is the dipole–dipole interaction energy between neighbouring sites and $\hat{a}_n^\dagger$ ($\hat{a}_n$) is the creation (annihilation) operator for an excitation in site $n$. The interaction Hamiltonian $\hat{H}_{\mathrm{int}}$ is

$$\hat{H}_{\mathrm{int}} = \sum_{n=1,N} [\chi^+ (u_{n+1} - u_n) \hat{a}_n^\dagger \hat{a}_n + \chi^- (u_n - u_{n-1}) \hat{a}_n^\dagger \hat{a}_n], \tag{3}$$

where $\chi^+$ ($\chi^-$) is an anharmonic parameter related to the coupling between the excitation and the displacement $u_n$ of the following (preceding) site. The Hamiltonian is called semiclassical because the lattice is considered a classical system, which is here expressed by treating the displacements $u_n$ (and below the momenta $P_n$) of the lattice as real variables, while the intramolecular

excitation is described by the operators $\hat{a}_n^\dagger$, $\hat{a}_n$. This difference is marked by the hats above the operators. Thus, the phonon Hamiltonian $H_{ph}$ is

$$H_{ph} = \sum_{n=1,N} \left( \frac{P_n^2}{2M} + \tfrac{1}{2}\kappa\,(u_n - u_{n-1})^2 \right), \quad (4)$$

where $P_n$ is the momentum of site $n$, $M$ is the mass of each site and $\kappa$ is the elasticity constant of the lattice. The exact general solution of the Schrödinger equation for the first excited state of the semiclassical Davydov (1–4) is

$$| \psi(\{u_n\},\{P_n\},\{\phi_n\},t)\rangle =$$

$$\sum_{n=1,N} \phi_n(\{u_n\},\{P_n\},t)\,\hat{a}_n^\dagger \mid 0\rangle, \quad (5)$$

where no form is set a priori to $\phi_n$. Exact minimum energy first excited states of the semiclassical Davydov Hamiltonian can be determined by numerically minimizing the Hamiltonian (1–4) averaged over state (5), subject to the normalization condition $\sum_{i=1,N} |\phi_n|^2 = 1$. In the case $J = 0$, the full quantum mechanical Davydov model can be solved exactly and it has been shown that the semiclassical Davydov model leads to the same minimum energy states [7]. For $J \neq 0$ the solution to the full quantum model is not known. In fig. 1 the minimum energy, as determined numerically for the semiclassical Davydov model is compared with that determined approximately by Venzl and Fischer [5] for the full quantum mechanical Hamiltonian. Letting $\mu = \chi^2 \kappa J$, (with $\chi = \chi^+ = \chi^-$), formula (39) of [5], p. 6093, was used when $\mu < 1$ and formula (55) of [5], p. 6094, was used when $\mu > 1$. Fig. 1 shows that, also for $J \neq 0$, the semiclassical Davydov model leads to approximately the same states as the full quantum mechanical model.
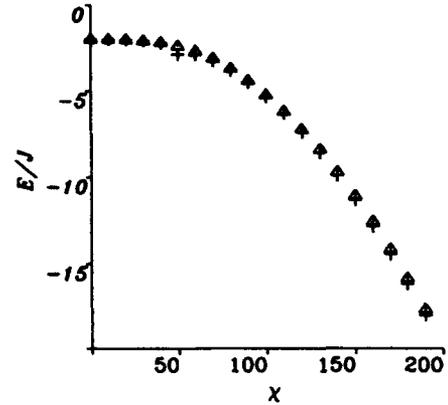


Fig. 1. Variation with $\chi$ of the energy of minimum energy excited states determined with semiclassical Davydov model (triangles) and with the full quantum mechanical model (crosses). $N = 50$, $J = 1.549 \times 10^{-22}$ J, $\kappa = 13$ N/m.

## 3. Monte Carlo simulations and Langevin dynamics

The variable selected to compare the results between MC and LD simulations is the thermodynamic average state, as specified by the average probability per site $n$, $\langle\langle|\phi_n|^2\rangle\rangle$, and the corresponding average displacement differences, $\langle\langle u_n - u_{n-1}\rangle\rangle$. This average state is determined by summing $|\phi_n|^2$ and $u_n - u_{n-1}$ over all sampled states, after having located, for each sampled state, the site for which the probability of excitation is maximum and translated this maximum to the middle of the chain and performing ·tly the same rotation with the displacement ·erences. Fig. 2 displays the average state as calculated from the MC (solid line) and from the LD (broken line) simulations, respectively, for the same values of the parameters (see figure caption).

The MC simulations involved more than 1 million passes and the LD were run for more than 100 nanoseconds. Fig. 2 shows that the correlation between the site of maximum excitation and the lattice deformation that was observed in the MC simulations is not destroyed by the LD. In fact, it is present in the LD with approximately the same strength (i.e. the value of the lat-

as the MC simulations indicates that the former constitutes indeed a faithful representation of the thermal bath. This supports the conclusion that the initial solutions used by other authors do lead to a soliton lifetime of a few picoseconds at 310 K [2,3]. These conclusions are valid within the semiclassical Davydov model. However, the results in section 2 indicate that they are approximately valid for the full quantum mechanical model. A few points remain however, to be studied. They include the temperature dependence of the parameters, a decrease in $J$ as temperature increases will lead to more stable solitons at higher temperatures and especially the second excited state has yet to be explored.

## References

[1] A.S. Davydov, J. Theor. Biol. 38 (1973) 559.
[2] P.S. Lomdahl and W.C. Kerr, Phys. Rev. Lett. 55 (1985) 1235.
[3] A.F. Lawrence, J.C. McDaniel, D.B. Chang, B.M. Pierce and R.R. Birge, Phys. Rev. A 33 (1986) 1188.
[4] W. Förner and J. Ladik, Influence of heat bath and disorder on Davydov solitons, in: Davydov's Soliton Revisited, P.L. Christiansen and A.C. Scott, eds. (Plenum, New York, 1990).
[5] G. Venzl and S.F. Fischer, J. Chem. Phys. 81 (1984) 6090.
[6] L. Cruzeiro-Hansson, V.A. Okhonin, R.G. Khlebopros and I.N. Yassievich, Nanobiology (1993), in press.
[7] L. Cruzeiro-Hansson, Phys. Rev. A 45 (1992) 4111.

Fig. 2. Plot of the thermally averaged probability (upper) and displacement differences (lower) determined from the MC simulations (solid line) and the LD (broken line). $N = 50$, $J = 1.549 \times 10^{-22}$ J, $\kappa = 13$ N/m, $\chi = 62$ pN.

tice deformation is the same) as in the MC simulation. The conclusion is thus that the LD does produce an accurate representation of the thermal bath in the semiclassical Davydov model.

## 4. Discussion

The semiclassical Davydov model can be solved exactly and thus the results presented here do not suffer from the ambiguities that plague the full quantum mechanical model, where ansätze for the wavefunctions are proposed whose degree of approximation is not known. The MC simulations also provide an exact simulation of the system at finite temperatures. The fact that LD leads to the same results

PHYSICA □

# Effects of temperature and interchain coupling on Davydov solitons

Wolfgang Förner

*Chair for Theoretical Chemistry and Laboratory of the National Foundation of Cancer Research, Friedrich-Alexander University Erlangen-Nürnberg, Egerlandstr. 3, W-8520 Erlangen, FRG*

*Dedicated to Professor Alwyn C. Scott
on the occasion of his 60th birthday*

After a short review of our previous results on temperature effects on Davydov soliton dynamics, we give some comparisons between results obtained with Davydov's average Hamiltonian method within the $|D_1\rangle$ ansatz state in one chain with quantum Monte Carlo results found in the literature. We find that Davydov's method leads to quantitatively incorrect results, but reproduces the qualitative trends correctly. Our results obtained with different models point into the direction that Davydov solitons should be stable, if the value of the spring constant for hydrogen bonds is larger than usually assumed. Therefore we return to Scott's suggestion, that in order to simulate the three coupled hydrogen bonded chains present in protein $\alpha$-helices with one chain, the spring constant and the mass should be enlarged. In case of the $|D_1\rangle$ state we present numerical simulations for three coupled chains including temperature with different initial states and for one chain using Scott's revised parameters at 300 K.

## 1. Introduction

For the mechanism of energy transport through proteins Davydov [1,2] suggested that the energy of $\approx 0.4$ eV released by hydrolysis of adenosine-triphosphate (ATP) could be transported in quanta of the amide-I (mainly C=O stretch) vibration ($\approx 0.2$ eV). The CO groups participate in hydrogen bonds which form chains parallel to the axis of $\alpha$-helical proteins. Thus the amide-I vibration interacts with the acoustic phonons in these chains. The [1,2] excitation of an amide-I oscillator causes a distortion in the lattice which in turn stabilizes the amide-I excitation. It was found that for certain regions of the parameter space of the model this effect can prevent the excitation from dispersion via the dipole – dipole coupling between neighboring CO-groups in the lattice. The region in which the vibrational energy is localized can travel as a soliton along the chain.

In his original theory Davydov [1] used an ansatz for the wave function ($|D_2\rangle$) which treats the lattice classically. At zero temperature is has been confirmed that Davydov solitons exist for parameter values appropriate for proteins [3]. Also their stability against disorder along the chain was studied [4]. The investigation of temperature lead to controversial results. Halding and Lomdahl [5] found stable pulses at $T = 310$ K using classical molecular dynamics for peptide units moving in a Lennard-Jones potential. Lomdahl and Kerr [6a] and others [7] used the $|D_2\rangle$ ansatz together with a damping and a noise term to introduce temperature and found no stable solitons at 310 K at a specific set of parameters. Bolterauer [8] argued that their classical thermalization scheme might lead to a too large transfer of energy into the quantum system (oscillators). Cottingham and Schweitzer [9] applied perturbation theory to the Hamiltonian after partial diagonalization and could show (again for

one set of parameters) that the soliton life-time at 300 K is too short for biological processes. In our previous work [4,10,11a] we prepared the lattice in a thermally excited state prior to the soliton start. We compared our results to those of [6a] and found agreement between the models if in the Langevin model [6a] the lattice is thermally equilibrated before the soliton starts. We could show that in a window in the parameter space which might well be realistic for proteins travelling solitons exist at 300 K.

Recently Brown et al. [12] have shown that the $|D_2\rangle$ state ansatz does not reproduce the dynamics of the exactly solvable small polaron limit (dipole – dipole coupling neglected). Davydov [2] introduced a more sophisticated ansatz state ($|D_1\rangle$) which allows for quantum effects in the lattice. However, he used the energy expectation value for $|D_1\rangle$ as a classical Hamiltonian function to derive equations of motion [2]. It was shown that with these equations $|D_1\rangle$ does not reproduce the small polaron limit [12] either. With these equations of motion and a thermally averaged Hamiltonian Davydov [2] could show within the continuum limit that solitons exist at 300 K. Cruzeiro et al. [13] reached at the same conclusion numerically without making use of Davydov's approximations, but using also the thermally averaged Hamiltonian as a classical Hamiltonian function.

Most recently Mechtly and Shaw [14] and Skrinjar et al. [15] could derive new equations of motions for $|D_1\rangle$ with help of quantum mechanical methods. These equations of motion reproduce the small polaron limit. We also used the Langrangian method described in [15] to obtain correct equations of motion for the $|D_1\rangle$ ansatz state from the thermally averaged Hamiltonian derived in [2,13]. In this investigation [16–18], as well as in our previous studies within the $|D_2\rangle$ state, summarized in [17] we found that Davydov solitons should be stable at 300 K if the spring constant of the hydrogen bonds is larger than previously assumed. These results are shortly reviewed in the first two parts of section 2 of

this work. There are doubts, if the Davydov concept of using a thermally averaged Hamiltonian to derive equations of motion from it, is in agreement with statistical mechanics. There is the possibility that it may lead to results which are even qualitatively misleading. Therefore we present in the third part of section 2 a comparison of our results obtained with the averaged Hamiltonian method with the exact quantum Monte Carlo results of Wang et al. [19].

Since all our results point into the direction, that the hydrogen bond spring constant should be large to allow soliton formation at 300 K, we go back to the suggestion of Scott [3,20,21] that in one chain simulations the spring constant should be larger by a factor of three, in order to simulate the three coupled chains present in real protein $\alpha$-helices within a one chain model. For an excellent review of the state of art of work on Davydov solitons the reader should consult Scott's recent paper [21]. In the first part of section 3 we present one chain simulations with revised parameters according to Scott's suggestion. In the second part of section 3 we present simulations on three chains using different excitations (symmetric as well as asymmetric ones) within the $|D_2\rangle$ ansatz state and compare them with one chain results, using different parameter values. Finally in the third part some examples of three chain simulations including temperature are discussed.

## 2. One-chain simulations at finite temperature

### 2.1. The $|D_2\rangle$ ansatz state

The Hamiltonian used for this study is in the most simple form for the system investigated by Davydov [1]. More sophisticated forms of the Hamiltonian which incorporate more details of the protein structure have lead qualitatively to the same results [3].

$$\hat{H} = \sum_n (E_0 \hat{a}_n^+ \hat{a}_n - J(\hat{a}_n^+ \hat{a}_{n+1} + \hat{a}_{n+1}^+ \hat{a}_n)$$

$$+ \frac{\hat{p}_n^2}{2M} + \tfrac{1}{2} W(\hat{q}_n - \hat{q}_{n-1})^2$$

$$+ X \hat{a}_n^+ \hat{a}_n (\hat{q}_n - \hat{q}_{n-1}) \,. \qquad (1)$$

In eq. (1) $\hat{a}_n^+$ ($\hat{a}_n$) are the usual boson creation (annihilation) operators [3] for the amide-I oscillators at sites $n$ (see fig. 1).

From infrared spectra the ground state energy of an isolated amide-I oscillator can be deduced ($E_0 = 0.205$ eV). Usually for all parameters in eq. (1) site-independent mean values are used. The average value for the dipole – dipole coupling between neighboring amide-I oscillators is $J = 0.967$ meV. The average spring constant of the hydrogen bonds is taken usually to be $W = 13$ N/m. $\hat{p}_n$ is the momentum and $\hat{q}_n$ the position operator of unit $n$. The average mass $M$ is taken as that of myosine ($M = 114m_p$; $m_p = $ proton mass). The energy of the CO stretching vibration in hydrogen bonds is a function of the length $r$ of the hydrogen bond ($E = E_0 + Xr$). For $X$ the experimental estimates are 35 pN and 62 pN.

For the solution of the time dependent Schrödinger equation we used the displaced oscillator state ansatz ($|D_2\rangle$) of Davydov [1]. Davydov [1,2] formed the expectation value of the Hamiltonian (1) with $|D_2\rangle$ and used this expectation value as classical Hamiltonian function. In this way he obtained the equations of motion. Explicit forms of the equations of motion used can be found in ref. [11a]. The $|D_2\rangle$ state reproduces the lattice dynamics for $J = 0$ correctly, but leads to an incorrect phonon energy [12]. From our
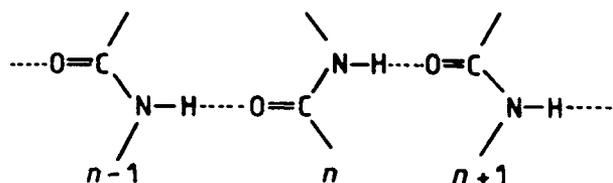


Fig. 1. Schematic picture of a hydrogen bonded channel in a protein.

results we deduced that within the $|D_2\rangle$ model Davydov solitons should be stable at 300 K if the spring constant $W$ of the hydrogen bonds is larger than roughly 30 N/m (see [11a,17] for details).

A very appealing model for treating temperature effects on Davydov solitons was recently introduced by Cottingham and Schweitzer [9]. They [9] could diagonalize the Hamiltonian partially. Then one can define $\hat{H} = \hat{H}_0 + \hat{W}$ where $\hat{H}_0$ is the diagonal part of $\hat{H}$ and $\hat{W}$ the non-diagonal one. Thus $\hat{W}$ can be treated as a perturbation. This partitioning is reasonable, and for their purpose unique, because the Davydov soliton state ($|D_2\rangle$) is an exact eigenstate of $\hat{H}_0$ and thus first order perturbation theory using $\hat{W}$ as perturbation allows the calculation of soliton lifetimes. They could compute the transition probability from an initial state containing a Davydov soliton and a thermal distribution of phonons to a final state without the soliton. From this probability they could derive an explicit expression for the life-time of a pinned soliton.

To compare our model with that of Cottingham and Schweitzer [9] we took a cyclic chain as in ref. [9] where only pinned solitons are found numerically [9]. We used $W = 13$ N/m. $X = 62$ pN and $a_n(0) = A \operatorname{sech}[(a_n - 100)X^2/WJ]$ as initial condition as in [9]. In addition we implemented the symmetric interaction into our program for the purpose of comparison, since in ref. [9] this model was used. As shown in [13] the two interaction models lead to rather different results. For comparison we varied $J$ and estimated soliton life-times from our model. For this purpose the initial excitation ($|D_2\rangle$ soliton solution) was centered around site 100 and we used the soliton detector plot of Lomdahl and Kerr [6a]. We found [11a,17] that the two models agree very well. However, one should keep two restrictions in mind: in case of small $J$ the continuum limit used in [9] does not apply and the formula for the life-time given in [9] still contains approximations which were removed in a later paper [23]. Thus the nearly quantitative

agreement [11a,17] between the two models should be considered as accidental. However, we expect that also with the improved perturbation treatment [23] qualitative agreement should still be found. With the new formula [23] for the parameters life-times between 0.30 ps and 322 ps for $J$ between 1 meV and 0.1 meV were found [24]. This still agrees qualitatively with the range of 0.6–96 ps from our model, at least for larger $J$ values. For the smaller ones the continuum approximation used in [23] breaks down.

However, studies on Davydov soliton life-times should not be based on the symmetric interaction ansatz, where the length of a hydrogen bond in which the C=O group does not take part is considered as to influence the C=O group as much as the hydrogen bond in which it participates. This seems to be rather unrealistic. To give some examples, we performed similar calculations as reported above for parameter values where we found stable solitons at 300 K with asymmetric interaction and open (instead of cyclic) boundaries. In case of $X = 110$ pN and $W = 90$ N/m we find a stable soliton in our simulations, observed in the computer experiment for more than 100 ps. However, if we perform the calculation as described above with symmetric interaction and cyclic boundaries we obtain a life-time of only 3–4 ps. This again is qualitatively in the same order of magnitude as perturbation theory which gives 0.8 ps in this case [24].

In general it seems that $|D_2\rangle$ dynamics which approximates the lattice classically leads to solitons which are more stable against temperature than in theories which apply the exact Hamiltonian, like perturbation theory [9,23]. Also Bolterauer [25] who studied the uncertainty of the soliton energy in order to obtain lifetimes found that both quantum and temperature fluctuations should destroy the soliton within less than 1 ps. However, these studies have in common, that the decay time of a soliton as given by the exact continuum limit solution for the $|D_2\rangle$ ansatz state is investigated within the exact Hamilto-

nian. It seems to be more realistic to investigate the dynamics of an initial one- or two-site excitation than that of a preformed soliton obtained from $|D_2\rangle$ continuum theory. As we found in our studies, the solitons formed from such an excitation have properties which can differ considerably from the continuum solutions. To study quantum effects on solitons one has to compute the dynamics of such initial excitations with better ansatz states, instead of calculating the life-time of $|D_2\rangle$ solitons within the exact Hamiltonian, since these $|D_2\rangle$ solitons are only solutions of $\hat{H}_0$ and not of $\hat{H}$. The full Hamiltonian $\hat{H}$ can have soliton solutions which might differ drastically from these approximative $|D_2\rangle$ solitons. Since the Schrödinger equation for $\hat{H}$ cannot be solved exactly, besides in special cases, we applied as a first step the improved $|D_1\rangle$ ansatz state of Davydov [2] which allows quantum effects in the lattice and also reproduces the exact solutions in the small polaron limit ($J = 0$) if the equations of motion are correctly derived [14,15] but is still an approximation for $J \neq 0$. The results of these investigations are reported in the next section.

## 2.2. The $|D_1\rangle$ ansatz state

In this chapter we want to report on our results at $T = 300$ K using the $|D_1\rangle$ state ansatz [17,18]. The Hamiltonian [1,2] in second quantized form including disorder is given e.g. in [17,18]. The $|D_1\rangle$ ansatz for an inclusion of temperature in Davydov's approximation for solution of the time dependent Schrödinger equation is [22]

$$|D_1, \nu\rangle = \sum_n a'_n(t)\hat{a}_n^+ |0\rangle_c |\beta_n, \nu\rangle. \qquad (2)$$

Here $|0\rangle_c$ is the exciton vacuum, and $|\beta_n\rangle$ a coherent phonon state. For the one-quantum oscillator states used here $\Sigma|a_n|^2 = 1$ holds. To include temperature approximately we assume, as in [13] that a phonon distribution is present in the lattice where each normal mode is occupied

by $\nu_k$ quanta. We do not consider a thermal distribution of amide-I quanta since at 300 K the Boltzmann factor implies that only 3 of 10 000 amide-I oscillators would be thermally excited. Thus one can neglect a possible thermalized soliton distribution in the system too, since presence of solitons requires first of all amide-I excitation. Then the expectation value of the Hamiltonian in this state is formed and thermally averaged over all phonon distributions. From this equations of motion are derived using the Euler–Lagrange formalism [17,18].

This method introduced by Davydov was criticized by several authors as being inconsistent with statistical mechanics, since equations of motion are obtained from a thermally averaged Hamiltonian. However, we feel that it might still be a reliable approximation to the real dynamics under physiological temperature. To investigate this we also give a comparison with quantum Monte Carlo simulations in the next section. However, since in previous work on Davydov's method [2,13] his incorrect derivation of equations of motion was used, it seems to be interesting in itself, to study the dynamics of the system using correct equations derived from the averaged Hamiltonian.

The equations of motion have been solved using a fourth order Runge–Kutta method [19] and the dynamics were calculated through $\approx 26$ ps. We performed a survey of the $(X, W)$ parameter space. At $T = 0$ K solitons occur at much larger values of $X$ ($>150$ pN [16]) and smaller values of $W$ than at 300 K [17,18]. The reason for that is, that effectively the dipole – dipole coupling $J$ which is responsible for dispersion decreases with increasing temperature. In this way a larger temperature can stabilize solitons at smaller values of $X$. This is consistent with results reported by Cruzeiro et al. [13], although the effect is far more pronounced in our case using the quantum mechanically correct equations of motion. Thus we arrive with $|D_1\rangle$ dynamics at the same conclusion as with $|D_2\rangle$ dynamics: Davydov solitons should be stable at

300 K if the spring constant of the lattice is much larger than the usually applied 13 N/m. This leads us first of all to the suggestion of Scott [3,21] that one has to take into account that in an $\alpha$-helix three coupled chains are present. Scott argued that in one chain dynamics one has to use a larger effective value for $W$ and $M$ in order to simulate the three chain case with a one chain model. We will investigate this in section 3 in more detail. But before turning to this point, we want to discuss some basic problems of the models applied. One may ask if it is correct to introduce temperature in an averaged way as done here and if not fluctuations taken explicitely into account could destroy the solitons. However, we have shown in our previous work [11,17] that in case of the $|D_2\rangle$ ansatz models with explicit random fluctuations and our model using a lattice prepared to $T = 300$ K in a deterministic way lead to comparable results. Thus we are confident that in the $|D_1\rangle$ case the same conclusion holds. The other problem is the approximative nature of the $|D_1\rangle$ state. To improve the ansatz in this respect we plan to take two-phonon terms into account in the generator of the unitary transformation [14]. A more severe problem is the inconsistency of Davydov's method for incorporation of temperature with statistical mechanics. Thus we want to compare $|D_1\rangle$ results with exact quantum Monte Carlo simulations in the next section, to be able to decide if Davydov's method can still be considered as an at least qualitatively correct approximation.

### 2.3. Comparisons with exact quantum Monte Carlo results

Wang et al. [19] reported the results of quantum Monte Carlo simulations on the Davydov Hamiltonian. These results should describe the equilibrium state in principle exactly, restricted only by numerical inaccuracies which can be controlled. They applied the parameters usually used in the literature ($J = 0.967$ meV,

$W = = 13 \, \text{N/m}$, $X = 62 \, \text{pN}$, $M = 114 m_p$), cyclic boundaries, the symmetric interaction, and rings of 24 sites. They determined for each of their configurations the excitation site $n_0$, and rotated all coordinates such that this site is in the middle of their lattice. Then they computed the average $\langle A_n \rangle$ of the lattice displacements $A_n = q_{n+1} - q_{n-1}$. Here $n$ refers to the rotated coordinate system, where the excitation site $n_0$ is always in the center. We implemented cyclic boundaries and symmetric interaction into our time simula-

tion programs, started from a random distribution of one amide-I vibrational quantum and performed a time average of the same quantity with which Wang et al. performed their ensemble average. We determined $n_0$ in each time step as the site where the excitation probability $|a_n(t)|^2$ is largest. In this way after a sufficient number of time steps (convergence of $\langle A_n \rangle$) we should obtain results comparable with those from Wang et al. [19].

In [19] it is reported, that at 2.8 K coherent
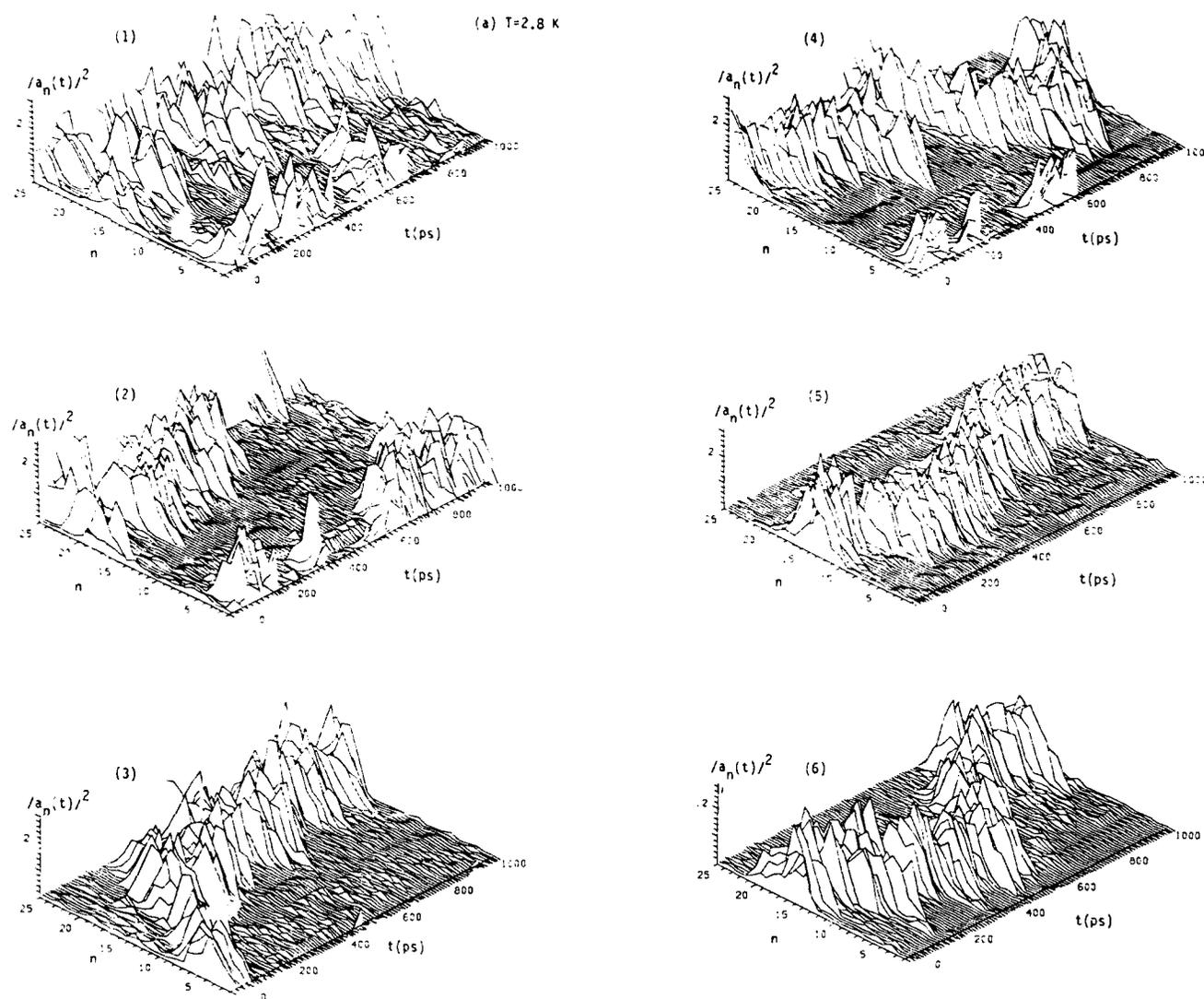


Fig. 2. Evolution of the excitation probability $|a_n(t)|^2$ from an initially randomly distributed amide-I quantum as function of site $n$ and time $t$ in cyclic chains of 25 units and with symmetric interaction, using the $|D_1\rangle$ ansatz state and Davydov's model for temperature effects. Each subfigure (1–6) shows every 5 000th of 500 000 time steps each ($\tau = 2 \, \text{fs}$). (a) $T = 2.8 \, \text{K}$; (b) $T = 7.0 \, \text{K}$; (c) $T = 11.2 \, \text{K}$.
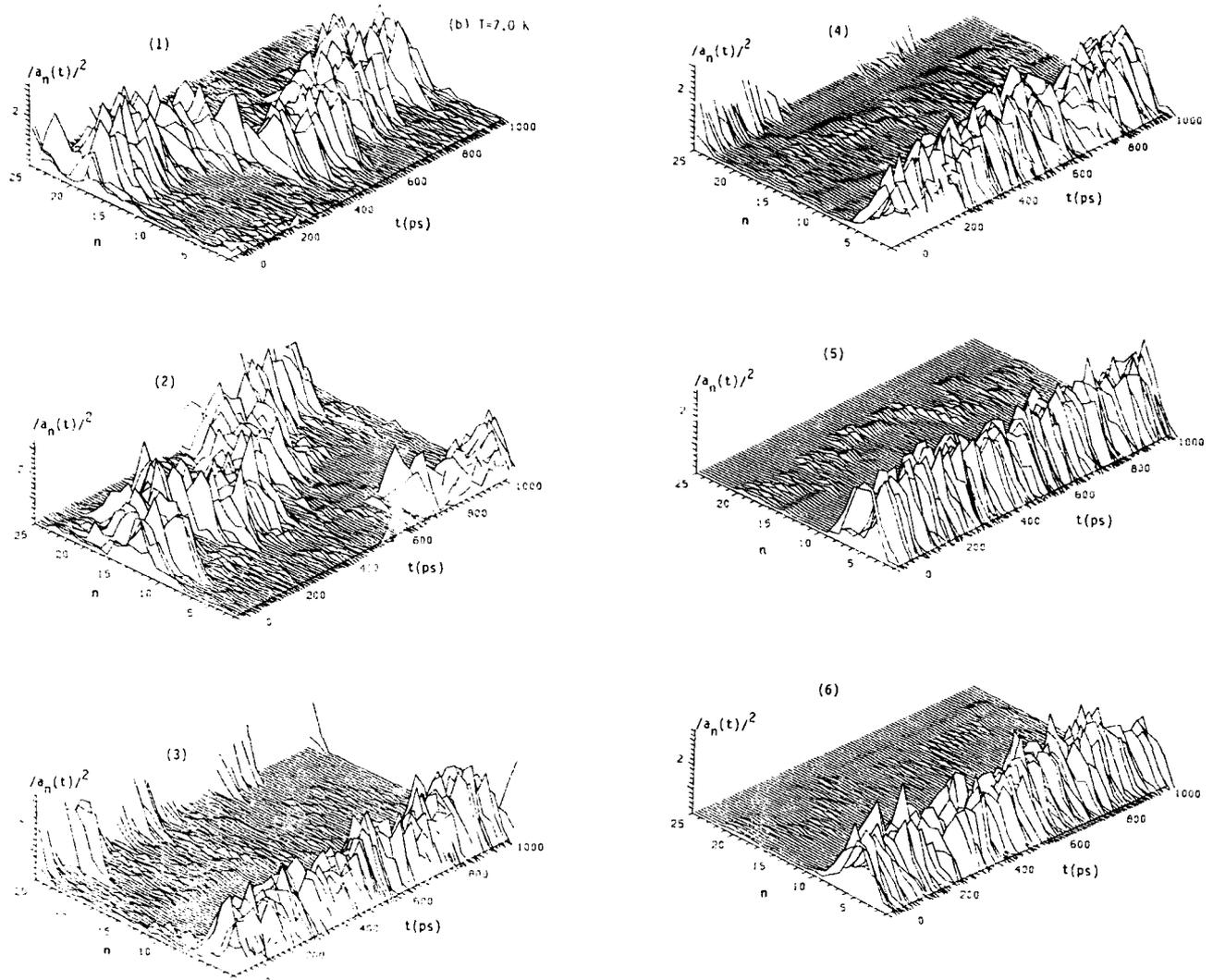
Fig. 2 (cont.).

structures are obtained. In this case from their fig. 1 we estimate a maximum of $\langle A_n \rangle$ at $n = n_0$ of $\approx 0.07$ Å. For $T = 7.0$ K they observe that this coherent structure starts to break down, leading at 11.2 K to localized structures comparable to small polarons (this might also be interpreted as an Anderson-like localization originating from increased disorder due to thermal fluctuations). For 7.0 K the maximum of $\langle A_n \rangle$ is found to be $\approx 0.08$ Å and $\approx 0.09$ Å for 11.2 K. The latter value is already close to the infinite temperature limit of $\approx 0.095$ Å [19].

In fig. 2 we show our result obtained with the

$|D_1\rangle$ state and Davydov's model for temperature effects, however, using correct equations of motion as given in the previous section. We followed the dynamics over a period of 6 ns, corresponding to 3 000 000 time steps. The calculation for one temperature had to be done in 6 runs, where one requires 7.7 CPU hours computation time on a Cyber 995E computer from Control Data Corporation (500 000 time steps each run). Typically $(T = 11.2$ K) the error in total energy in such runs is $\approx 3$ neV and the norm error is less than 0.05 ppm (parts per million). We see from fig. 2a that at $T = 2.8$ K after roughly 1 ns a
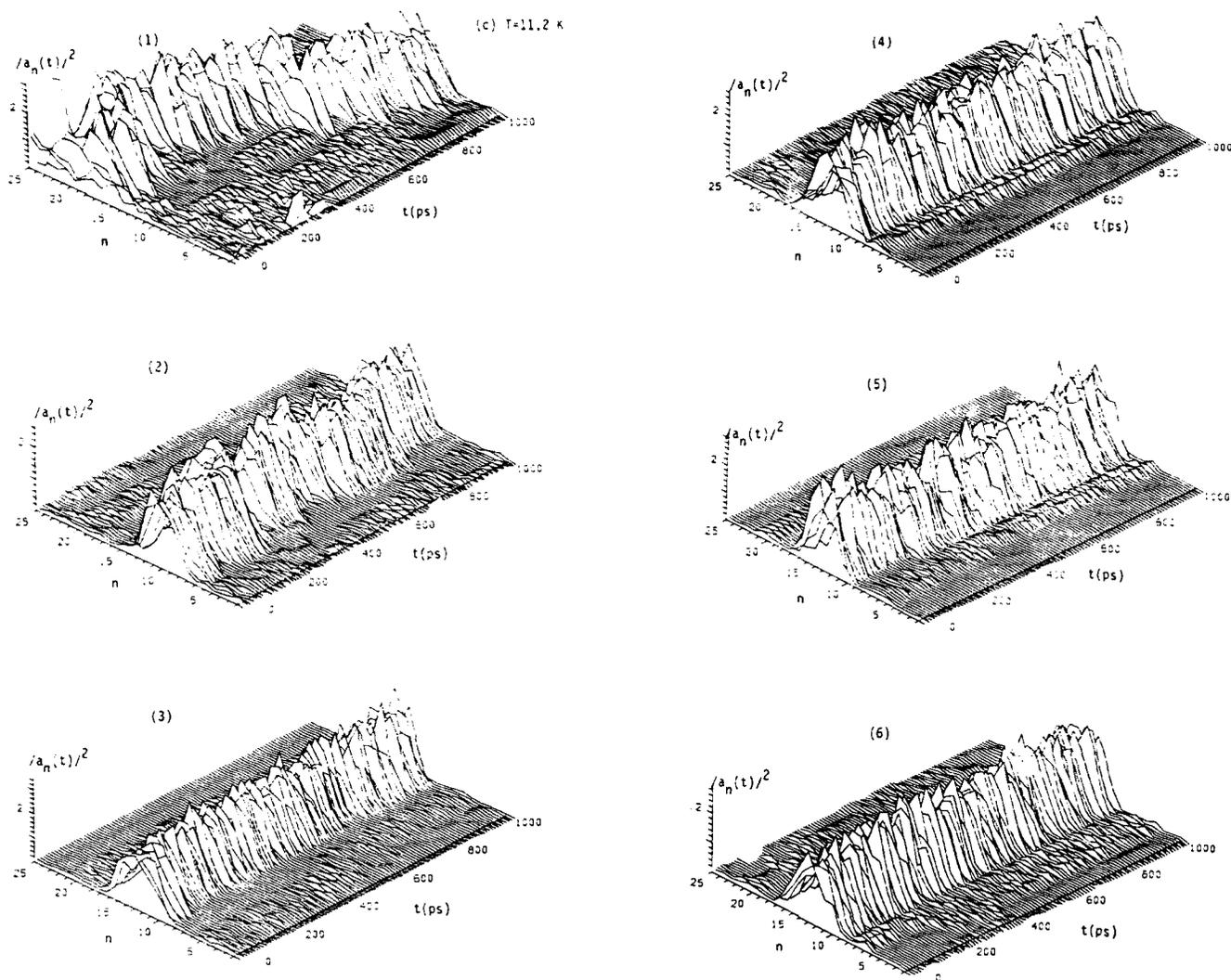
Fig. 2 (cont.).

soliton nucleates from the random initial conditions and performs a random walk in the system. This corresponds qualitatively to the coherent structure found by Wang et al. [19] at this temperature. At $T = 7.0$ K (fig. 2b) also such a localized packet forms, but it is first of all smaller and after 2 ns it remains confined within a few lattice sites, again corresponding to the destruction of the coherent structure towards a localized state reported in [19] for this temperature. Finally at 11.2 K the packet remains in the middle of the chain and becomes smaller, again in qualitative agreement with [19], although it seems that in $|D_1\rangle$ theory the building of small polaron-

like structures is shifted to higher temperatures. Thus Davydov's model appears to be in qualitative agreement with quantum Monte Carlo results. However, it does not agree quantitatively as fig. 3 shows where we present the time average $\langle A_n \rangle$ of the lattice displacements through 3 000 000 time steps. Obviously the structure found has a decreasing width with increasing temperature, but the peak values are $\approx -0.026$ Å (2.8 K), $\approx -0.028$ Å (7.0 K) and $\approx -0.029$ Å (11.2 K) in contrast to the much larger values found in [19] ($-0.07$ Å, $-0.08$ Å, $-0.09$ Å, respectively). Also the increase of this peak value with temperature is much less pronounced than
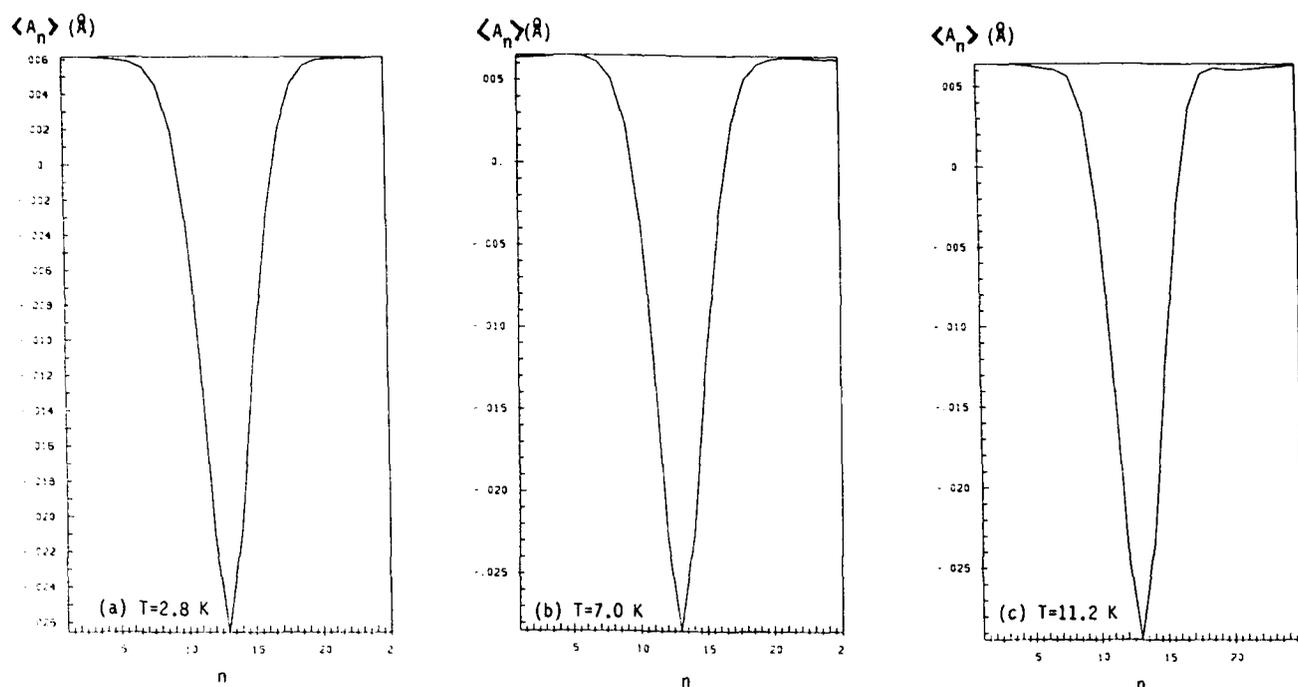
Fig. 3. Time average $\langle A_n \rangle$ over 3 000 000 time steps (6 ns) of the lattice displacements $A_n$ in the rotated coordinate system for the system of fig. 2 at the same temperatures.

in the quantum Monte Carlo case. Therefore we conclude that Davydov's $|D_1\rangle$ model gives a qualitatively correct picture, but fails to reproduce exact results quantitatively.

Let us turn now to the partial dressing state of Brown and Ivic [26] which is a special case of the $|D_1\rangle$ ansatz with a fixed dependence of the coherent state amplitudes $b_{nk}$ on the site $n$. Our simulations at 0 K for this state have shown that the results are rather similar to $|D_1\rangle$ dynamics, however, with solitons appearing at even higher values of $X$ than in zero-temperature $|D_1\rangle$ theory [16]. Our implementation of the partial dressing theory including temperature is described in [16] and we do not repeat it here. In case of this state already after 900 000 time steps of 2 fs convergence of the average $\langle A_n \rangle$ is obtained. In this case no solitons are nucleated. We found that here the infinite temperature value from [19] for the peak is underestimated by all three curves. Thus even at 2.8 K where in [19] coherent structures were found, the partial dres-

sing ansatz shows a narrow and too high peak for all three temperatures. The peak does not change for increasing temperature and thus for this ansatz we even do not find qualitative agreement with quantum Monte Carlo results.

Finally we computed again the average $\langle A_n \rangle$ for the $|D_2\rangle$ ansatz with our temperature model. In the $|D_2\rangle$ case the peak values of $\langle A_n \rangle$ are again smaller than the corresponding quantum Monte Carlo values, however, with the wrong tendency: they decrease somewhat with increasing temperature (2.8 K: $-0.0107$ Å; 7.0 K: $-0.0106$ Å; 11.2 K: $-0.0098$ Å). The decrease is not very pronounced and it seems that all three values are far off the infinite temperature value given in [19]. Since there are no solitary or small polaron like structures visible in the $|a_n|^2$ plots and the infinite temperature behavior shows already up at $T = 2.8$ K it seems that here temperature effects are overestimated in agreement with Bolterauer's [8] argument. In conclusion it seems that among the models studied, the

$|D_1\rangle$ ansatz together with Davydov's treatment of temperature is the only one which gives an at least qualitatively correct picture, although it is quantitatively incorrect and furthermore there are doubts on the validity of Davydov's ansatz for the description of temperature. However, it seems that Davydov's ansatz can be viewed as a qualitatively valid approximation (see [27] for details).

## 3. Three chain dynamics

### 3.1. One chain dynamics at $T = 300$ K with enlarged mass

In this paragraph we want to follow Scott's suggestion to use revised parameters in one chain dynamics in order to simulate three chain dynamics with them. Since we want to survey the $(X, W)$ parameter space, we only need to change the site mass from $114m_p$ to $342m_p$. However, we have to keep in mind, that then in these simulations $W$ is an effective spring constant and no longer the spring constant of an individual hydrogen bond. We used chains of 50 units and an initial excitation at one site (49) as in the calculations reported before. The temperature was 300 K using Davydov's model for temperature effects and the $|D_1\rangle$ ansatz state. The time step was 0.15 fs and we followed the dynamics over roughly 26 ps. In case of the $|D_1\rangle$ state Scott's analytical considerations for an $A$-mode excitation within the $|D_2\rangle$ state which are repeated shortly in the next section do not hold. Thus we cannot expect that also for the $|D_1\rangle$ ansatz the three chain case is reproduced by one chain dynamics if we use $3W$ and $3M$ instead of $W$ and $M$.

The results of our calculations are displayed in fig. 4 Since the enlarged mass influences considerably the phonon frequencies we have to expect that interactions between the soliton and sound waves in the lattice might be different from the one chain case with smaller site mass. Indeed
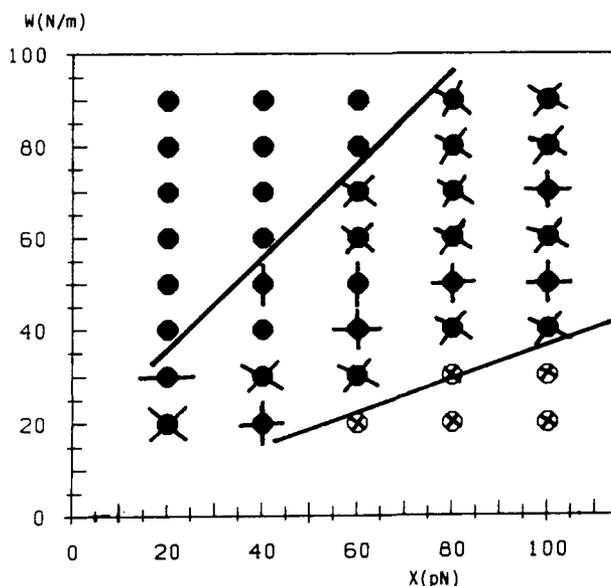


Fig. 4. Survey of the $(X, W)$ parameter space at $T = 300$ K using Davydov's temperature model and the $|D_1\rangle$ ansatz state together with a mass of $3M = 342m_p$ (initial excitation at site 49, time step 0.15 fs, chains of 50 units, crossed open circles: pinned soliton, horizontally dashed black circles: pinning or reflection after roughly 40 sites, vertically dashed black circle: slowly dispersive solitary wave, horizontally crossed black circle: soliton which becomes pinned after a few sites, diagonally crossed black circle: moving soliton which becomes pinned after interaction with the shock wave and then slowly disperses, black circle: travelling soliton).

this is the case as the figure shows. We find many parameter values where the solitons become pinned after interaction with the sound wave reflected from the chain end. After that interaction in most cases the solitons disperse slowly and perform a random walk around the pinning site. For $W = 80$ N/m and $X = 60$ pN the soliton is even reflected from the shock wave, but not destroyed. Therefore we marked this point in the parameter space as a point where travelling solitons exist. In general we see that the picture is very similar to that for the reduced mass discussed in section 2.2. The region of stable solitons is again shifted to small $X$- and rather large $W$-values, and the boundary of this region appears to be roughly linear. However, here $W$ has another meaning than in section 2.2. Thus the threshold for soliton formation of $W \approx 40$ N/m

should correspond to a spring constant for single hydrogen bonds of roughly 13 N/m which is the usually cited value, while the also used value of 19 N/m corresponds here to $W \approx 60$ N/m. However, the factor of 3 holds only for $|D_2\rangle$ theory, while for $|D_1\rangle$ this is not the case. Further the figure shows that for a value $X = 60$ pN $W$ has to be larger than 70 N/m to allow travelling solitons, which corresponds to a hydrogen bond spring constant of $\approx 25$ N/m.

### 3.2. Three chain simulations at $T = 0$ K

As Scott pointed out already (see [21] for a recent review) a long time ago [3], in an $\alpha$-helix three parallel chains of hydrogen bonds exist, which are coupled by dipole–dipole interactions. If $\alpha$ is an index which specifies the chains ($\alpha = 1$, 2, or 3), and if $L$ is the coupling parameter between two neighboring C=O oscillators on different chains, the equations of motion for the simple $|D_2\rangle$ ansatz state are [3,21]:

$$i\hbar a_{n\alpha} = -J(a_{n+1,\alpha} + a_{n-1,\alpha})$$
$$+ L(a_{n,\alpha+1} + a_{n,\alpha-1})$$
$$+ X(q_{n+1,\alpha} - q_{n\alpha})a_{n\alpha}, \qquad (3)$$

$$M\ddot{q}_{n\alpha} = W(q_{n+1,\alpha} - 2q_{n\alpha} + q_{n-1,\alpha})$$
$$+ X(|a_{n\alpha}|^2 - |a_{n-1,\alpha}|^2). \qquad (4)$$

Scott's argument is now, that the usually used values for the parameters $W$ and $M$ apply only for the case of three coupled chains. He found [3] for the interchain coupling parameter a value of $L = 1.54$ meV. For this value of $L$ and standar.' values for the other parameters ($W = 13$ N/m, $M = 114m_p$, $X = 62$ pN, and $J = 0.967$ meV) we performed simulations for a three chain system, as Scott did for one special case of excitation [3]. We used the symmetric $A$-mode ($a_{n1} = a_{n2} = a_{n3}$), the linear combination of the two degenerate $E$-modes ($a_{n1} = 0$, $a_{n2} = 1/\sqrt{2}$, $a_{n3} = -1/\sqrt{2}$) and an asymmetric local excitation

($L$) of one unit on a single chain. In the case of the $A$-mode the soliton consists of three identical, parallel moving localized excitations on all three chains, while in the $E$-mode the soliton moves only on two chains. In the case of the local excitation the soliton is found mainly on one chain, with a small fraction of the excitation transferred to the others. Scott [3,21] found for the $A$-mode that the equations of motion for three chains are identical to the equations for one chain but with $M' = 3M$ and $W' = 3W$. The important conclusion is that numerically this holds not only for the symmetric $A$-modes, for which Scott's considerations outlined above hold, but also for the $E$-mode and even for a single chain excitation [28].

In the more complicated $|D_1\rangle$ case the equations of motion for the symmetric $A$-mode for three chains and for one chain are identical as shown in [28]. We performed three chain simulations with an $A$-mode excitation and again the two single chain simulations (using $W$, $M$ and $3M$, $3W$, respectively), but now computed in the $|D_1\rangle$ model ($X = 62$ pN). We found that the one chain results for parameters $M$ and $W$ are the same as those from the three chain calculations as expected from the considerations outlined in [28]. However, an $A$-mode excitation does not necessarily occur in reality and thus we also have to study initial excitations in the $E$-mode and ones which are localized on one chain where the equations for one chain are not the same as those for three interacting chains. This is discussed in the next section.

### 3.3. Three chain dynamics at $T = 300$ K

Here we want to show results obtained with the equations for three chains with general initial excitations in the $|D_1\rangle$ case using Davydov's model for temperature effects. The equations are given in [28]. Let us first discuss a localized excitation at one of the terminal sites of just one spine. We have chosen a chain length of 20 units for each chain (excitation at site 19 of chain 1)

and the usual value of 1.5373 meV for the inter-chain coupling $L$ [21]. The time step was chosen as 0.25 fs. In this case in a typical calculation on localized initial excitations the total energy was conserved within roughly 2 μeV and the norm to better than 5 ppb (parts per billion). Repetition of one of the calculations with a time step of 0.1 fs lead to no changes in the results. In fig. 5 we show the results of these simulations. We use for $W$ the values 13 N/m from measurements on formamide crystals and 19 N/m from theoretical calculations. For $X$ we use 35 N/m which was found by Scott and also the usually applied value

of 62 pN (see [21] for references). Figures 5a and b show that for $W = 13$ N/m in case of the lower $X$ value a solitary wave is formed first which becomes trapped roughly in the middle of the chain, while for the larger $X$ value the excitation is trapped close to the chain end. In both cases a considerable part of the excitation goes over to the initially unexcited chains. In case of $W = 19$ N/m (5c, d) at the lower $X$ value a clear solitary wave is formed and travels through the chain. If one looks closer at the pulse like structure, one sees that the excitation is oscillating between the initially excited chain and the other
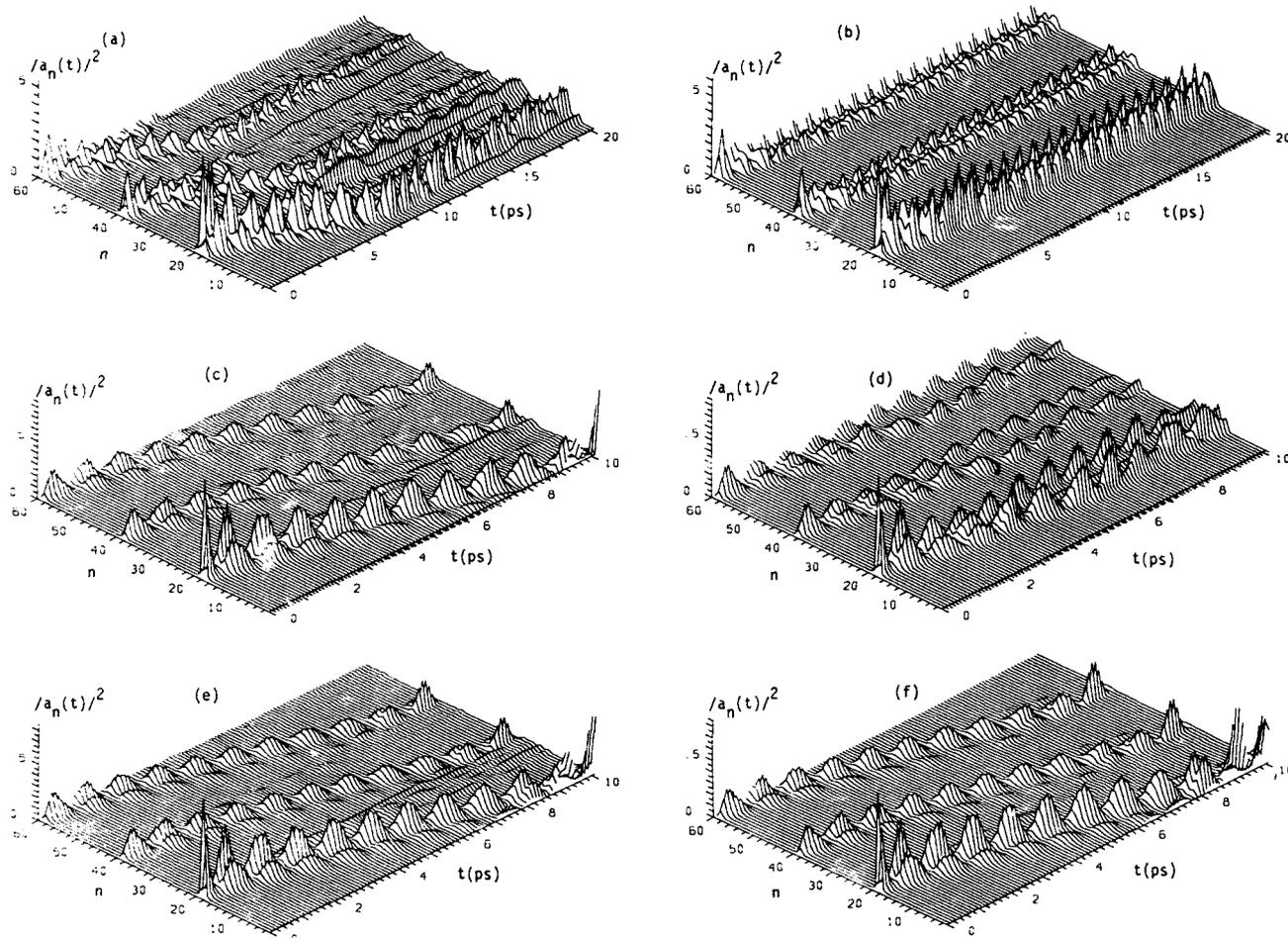


Fig. 5. Time evolution of $|a_n(t)|^2$ in the $|D_1\rangle$ ansatz state at 300 K using Davydov's model for temperature effects and a localized initial excitation $(a_{19,1}(0) = 1$, all other $a$'s equal 0.005, then normalized to 1, all three chains are shown: chain 1: $n = 1$–20, chain 2: $n = 21$–40, chain 3: $n = 41$–60, time step: 0.25 fs) for different values of $W$ and $X$:
(a) $W = 13$ N/m, $X = 35$ pN; (b) $W = 13$ N/m, $X = 62$ pN; (c) $W = 19$ N/m, $X = 35$ pN; (d) $W = 19$ N/m, $X = 62$ pN; (e) $W = 40$ N/m, $X = 62$ pN; (f) $W = 60$ N/m, $X = 62$ pN.

two. When the excitation probability is small on chain 1, it is large on chains 2 and 3 and vice versa. At the larger $X$ value the excitation is again trapped close to the initial excitation site. For $X = 62$ pN we have increased $W$ to 30 N/m, 40 N/m, 50 N/m and 60 N/m. Two cases are shown in fig. 5e (40 N/m) and f (60 N/m). As in case of one chain simulations an increase in $W$ favors soliton formation and we found soliton formation between $W = 40$ and 60 N/m, even quantitatively in fair agreement with one-chain and even with $|D_2\rangle$ results. The results for the $E$-mode are more or less similar [28]. Again from $W = 40$ N/m solitary wave formation can be expected. In case of $A$-mode excitations a strong tendency to trapping after a few sites shows up. However, for $W = 19$ N/m and $X = 35$ pN again a solitary wave shows up. In the cases of larger $W$ values we observe again solitary wave formation, but at 40 N/m after the reflection the soliton oscillates around the chain end. Increasing $W$ to 60 N/m leads again to a solitary wave, but it moves erraticly after reflection [28]. Thus we reach at the same basic conclusion as in case of most other models (with the exception of the partial dressing state) we studied: If $W$ is larger then 30–40 N/m solitary wave formation should be possible in proteins also at physiological temperature.

## 4. Conclusion

With the classical $|D_2\rangle$ ansatz state we performed dynamic simulations on one chain using different models for the incorporation of a finite temperature into the Davydov soliton theory. We varied the parameters $W$ (hydrogen bond spring constant) and $X$ (oscillator–lattice coupling constant). We found that at 300 K the spring constant of the hydrogen bonds should be larger than roughly 30–40 N/m to allow formation and propagation of Davydov solitons in the system. From calculations with the $|D_1\rangle$ ansatz state which allows for quantum effects in the

lattice and Davydov's method for incorporation of temperature, we arrived at roughly the same conclusion. Since Davydov's method to account for temperature effects is believed by several workers to be inconsistent with statistical mechanics, we performed comparisons with exact quantum Monte Carlo results (QMC) [19]. Since these results are available only for one set of parameters and with the rather unrealistic symmetric interaction ansatz, we incorporated these features into our program to be able to perform comparable calculations. Instead of the ensemble average in QMC we performed a time average over the lattice displacements in a rotated coordinate system. We found that the transition from coherent to localized structures between 2.8 K and 11.2 K reported in [19] is reproduced by Davydov's method. However, quantitatively the averaged lattice displacements and their variation with temperature are underestimated. Thus the method serves as a qualitatively correct approximation to the dynamics at physiological temperature. The classical $|D_2\rangle$ ansatz seems to overestimate temperature effects, while the partial dressing ansatz [26] leads neither to coherent structures below 7.0 K nor to localized ones above 7.0 K and also underestimates the averaged lattice displacements and their temperature variation.

For the $|D_2\rangle$ ansatz state we performed soliton dynamics in a system of three coupled chains as suggested and done earlier by Scott et al. [3]. We could verify Scott's proposal that three chain dynamics can be simulated with one chain calculations if one takes as site mass three times the value applied in the three chain dynamics and for the spring constant of the hydrogen bonds also three times the value for one bond. We obtained this result independent of the initial excitation, which can correspond to the symmetric $A$ mode or to the degenerate $E$ mode of the three chain system. Even for a localized excitation on one chain only we obtained agreement between three chain dynamics and those using one chain but changed parameters.

For the $|D_1\rangle$ state we surveyed again the $(X, Y)$ parameter space at 300 K using one chain and $3M$ as site mass for stable solitons within Davydov's method. We found that $W$ has to be larger than roughly 70 N/m to allow soliton formation around $X = 60$ pN. Since this value would correspond to a spring constant of about 25 N/m, we arrive more or less at the same conclusion as discussed above for one chain, if the suggestion, that three chain dynamics could be reproduced by one chain simulations if a site mass of $3M$ is used, is correct also in the $|D_1\rangle$ case. However, at 0 K we found that in the $|D_1\rangle$ case three chain dynamics with parameters $M$, $W$ and $X$ are identical to one chain simulations with the same parameters in contrast to $|D_2\rangle$ dynamics.

With explicit three chain simulations at $T = 300$ K using Davydov's temperature model we found again that solitary waves are formed at $W$ values larger than 30–40 N/m. However, if $X$ is around 35 pN instead of 62 pN already from $W$ around 19 N/m solitary waves were observed. Thus the conclusion drawn earlier by us remains unchanged: If the spring constant of the hydrogen bonds in protein $\alpha$-helices is larger than 30–40 N/m the Davydov soliton should be able to function at 300 K. Interestingly this conclusion is reached with both of Davydov's ansatz states and with different models for temperature effects. Since the usually quoted value of 13 N/m derives from formamide crystals where the hydrogen bonded molecules vibrate freely it should be too small for proteins. In proteins the hydrogen bonded sites are embedded in the covalent backbone of the helix which becomes distorted due to the vibration. Thus we expect the spring constant of a protein normal mode corresponding to hydrogen bond stretch to be much larger than that of crystalline formamide, and thus probably allowing for Davydov solitons to be formed in proteins. However, detailed calculations or measurements on the spring constant in proteins are necessary to decide finally on the question of existence of Davydov solitons.

## Acknowledgement

## References

[1] A.S. Davydov and N.I. Kislukha, Phys. Stat. Sol. B 59 (1973) 465;
A.S. Davydov, Phys. Scr. 20 (1979) 387.

[2] A.S. Davydov, Zh. Eksp. Teor. Fiz. 78 (1980) 789; Sov. Phys. JETP 51 (1980) 397.

[3] A.C. Scott, Phys. Rev. A 26 (1982) 57; Physica Scr. 29 (1984) 279;
L. MacNeil and A.C. Scott, Phys. Scr. 29 (1984) 284;
Phil. Trans. R. Soc. London A 315 (1985) 423.

[4] W. Förner and J. Ladik, in: Davydov's Soliton Revisited, eds. P.L. Christiansen and A.C. Scott, NATO ASI, Series B – Physics, Vol. 243 (Plenum, New York, 1991).

[5] J. Halding and P.S. Lomdahl, Phys. Lett. A 124 (1987) 37.

[6] (a) P.S. Lomdahl and W.C. Kerr, Phys. Rev. Lett. 55 (1985) 1235; in: Davydov's Soliton Revisited, eds. P.L. Christiansen and A.C. Scott, NATO ASI, Series B – Physics, Vol. 243 (Plenum, New York, 1991);
(b) W.C. Kerr and P.S. Lomdahl, Phys. Rev. B 35 (1987) 3629;
(c) W.C. Kerr and P.S. Lomdahl, in: Davydov's Soliton Revisited, eds. P.L. Christiansen and A.C. Scott, NATO ASI, Series B – Physics, Vol. 243 (Plenum, New York, 1991).

[7] A.F. Lawrence, J.C. McDaniel, D.B. Chang, B.M. Pierce and R.B. Birge, Phys. Rev. A 33 (1986) 1188.

[8] H. Bolterauer, in: Structure Coherence and Chaos, Proc. MIDIT 1986 Workshop (Manchester University Press).

[9] J.P. Cottingham and J.W. Schweitzer, Phys. Rev. Lett. 62 (1989) 1792.

[10] H. Motschmann, W. Förner and J. Ladik, J. Phys.: Cond. Matter 1 (1989) 5083.

[11] (a) W. Förner, J. Phys.: Cond. Matter 3 (1991) 4333.
(b) W. Förner, J. Comput. Chem. 13 (1992) 275.

[12] D.W. Brown, K. Lindenberg and B.J. West, Phys. Rev.

A 33 (1986) 4104, 4110; Phys. Rev. B 35 (1987) 6169; B 37 (1988) 2946;

D.W. Brown, Phys. Rev. A 37 (1988) 5010.

[13] L. Cruzeiro, J. Halding, P.L. Christiansen, O. Skovgaard and A.C. Scott, Phys. Rev. A 37 (1988) 880.

[14] B. Mechtly and P.B. Shaw, Phys. Rev. B 38 (1988) 3075.

[15] M.J. Skrinjar, D.V. Kapor and S.D. Stojanovic, Phys. Rev. A 38 (1988) 6402.

[16] W. Förner, Phys. Rev. A 44 (1991) 2694.

[17] W. Förner, Nanobiology 1 (1992) 413.

[18] W. Förner, J. Phys.: Cond. Matter 4 (1992) 1915.

[19] X. Wang, D.W. Brown and K. Lindenberg, Phys. Rev. Lett. 62 (1989) 1796.

[20] A.C. Scott, presented at Conf. on Nonlinear Sciences: The Next Decade, Los Alamos, National Laboratory (May 1990).

[21] A.C. Scott, Phys. Rep. 217 (1992) 1.

[22] A.S. Davydov, Phys. Scr. 20 (1979) 387.

[23] J.W. Schweitzer and J.P. Cottingham, in: Davydov's Soliton Revisited, eds. P.L. Christiansen and A.C. Scott, NATO ASI, Series B – Physics, Vol. 243 (Plenum, New York, 1991).

[24] J.P. Cottingham and J.W. Schweitzer, unpublished.

[25] H. Bolterauer, in: Davydov's Soliton Revisited, eds. P.L. Christiansen and A.C. Scott, NATO ASI, Series B – Physics, Vol. 243 (Plenum, New York, 1991).

[26] D.W. Brown and Z. Ivic, Phys. Rev. B 40 (1989) 9876.

[27] W. Förner, J. Phys.: Condensed Matter 5 (1993) 803.

[28] W. Förner, J. Phys.: Condensed Matter 5 (1993) 823.

PHYSICA ⑩

# Thermodynamics of Toda lattice models: application to DNA

R.K. Bullough[a], Yu-zhong Chen[b] and J.T. Timonen[c]

[a]*UMIST, P.O. Box 88, Manchester M60 1QD Manchester, UK*
[b]*Purdue University, Indianapolis, IN 46202, USA*
[c]*Department of Physics, University of Jyväskylä, SF-40351 Jyväskylä, Finland*

Our generalised Bethe ansatz method is used to formulate the statistical mechanics of the classical Toda lattice in terms of a set of coupled integral equations expressed in terms of appropriate action–angle variables. The phase space as coordinatised by these action–angle variables is constrained; and both the soliton number density and the soliton contribution to the free energy density can be shown to decouple from the phonon degrees of freedom and to depend only on soliton–soliton interactions. This makes it possible to evaluate the temperature dependence of the soliton number density which, to leading order, is found to be proportional to $T^{1/3}$.

## 1. Introduction

This paper is concerned with Alwyn Scott's contributions to molecular biology – particularly with his recent work [1,2] on the thermal excitations of the Toda lattice [3,4] viewed as a model of the DNA *molecule at biological* (i.e. *body*) *temperatures* $\approx 310$ K. It connects in a natural way with his paper to this meeting [5] 'There's more than one way to skin Schrödinger's cat' because even *classical* statistical mechanics is intimately related to quantum mechanics (recall [6] the Gibbs' paradox e.g.). Thus just as one can rewrite the title of Alwyn Scott's (ACS's) paper [5] in another language, Danish, as 'Man kan flå Schrödinger's kat på mere end en facon' (the title actually reported during the meeting in Copenhagen's newspaper 'Berlinske Tidende') and one can also rewrite this in many different ways, so in quantum mechanics there are many technically different lines of argumentation using the many different representations.

Thus for the many quantum integrable lattice models [7] which have a commuting number operator $\hat{N}$, $[\hat{H}, \hat{N}] = 0$ and $\hat{H}$ is the Hamiltonian, there is the standard number state method (SNSM) [5,7] and there is the quantum inverse scattering method (QISM) [5,7] equivalent to the Bethe ansatz (BA) method as conceived by Bethe [8]. The QISM as well as the BA can also be developed for field theories, while there is the Feynman propagator method for both integrable and non-integrable field theories and lattices in which $K(\phi, \phi_0; T)$ ($t$ is time and $0 \leq t \leq T$) is a Green's function expressible in terms of the classical action $S(\phi, \phi_0; t)$ as a path (or functional) integral on a Hilbert space [9]: $K(\phi, \phi_0; T)$ can be expressed (with advantage [10]) as a functional integral on a symplectic manifold M (a phase space) infinite dimensional in the case of field theories [10]. Wick rotation on the time, so that $t \rightarrow -\mathrm{i}t$, $T \rightarrow -\mathrm{i}\beta$ yields the partition function $Z(\beta) = \mathrm{Tr}\, K(\phi, \phi_0; -\mathrm{i}\beta) = \int \mathcal{D}\phi_0\, K(\phi_0, \phi_0; -\mathrm{i}\beta)$ where $\mathcal{D}$ means functional integration [10] and $\beta^{-1} = k_\mathrm{B}T$ with $T$ now the temperature. The free energy $F = -\beta^{-1} \ln Z$, and $\beta^{-1} = k_\mathrm{B}T \rightarrow 0$ gives the ground state energy in thermodynamic limit [7,11]. Alternatively and equivalently $F = E - \beta^{-1}S$ where $E$ is an energy defined for the model at temperature $\beta^{-1}$ and $S$ is an entropy. $F$ is then minimised under the constraint of the algebraic Bethe ansatz described by ACS [5]. Yang and Yang [12] introduced this particular method for the quantum SM in the case of the integrable quantum repulsive nonlinear Schrödinger model (the 'bose gas'). Both

the 'bose gas' and the discrete lattice bose gas can be handled this way [7]. With $[\hat{H}, \hat{N}] = 0$, $E - \beta^{-1}S - \bar{\mu}N = -pL$, where $L$ is the 'volume' in one dimension, $\bar{\mu}$ is the chemical potential, and $p$ is the pressure. All of these methods for the quantum integrable systems rest on algebraic methods and most especially on the quantum groups, non-commutative and non-cocommutative for the quantum integrable models, commutative and cocommutative for the classical integrable models [7,13,14]. Fig. 1 in ref. [14] shows some of these algebraic connections: the classical inverse method (top left) and the QISM (third column down from the right) are shown, while our work, e.g., refs. [15–22], on the quantum and classical SM of the integrable models is summarised by the route from the Lax pair of the inverse method to periodic boundary conditions (b.c.s.) across the top and then down to the partition function $Z$ given by eq. (4) below and out to EXPERIMENTS.

The functional integral form of quantum $Z$ we have used is defined on the symplectic manifold M as

$$Z = \text{Tr} \int \mathscr{D}\Pi \, \mathscr{D}\phi \, \exp S[\Pi, \phi] \,, \quad \{\Pi, \phi\} = \delta(x - x') \,. \tag{1}$$

The classical action $S$ (for quantum $Z$) is [10]

$$S[\Pi, \phi] = \hbar^{-1} \int_0^{\beta\hbar} d\tau \left( i \int \Pi\phi_\tau \, dx - H[\Pi, \phi] \right) \,, \tag{2}$$

where $i = \sqrt{-1}$ and $H[\Pi, \phi]$ is the classical Hamiltonian. The classical limit of this quantum $Z$ is the classical $Z$ given by (1) with

$$S[\Pi, \phi] = -\beta H[\Pi, \phi] \,. \tag{3}$$

This limit must be taken so that $\hbar \to 0$ with zero winding number [10].

Natural *coordinates* for the integrable models are the action–angle variables (since $S$ is classical for both quantum and classical $Z$ we can use the action–angle variables as the natural coordinates on M in both cases). Under this choice, classical $Z$, eq. (1) with (3) becomes

$$Z = \text{Tr} \int \mathscr{D}\mu \, \exp -\beta H[p] \,, \tag{4}$$

where $H[p]$ depends only on the action variables and $\mathscr{D}\mu$ is a measure for functional integration on M to be determined [10,15–22]. These action–angle variable coordinates enable us to *define* a soliton number $N_s$ and its mean value $\langle N_s \rangle$ (note that in terms of canonical coordinates $\Pi$, $\phi$ there is no means of identifying soliton or other particular contributions (the phonons – see below) to $Z$).

The calculation of $Z$ simplifies to some extent for *lattices* (i.e. for integrable lattices, quantum or classical) where the functional integral for *classical* $Z$ eq. (1) with the action (3), becomes a product of ordinary integrals. Generically the action–angle variables, under vanishing b.c.s at $\pm\infty$, split into variables $P(k)$, $Q(k)$ labelled by modes $k$, together with 'soliton coordinates' $p_i$, $q_i$ ($i = 1, 2, \ldots, N_s$). Our method of calculation exploits the fact that in thermodynamic limit there is a connection between action-angle variables under these b.c.s. and action–angle variables under periodic b.c.s. Under periodic b.c.s. the integrable lattice has a total number $N$ of degrees of freedom (the total number of lattice points in the period). The thermodynamic limit is then taken so that e.g. $\lim_{N \to \infty} N_s N^{-1}$ is a non-zero soliton number density.

In the action–angle variable description the number of modes $N_{\text{modes}}$ (say) of the modes $k_i$ and the number of solitons $N_s$ satisfy $N_s + N_{\text{modes}} = N$ for fixed $N$. The limit $N \to \infty$ is then taken at finite density as described. For fixed $N \leq \infty$ one finds for the classical integrable models that generically

$$H[p] = \sum_{i=1}^{N_{\text{modes}}} \omega(k_i) P(k_i) + \sum_{j=1}^{N_s} H[p_j] \, . \tag{5}$$

The mode sum is apparently the linearised part and $\omega(k_i)$ is the linearised dispersion relation. However, the mode sum also provides *nonlinear* contributions to classical $Z$, eq. (4) – as is exemplified in the calculations [16] for the sinh-Gordon model. This is because the $k_i$ couple to the $P_i$ in our *classical generalization* of the Bethe equations as we explain for the Toda lattice below. We call this contribution to (5) the 'phonon contribution' in this paper. The remaining part of (5) provides the soliton contribution to $Z$.

In these terms the classical lattice partition function $Z$ and mean soliton number $\langle N_s \rangle$ are given by

$$\langle N_s \rangle = Z^{-1} \left( \sum_{N_s=1}^{N} N_s (N_s!)^{-1} \int \prod_{i=1}^{N_{\text{modes}}} dP(k_i) \prod_{j=1}^{N_s} dp_j \exp - \beta H[p] \right) ,$$

$$Z = \sum_{N_s=1}^{N} (N_s!)^{-1} \int \prod_{i=1}^{N_{\text{modes}}} dP(k_i) \prod_{j=1}^{N_s} dp_j \exp - \beta H[p] \, , \tag{6}$$

with $N_s + N_{\text{modes}} = N$. The $(N_s!)^{-1}$ is due to Willard Gibbs [6]. In thermodynamic limit $N_s, N \to \infty$ so that $N_s N^{-1} > 0$. We can execute this limit best by going to a 'number representation' in terms of densities $\rho_s(p)$ for the solitons ($\int \rho_s(p) \, dp = \lim N_s N^{-1}$) and defining a 'phonon' density $\rho(k)$. Total particle density is then $\int \rho(k) \, dk + \int \rho_s(p) \, dp$, and we can then work with variable numbers of particles by introducing chemical potentials $\mu_{\text{ph}}$ and $\mu_s$ (say). The problem posed and largely solved in this paper is to carry out this calculation for the classical Toda lattice and most particularly to find the mean soliton number density $\langle N_s \rangle N^{-1}$.

This problem is of particular interest for at least two reasons. Most importantly for this meeting, in refs. [1,2] V. Muto, A.C. Scott and P.L. Christiansen calculate this density numerically for finite but quite large values of $N$ and they find unequivocally that $\langle N_s \rangle N^{-1} \propto T^{1/3}$ where $T$ is the temperature. A second reason is that we had believed that our methods for $Z$ just sketched could be applied to *all* of the integrable models in $1+1$ dimensions, lattices and field theories both quantum and classical. Our methods have been the functional integral methods sketched, or our method of 'generalised Bethe ansatz' [15,16,19] which apparently extends the Yang and Yang BA method [12] to the quantum *and classical* SM of all of the integrable models in $1+1$ and finds in many cases alternative descriptions in terms of fermi, bose, or other statistics [10] for each model. We reported solutions for $Z$ and $F$ for some 14 integrable models in ref. [19] and it was therefore surprising to discover [20] that the classical Toda lattice did not seem to fit within these schemes. Particularly $\langle N_s \rangle N^{-1} \propto T^{1/3}$ seemed impossible given the expressions for $Z$ we were able to find [20]. There has been a considerable variation in results already obtained for the Toda lattice: ref. [2] quotes an unpublished $T^{1/3}$ found by Schneider and Stoll and an error in this calculation found by Bolterauer and Opper [23], as well as $\langle N_s \rangle N^{-1} = (\ln 2/\pi^2)\tilde{T}$, $\tilde{T}$ a normalised temperature by Mertens and Büttner [24,25]. The paper most relevant to ours is ref. [26] which finds $T^{1/3}$. This paper used the classical action–angle variables but uses these inconsistently. We refer in more detail to this paper ref. [26] elsewhere.

## 2. Toda lattice

In our paper we sketch what appears to be a correct *ab initio* non-phenomenological analytic theory for the quantity $\lim_{N \to \infty} \langle N_s \rangle N^{-1}$ and show that the $T^{1/3}$ behaviour found numerically [1,2] can still be

accommodated in our expression for the free energy density, $\lim FN^{-1}$ which also shows *no* non-analytical behaviour in $T$. Incidentally it is a particular pleasure to work once again on the Toda lattice because Professor Morikazu Toda was present at this meeting and himself gave a paper on it [4].

In refs. [1,2] two numerical method were used: one of these was to create Gaussian chaotic Brownian motion at a chosen temperature on a lattice of $N$ points under periodic b.c.s. $T$ was identified by setting the mean kinetic energy per particle $E_{kin} = \frac{1}{2} k_B T$. In fact remarkably similar results were found [2] for a number of different initial distributions of energy between the kinetic and potential energy parts of the total energy. For canonical $P_n, Q_m$ the general results of classical SM for both linear *and* nonlinear systems are, of course, the 'virials' $\langle P_n \partial H / \partial P_n \rangle = \langle P_n \dot{Q}_n \rangle = \beta^{-1}$; $\langle Q_n \partial H / \partial Q_n \rangle = -\langle \dot{P}_n Q_n \rangle = \beta^{-1}$. The number of solitons was *defined* and counted by an ingenious 'spectral analyser' method which counts the number of eigenvalues $\lambda$ of an associated spectral problem lying in $\lambda > +1$ or $\lambda < -1$. Our definition of soliton number is through the action–angle variables and is therefore essentially identical to this definition of refs. [1,2]. Thus we should be in complete agreement. (Note that the *definition* of a soliton in equilibrium or non-equilibrium SM is not necessarily unique – since (a) periodic b.c.s. are used and (b) even under vanishing b.c.s at $\infty$ the soliton in the equilibrium SM (thermal equilibrium) is not the 'bare' soliton calculated through the classical inverse method). The numerical results found by this first method were unequivocally $\langle N_s \rangle N^{-1} \propto T^{1/3}$ [1,2,27]. The second method was [2,27] to drive the Toda chain by a random force $\eta_n(t)$ at each lattice site $n$. Again $\langle N_s \rangle N^{-1} \propto T^{1/3}$ even for $N = 32$ (cf. fig. 4 of ref. [2]). The spectral analyser was again used to count the solitons.

It is possible to calculate the *partition function* $Z$ for the Toda lattice exactly. Toda gave $Z$ for *one* fixed end of the chain, i.e., at constant pressure, in ref. [3]. We gave a closed form for $Z$ at constant volume (two fixed ends) in ref. [20] as $Z = Z_{kin} Z_{pot}$: $Z_{kin} = (2\pi b a^{-1} \beta)^{-N}$ in which $a, b$ are the usual Toda parameters and $Z_{pot}$ was

$$Z_{pot} = e^{(2N+1)ab^{-1}\beta} \left\{ \prod_{i=1}^{2N} \int_0^{\infty} \frac{dy_i}{y_i} \exp\left[ -ab^{-1}\beta \left( \sum_{j=1}^{2N} y_j + \sum_{j=1}^{2N} y_j^{-1} \right) \right] \right\}. \tag{7}$$

There are now $2N + 1$ points in a period under periodic b.c.s and ($\hbar = 1$)

$$Z = (2\pi)^{-2N} \int \prod_{n=-N+1}^{N} dP_n \, dQ_n \, e^{-\beta H[P_n, Q_n]},$$

where

$$H[P_n, Q_n] = ab^{-1} \sum_{n=-N}^{N} (\frac{1}{2} b^2 a^{-2} P_n^2 + e^{-(Q_n - Q_{n-1})} - 1) \tag{8}$$

is a Hamiltonian for a scaled Toda lattice: $\{P_n, Q_m\} = \delta_{nm}$. The $Q_n$ are dimensionless and $ba^{-1}P_n$ is dimensionless; $ab^{-1}$ is an energy and $t$ is scaled as $\sqrt{ab} M^{-1} t \rightarrow t$. The closed form (7) is exact and has a low temperature asymptotic expansion in the dimensionless temperature $ba^{-1}\beta^{-1}$ of the form [20]

$$\lim_{N \rightarrow \infty} F(2N + 1)^{-1} \sim \beta^{-1} \ln \beta - \beta^{-1} \ln[1 + \mathcal{O}(ba^{-1}\beta^{-1})]$$

$$+ \mathcal{O}(\ln(2N)/(2N + 1)) + \mathcal{O}(1/(2N + 1)), \tag{9}$$

in which the $[1 + \mathcal{O}(ba^{-1}\beta^{-1})]$ is an asymptotic power series. The $\beta^{-1} \ln \beta$ is the linear part of the phonon contribution and the $\beta^{-1} \ln[\cdots]$ is the nonlinear contribution provided by both solitons and phonons. The error $\mathcal{O}(1/(2N + 1))$ is usual for the thermodynamic limit $2N + 1 \rightarrow \infty$. The $\mathcal{O}(\ln(2N)/$

$(2N + 1)$) is *not* usual and conceals a divergent ($\sim \ln N$) phonon number density. However the free energy (9) is well defined.

It is worthwhile comparing (9) with a corresponding low $T$ expansion for the integrable classical sine-Gordon (sG) model [15,17,19]. One can calculate the sG soliton density to be $m[8/\pi t]^{1/2} e^{-1/t} +$ corrections $\mathcal{O}(t)$; $t = M\beta^{-1}$, $m$ is the sG phonon mass and $M = 8m\gamma_0^{-1}$ is the soliton mass ($\gamma_0$ is the coupling constant). This is $-\beta$ times the leading term in the soliton contribution to the free energy density. On this line of reasoning we see that, since $\ln[1 + \mathcal{O}(ba^{-1}\beta)]$ in (9) is itself a simple-power series in $ba^{-1}\beta$, a behaviour for $\lim_{N\to\infty}\langle N_s\rangle/(2N + 1)$ nonanalytic in $\beta^{-1}$ and proportional to $\beta^{-1/3}$ cannot apparently be accommodated. Despite this situation the main result reported in this paper is, nevertheless, that 'to first order',

$$\lim_{N\to\infty} \langle N_s\rangle/(2N + 1) = \frac{1}{\pi}\Gamma(\tfrac{4}{3})\left(\frac{T}{T_0}\right)^{1/3}, \quad T_0 \equiv \frac{4a}{3bk_B}. \tag{10}$$

The result (10) has also been given by F. Marchesoni and G. Lucheroni [28]. Unfortunately we are unable (*so far*) to justify any part of the argument these authors give to reach this result though ref. [26] may be suggestive. The analysis of ref. [28] is based on phonons and is modelled in effect on the sG analysis [15]. We show here next that the thermodynamics of the Toda lattice *is very different* from that of sG and that phonons play no role in the calculation leading to (10). Fig. 1 in ref. [28] shows how well formula (10) corresponds to the data of fig. 4 in ref. [2] (for which $2N + 1 \approx 32$). The line through the actual data in this figure 1 of ref. [28] was reached by considering additional corrections due to soliton–soliton interactions otherwise totally neglected. Note that our result (10) is in thermodynamic limit where $2N + 1 \to \infty$: in ref. [2] it seems that the factor $\pi^{-1}\Gamma(\tfrac{4}{3})(T_0)^{-1/3}$ changes so as to move a line through the data on the $\ln(\langle N_s\rangle/(2N + 1)$ against $\ln T$ plot to the left as $2N + 1$ increases from $\approx 32$ to $\approx 128$ whereas (10) already differs from the data in fig. 4 of ref. [2] by a missing factor of $\approx 1.5$, pushing the line to the left. On the other hand the first of the two methods which has $2N + 1 \approx 64\,000$ is much closer to (10). For the exact comparison between the numerical results and our analytic methods more numerical work like that of ref. [2] is now needed.

In ref. [20] we give more detail of the calculation of $\lim_{N\to\infty} \langle N_s\rangle/(2N + 1)$ for sG: we gave explicit integral equations whose iteration yields

$$\rho_s(p) \sim (2\pi)^{-1} e^{-\beta E_s(p)}, \tag{11}$$

where $E_s(p)$ is the excitation energy of solitons at the temperature $\beta^{-1} = k_B T$ and is found from the classical Yang and Yang type integral equations given in refs. [15,17,19]. From (11) we can see that the soliton contribution to the free energy density of sG is $-2\beta^{-1}\int_{-x}^{x}\rho_s(p)\,dp \equiv -2\beta^{-1}\ln\langle N_s\rangle/(2N + 1) = m[8/\pi t]^{1/2} e^{-1/t} + \mathcal{O}(t)$ as quoted before. The same connection between soliton number density and free energy contribution does obtain for the Toda lattice – providing we push all difficulty in calculating the resultant total free energy density into the remaining phonon contributions. We noted already that there is no definable phonon number density for the unconstrained linearised Toda lattice in thermodynamic limit. We explain the phrase 'unconstrained' in the following.

## 3. Action–angle variables

The key to the situation is that the Toda lattice has a constrained phase space in terms of its action–angle variables under vanishing b.c.s at infinity. (We have not yet checked however that we can exploit the connection between action–angle variables under periodic b.c.s and those under vanishing

b.c.s in thermodynamic limit in this case of the Toda lattice. Nor have we checked that our result (10) for the soliton density leads finally to the free energy density given by (7) and (9). Our calculations which lead to (10) are therefore incomplete in these respects.)

The essential of the matter under vanishing b.c.s is given in ref. [29]: solitons and phonons are coupled in terms of their *realisable* 'action–angle variables'. The Hamiltonian of the Toda lattice in terms of these action–angle variables can be put in the form [29]

$$H = 2 \int_0^\pi \sin k \, P(k) \, \mathrm{d}k + \tfrac{1}{2} \sum_{j=1}^{N_s} (z_j^{-2} - z_j^2 + 2 \ln z_j^2) \,, \tag{12}$$

but the variables $P(k)$ and $z_j$ are constrained by

$$\int_0^\pi \frac{P(k)}{\sin k} \, \mathrm{d}k = -\sum_{j=1}^{N_s} \ln z_j^2 \,. \tag{13}$$

Thus there is the alternative form for $H$ in these variables which is

$$\tilde{H} = -2 \int_0^\pi \frac{\cos 2k}{\sin k} \, P(k) \, \mathrm{d}k + \tfrac{1}{2} \sum_{j=1}^{N_s} (z_j^{-2} - z_j^2) \,. \tag{14}$$

The $z_j$ are the zeros of the 'transmission coefficient' $a(z)$ inside the unit circle $|z| = 1$ and describe the solitons: coefficients $a(z)$, $b(z)$ are defined on $|z| = 1$, $z \neq \pm 1$ so that $|a(z)|^2 - |b(z)|^2 = 1$ [29]. These coefficients near $z = \pm 1$ are singular or regular together [29] so that $b(z)$ and $z_j$ are not independent. This is tied to the fact that one half of the infinite set of constants of the motion are inadmissible since they do not conserve the phase space. Consequently the 'natural' soliton action–angle variables

$$\tilde{p}_j = z_j + z_j^{-1} \,, \qquad \tilde{q}_j = \ln|\gamma_j| \,, \qquad j = 1, \dots, N_s \tag{15}$$

(for the definition of $\gamma_j$ see ref. [29]) together with the natural action–angle variables for phonons

$$P(k) = \frac{1}{\pi} \sin k \, \ln(1 + |b(\mathrm{e}^{\mathrm{i}k})|^2) \,,$$
$$Q(k) = -\arg b(\mathrm{e}^{\mathrm{i}k}) \,, \qquad 0 < k < \pi \,, \tag{16}$$

have [29] the complex Poisson brackets

$$\{P(k), Q(k')\} = \delta(k - k') - \frac{\sin k}{\sin k'} [\delta(k) - \delta(k - \pi)] \,,$$
$$\{P(k), \tilde{q}_j\} = -2 \sin k \frac{z_j}{z_j^2 - 1} [\delta(k) - \delta(k - \pi)] \,,$$
$$\{\tilde{p}_i, \tilde{q}_j\} = \delta_{ij} \,, \tag{17}$$

while all other pairings commute. Evidently $P$, $Q$ are not canonical action–angle variables and $P(k)$ and $\tilde{q}_j$ are coupled.

With these variables as 'best possible' action–angle variables under vanishing b.c.s, a situation we are investigating further, one can make the canonical transformation $z_j = \exp(-p_j)$, $0 < p_j < \infty$. Then $\tilde{p}_j = 2 \cosh p_j$. Also $\tilde{q}_j = (\sinh p_j)^{-1} q_j$ preserves the symplectic form and

$$H = 2 \int_0^{\pi} \sin k \, P(k) \, dk + \sum_{j=1}^{N_s} (\sinh 2p_j - 2p_j) \, . \tag{18}$$

This splits as in (5) into phonon and soliton parts. It is important to our result (10) that $\sinh 2p_j - 2p_j$ expands as $\frac{4}{3} p_j^3 + \mathcal{O}(p_j^5)$: $\tilde{H}$ eq. (14) expands as $2p_j + \mathcal{O}(p_j^3)$.

## 4. Statistical mechanics and soliton density

The constraint (13) on the phase space coordinatised through these action–angle variables $P(k)$, $Q(k)$, $p_j$, $q_j$ need not be implemented for the calculation of $\lim_{N \to \infty} \langle N_s \rangle / (2N + 1)$ if we use the soliton coordinates $p_j$, $q_j$ as free coordinates. The difficulty is thus pushed into the calculation of the phonon number density (without constraints apparently undefined in thermodynamic limit – see by eq. (9)) and the contribution of the phonons to the total free energy density which *is* defined by (9). This procedure can work because, with this choice of variables and the choice (18) for $H$, *both the soliton number density and the soliton contribution to the free energy density, which to leading order* $= -\beta^{-1} \cdot$ (*number density*) *depend only on the soliton–soliton interactions.*

One can see this by looking at the classical generalised BA, namely the *classical* Bethe equations for the Toda lattice. These we find to be [21] as follows: Recall that, generically the quantum Bethe equations take the form

$$e^{i\tilde{k}_i(2N+1)} = \prod_{\substack{j \neq i}}^{N_0} e^{-i\Delta(\tilde{k}_i, \tilde{k}_j)} \tag{19}$$

so their logarithm is

$$(2N + 1)\tilde{k}_i = 2\pi n_i - \sum_{\substack{j \neq i}}^{N_0} \Delta(\tilde{k}_i, \tilde{k}_j) \, , \tag{20}$$

where the $n_i$ are integers $1, 2, \ldots, 2N + 1$, and the $\Delta$ are quantum $S$-matrix phase shifts in a fermion description. We need $N_0, 2N + 1 \to \infty$ in thermodynamic limit with $N_0(2N + 1)^{-1}$ finite, for a finite density thermodynamic limit. Our result [21] for the 'classical' Bethe equations in corresponding logarithmic form is the set of 'phase shift equations'

$$(2N + 1)\tilde{k}_i = 2\pi n_i - \mathcal{P} \int_0^{\pi} dk' \, \Delta(\tilde{k}_i, k') \, P(k')$$

$$+ \sum_{j=1}^{N_s^+} \Delta_s^+(\tilde{k}_i, p_j^+) + \sum_{j=1}^{N_s^-} \Delta_s^-(\tilde{k}_i, p_j^-) \, , \tag{21a}$$

$$(2N + 1)p_j^+ = 2\pi n_j^+ - \sum_{j'=1}^{N_s^+} \Delta_{ss}^{++}(p_j^+, p_{j'}^+) - \sum_{j'=1}^{N_s^-} \Delta_{ss}^{+-}(p_j^+, p_{j'}^-) \, , \tag{21b}$$

$$(2N + 1)p_j^- = 2\pi n_j^- - \sum_{j'=1}^{N_s^+} \Delta_{ss}^{-+}(p_j^-, p_{j'}^+) - \sum_{j'=1}^{N_s^-} \Delta_{ss}^{--}(p_j^-, p_{j'}^-) \, , \tag{21c}$$

with $j' = j$ excluded in (21b,c): $n_i$ and $n_j^\pm$ are integers and $\mathscr{P}$ means principal value. These two equations concern at their left sides positive ($+ve$) and negative ($-ve$) momentum solitons respectively. There is coupling of the two sorts of soliton to the phonons through the phase shifts $\Delta_s^\pm(k, p)$ in (21a). However, the soliton equations (21b,c) are closed to soliton–soliton interactions.

By our generalised BA method [15,16,18,19,21], this set (21) of 'phase shift equations' yields for the free energy density

$$\lim_{N \to \infty} \frac{F}{2N + 1} = \frac{1}{2\pi\beta} \int_0^\pi dk \, \ln[\beta\epsilon(k)] - \frac{1}{2\pi\beta} \int_0^\infty dp \, e^{-\beta E_s^+(p)} - \frac{1}{2\pi\beta} \int_0^\infty dp \, e^{-\beta E_s^-(p)} . \tag{22}$$

The phonon excitation energies $\epsilon(k)$ are found from a nonlinear integral equation coupling phonons and both sorts of soliton. But the corresponding integral equations for $E_s^\pm(p)$ are

$$E_s^+(p) = E_s(p) - \frac{1}{2\pi\beta} \int_0^\infty dp' \, \frac{\partial \Delta_{ss}^{++}(p, p')}{\partial p} e^{-\beta E_s^+(p')}$$

$$- \frac{1}{2\pi\beta} \int_0^\infty dp' \, \frac{\partial \Delta_{ss}^{+-}(p, p')}{\partial p} e^{-\beta E_s^-(p')} , \tag{23}$$

with the obvious corresponding equation for $E_s^-(p)$. These two coupled integral equations are closed to $E_s^\pm(p)$.

In (23) $E_s(p) = \sinh 2p - 2p$ and the phase shifts, which appear first in (21), are [21]

$$\Delta(k, k') = \{[\sin(\tfrac{1}{2}k) \cos(\tfrac{1}{2}k')]^2 - [\cos(\tfrac{1}{2}k) \sin(\tfrac{1}{2}k')]^2\}^{-1} ,$$

$$\Delta_s^+(k, p) = -2 \tan^{-1}[\tan(\tfrac{1}{2}k) \tanh(\tfrac{1}{2}p)] ,$$

$$\Delta_s^-(k, p) = 2 \tan^{-1}[\cot(\tfrac{1}{2}k) \tanh(\tfrac{1}{2}p)] ,$$

$$\Delta_{ss}^{+-}(p, p') = 2 \tanh^{-1}[\coth(\tfrac{1}{2}p) \tanh(\tfrac{1}{2}p')] = \Delta_{ss}^{-+}(p, p') ,$$

$$\Delta_{ss}^{++}(p, p') = -2 \tanh^{-1}[\tanh(\tfrac{1}{2}p) \tanh(\tfrac{1}{2}p')] = \Delta_{ss}^{--}(p, p') . \tag{24}$$

The two soliton number densities are then given by $\lim_{N \to \infty} \langle N_s^\pm \rangle / (2N + 1) = \int_0^\infty \rho_s^\pm(p) \, dp$ where $\rho_s^\pm(p)$ satisfy

$$\rho_s^\pm(p) \, e^{\beta E_s^\pm(p)} = \frac{1}{2\pi} - \frac{1}{2\pi} \int_0^\infty dp' \, \frac{\partial \Delta_{ss}^{\pm\pm}(p, p')}{\partial p} \rho_s^\pm(p')$$

$$- \frac{1}{2\pi} \int_0^\infty dp' \, \frac{\partial \Delta_{ss}^{\pm\mp}(p, p')}{\partial p} \rho_s^\mp(p') , \tag{25}$$

and this system is again closed to solitons. Note that $E_s^\pm(p)$ solve (23) and its companion. This solution is obtained by iteration (assuming iteration is possible). Then (25) must be solved by iteration (assuming *that* is possible). So there are *two* iterations. In this way we reach the formula (10) as

$$\lim_{N \to \infty} \frac{\langle N_s^+ + N_s^- \rangle}{2N + 1} = \int_0^\infty \rho_s(p) \, dp = \frac{1}{\pi} \Gamma(\tfrac{4}{3}) \left(\frac{T}{T_0}\right)^{1/3} - \mathcal{O}(T/T_0) . \tag{26}$$

We have managed to evaluate all but one of the contributions to the first correction to (26) which is then $(1/48\pi\sqrt{3})\beta^{-1}$ ($\beta$ measured in $(a/b)^{-1}$). Note that the contribution to the total free energy density of the constrained phonon modes must *eliminate* the free energy contribution

$$\lim_{N\to\infty} \frac{F_s}{2N+1} = -\beta^{-1} \lim_{N\to\infty} \frac{\langle N_s^+ + N_s^- \rangle}{2N+1} \tag{27}$$

determined by (26) since this goes as $T^{4/3}$. The result for the total free energy must then be equivalent to the asymptotic expansion (9). The problem of actually demonstrating all of this looks very difficult.

It would be very interesting now to make detailed comparisons between numerical estimates for the soliton density and the analytical theory as we develop the further corrections to (26). These corrections already tend to explain the turn-over of the numerical log–log plots at higher temperatures but, as noted, the numerical calculations found by the second method in ref. [2] are very far from the thermodynamic limit used in our calculations. Of course it will also be instructive to calculate the free energies numerically by an adaptation of the second method of ref. [2] if this can be devised.

*Conclusion:* The formula (10) for the soliton number density of the classical Toda lattice in thermodynamic limit seems to be a good estimate for small enough temperatures $T$ in the range for which classical statistical mechanics applies. It is not incompatible, apparently, with the free energy result at constant volume and this result is in agreement with Toda's free energy [3] taken at constant pressure.

We have still to check that our use of the constrained 'action–angle variables' (15) and (16) for the thermodynamic limit is compatible with the periodic b.c.s initially used for that thermodynamic limit, as it is for the other models concerned in ref. 19. The numerical results [1,2] nevertheless seem to provide yet another check on our methods, and vice-versa. There exist [15] already other good checks on these methods.

Evidently this work [1,2] by Alwyn Scott and his collaborators is yet once again of considerable interest and importance.

# References

[1] V. Muto, A.C. Scott and P.L. Christiansen, Phys. Lett. A 136 (1989) 33.
[2] V. Muto, A.C. Scott and P.L. Christiansen, Physica D 44 (1990) 75.
[3] M. Toda, Studies on a Nonlinear Lattice, Arkiv for Det Fysiske Seminar i Trondheim, No. 2 (1974).
[4] M. Toda, Solitons in discrete systems, presented at this meeting.
[5] A.C. Scott, V.Z. Enol'skii, M. Salerno and J.C. Eilbeck, There is more than one way to skin Schrödinger's cat, presented at this meeting.
[6] J. Willard Gibbs, Elementary Principles of Statistical Mechanics (Yale Univ. Press, Yale, 1902; Dover, New York, 1960).
[7] R.K. Bullough, N.M. Bogoliubov and G.D. Pang, The quantum Ablowitz–Ladik equation as a $q$-boson system, and references, presented at this meeting.
[8] H. Bethe, Z. Phys. 71 (1931) 205.
[9] R.P. Feynman and A.R. Hibbs, Quantum Mechanics and Path Integrals (McGraw-Hill, New York, 1965).
[10] R.K. Bullough and J. Timonen, in: Microscopic Aspects of Nonlinearity in Condensed Matter, eds. A.R. Bishop, V.L. Pokrovsky and V. Tognetti (Plenum, New York, 1991) pp. 263–280.
[11] H.B. Thacker, Rev. Mod. Phys. 53 (1981) 253.
[12] C.N. Yang and C.P. Yang, J. Math. Phys. 10 (1969) 1115.
[13] G. Vitiello, Squeezing and quantum groups, presented at this meeting.
[14] R.K. Bullough and N.M. Bogoliubov, in: Differential Geometric Methods in Theoretical Physics, eds. Sultan Catto and Alvany Rocha (World Scientific, Singapore, 1992) pp. 488–504.

[15] J. Timonen, M. Stirland, D.J. Pilling, Yi Cheng and R.K. Bullough, Phys. Rev. Lett. 56 (1986) 2233.

[16] R.K. Bullough, D.J. Pilling and J. Timonen, J. Phys. A 19 (1986) L955.

[17] J. Timonen, R.K. Bullough and D.J. Pilling, Phys. Rev. B 34 (1986) 6523.

[18] J. Timonen, Yu-zhong Chen and R.K. Bullough, Nucl. Phys. B (Proc. Suppl.) 5A (1988) 58.

[19] R.K. Bullough, Yu-zhong Chen and J. Timonen, in: Nonlinear World, eds. V.G. Bar'yakhtar, V.M. Chernousenko, N.S. Erokhin, A.G. Sitenko and V.E. Zakharov (World Scientific, Singapore, 1990) pp. 1377–1422.

[20] R.K. Bullough, J. Timonen, Yu-zhong Chen, Yi Cheng and M. Stirland, in: Nonlinear Evolution Equations, Integrability and Spectral Methods, eds. A. Degasperis, A.P. Fordy and M. Lakshmanan (Manchester Univ. Press, Manchester, 1990) pp. 605–617.

[21] Yu-zhong Chen, Ph.D. Thesis, University of Manchester (1989).

[22] R.K. Bullough, N.M. Bogoliubov and J. Timonen, Physica D 51 (1991) 43.

[23] H. Bolterauer and M. Opper, Z. Phys. B 42 (1981) 155.

[24] F. Mertens and H. Büttner, Phys. Lett. A 84 (1982) 335.

[25] F. Mertens and H. Büttner, J. Phys. A 15 (1985) 1831.

[26] F. Yoshida and T. Sakuma, Phys. Rev. A 25 (1982) 2750.

[27] V. Muto et al., Nonlinearity 4 (1990) 477; Phys. Rev. A 42 (1990) 7452;
P.L. Christiansen, DNA Models, presented at this meeting.

[28] F. Marchesoni and G. Lucheroni, Soliton Density in an Infinite Toda Lattice, preprint, Department of Physics, University of Perugia (Italy).

[29] L.D. Faddeev and L.A. Takhtajan, Hamiltonian Methods in the Theory of Solitons (Springer, Berlin, 1987).

PHYSICA ⓓ

# Nonlinear models of DNA dynamics

Peter L. Christiansen[a] and Virginia Muto[b]

[a]*Laboratory of Applied Mathematical Physics, The Technical University of Denmark, DK-2800 Lyngby, Denmark*
[b]*Departamento de Matemática Aplicada, Estadistica e Investigación Operativa, Facultad de Ciencias,
Apartado 644, Universidad del Pais Vasco, 48080 Bilbao, Spain*

The effect of nonlinearity in models of DNA with homogeneous and inhomogeneous strands is investigated. Numbers of solitons and numbers of long-lived open states as function of temperature are computed.

A number of researchers have suggested that anharmonicity may play a role in the dynamics of deoxyribonucleic acid (DNA) molecules at physiological temperatures. A summary of this controversial question with an extensive list of references was given in 1990 in [1]. At the same time studies of a dynamical model in which the hydrogen bonds in the base pairs of the DNA molecule were modelled by nonlinear springs, governed by Morse potentials, were presented in [2,3]. A multiple scale perturbation analysis leads to a continuum model approximation, based on nearest neighbors' interactions, in terms of the nonlinear Schrödinger equation. Inclusion of long range linear interactions, however, tends to disperse pulses propagating along the strands of the DNA molecule [4]. In the DNA molecule there is a succession of adenosine-thymine (A-T) and guanine-cytosine (G-C) base pairs. The A-T bond involves two hydrogen bonds whereas the G-C bond involves three hydrogen bonds. The influence of this type of inhomogeneity on the localization and propagation of energy along the DNA strand was investigated in [5].

The temperature problem for the nonlinear DNA model with a transversal Morse potential was treated in [2] by statistical mechanics methods. In [6,1] the DNA molecule is modelled by a Toda lattice in which each particle represents a single base pair and the nonlinear springs connecting neighbouring particles represent the Van der Waals potentials between adjacent pairs. The system was thermalized either by choosing the initial conditions from Gaussian random distributions [6] or by adding a damping force and a noise force which simulate the interaction with a thermal reservoir at finite temperature, $T$ [1]. The resulting Langevin equations are then perturbed Toda equations for each particle.

Due to the integrability of the Toda lattice, it becomes possible to measure the effect of the nonlinearity in a very concrete manner, namely by using the machinery of the inverse scattering transform to count the number of solitons e.g.

Numerical integration of the perturbed Toda equations and subsequent application of the soliton counter [1] shows that, when thermal equilibrium is reached, a significant number of solitons, $N_s$, is generated in the DNA model at physiological temperatures, namely $N_s \simeq 0.31N$, where $N$ is the number of base pairs. Moreover, from dynamical simulations at different temperatures it was found that $N_s \propto T^{1/3}$. This power law is observed independently of the coefficient used in the friction force.

In [7] this temperature dependence was derived by statistical mechanics for a free soliton gas. A linear dependence on temperature for low temperatures was found in [8,9] considering the

density of states. Using the interacting soliton phonon theory a $T^{1/3}$ law was found in [10,11]. However, the difficulty of physically distinguishing low-energy Toda solitons from long-wave length phonons is underlined in [12]. Further work based on the Bethe ansatz is reported in [13,14]. Recent analytical work, [15,16], arrives at a $T^{1/3}$ law for the soliton density.

In order to make a more realistic model of the DNA dynamics the molecule must be described by two chains, transversely coupled, where each chain simulates one of the two polynucleotide strands. The hydrogen bond between the two bases in a pair can be modelled by a Lennard–Jones potential as suggested in [17, 18]. In [19] this system was thermalized by inclusion of damping and noise forces. The masses, representing the bases, were permitted to perform longitudinal as well as transversal motions. Already, an unperturbed single Toda chain in which the masses can move in this manner, is not an integrable system. However, in the case where the longitudinal and transversal strains are of the same order of magnitude, a supersonic solitary hybrid wave can propagate for a while without dispersing, if the velocity and the parameter, describing the coupling between the longitudinal and the transversal modes, are small [20]. Correspondingly, in a continuum limit based on equations of improved Boussinesq type analytical solutions, travelling with a characteristic velocity, exhibit stable behaviour or blow-up (of the longitudinal component) [21]. Other supersonic analytical solutions, travelling with arbitrary velocity [22], can be shown to be unbounded. The existence of blow-up has recently been demonstrated by analytical methods [23].

The influence of the specific base sequences along the helix has been investigated in a number of papers. Here we mention a recent study of DNA-promoter dynamics based on a discrete sine-Gordon model [24]. Propagation and pinning of solitons were found in various segments of the inhomogeneous double strand.

In the present work we consider the influence of the inhomogeneity on the dynamics of the DNA helix described by the Toda–Lennard-Jones model. The inhomogeneous Toda–Lennard-Jones model is a more realistic description of the DNA helix since the two strands of the helix are described by two chains transversely coupled and the inhomogeneity of the molecule is taken into account. As in [19], the anharmonic Van der Waals potentials are described by a Toda potential with lattice constant $l$:

$$V_T(l) = \frac{a}{b_n} \exp(-b_n l) + al, \tag{1}$$

and the hydrogen bonds are modelled by a Lennard-Jones potential

$$V_{LJ}(d) = 4\epsilon_n \left[ \left(\frac{\sigma}{d}\right)^{12} - \left(\frac{\sigma}{d}\right)^6 \right]. \tag{2}$$

The index $n$ appears now explicitly to indicate the presence of inhomogeneity. Denoting the weak A-T base pairs by W, the strong G-C base pairs by S, and an averaged base pair by H, the presence of inhomogeneities will be described using different values for the linear spring constant, $k_n = ab_n$, in the Toda potential, and for the Lennard–Jones parameter, $\epsilon_n$, related to the stacking energy and the hydrogen bond strength, respectively. DNA sequences of the following type have been considered:

X...XWSWS...WSWSX...X ,

where X stands for W, S, or H. In the Toda–Lennard-Jones model [19,25,26] the linear spring constants in the Toda potential between W and S, W and W, S and S, and H and H are (a) $k_{W-S} = 1\,N/m$, $k_{W-W} = k_{S-S} = 29\,N/m$, and $k_{H-H} = 15.8\,N/m$ and (b) $k_{W-S} = 0.1\,N/m$, $k_{W-W} = k_{S-S} = 29.9\,N/m$ and $k_{H-H} = 15.8\,N/m$, respectively. In the Lennard-Jones potential for W, S, and H we use $\epsilon = \epsilon_W = 0.0176\,eV$, $\epsilon = \epsilon_S = 0.0264$ eV, $\epsilon = \epsilon_H = 0.022\,eV$, and $\sigma = 2.7 \times 10^{-10}$ m in case (a) and (b).

In fig. 1 the number of solitons as function of temperature is shown for the cases of inhomoge-
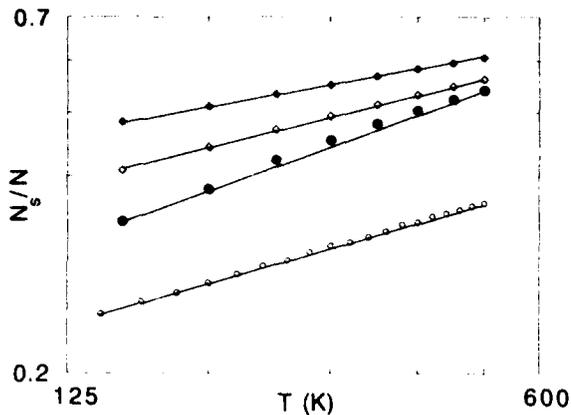
Fig. 1. Number of solitons, $N_s$, versus temperature, $T$, in logarithmic scale, for $N = 64$ masses. Circle: single strand model [6,1]. Bullet: homogeneous double strand model ($\epsilon = 0.022$ eV, $k = 15.8$ N/m, $\sigma = 2.7 \times 10^{-10}$ m). Diamond: inhomogeneous double strand model for parameter values of case (a) (empty diamond) and case (b) (full diamond).

Fig. 2. Number of open states of hydrogen bonds versus temperature, $T$, with life-times: 5–20 psec (open symbols), longer than 20 ps (full symbols). Homogeneous chain (circles), inhomogeneous chain with prevalence of bases of strong type (triangles), homogeneous type (diamonds), and weak type (squares).

neous chains with prevalence of H type (X = H) for parameter values of case (a) (empty diamond) and case (b) (full diamond), respectively. This soliton ratio is compared with the results of the homogeneous double strand model ($\epsilon = 0.022$ eV, $k = 15.8$ N/m, and $\sigma = 2.7 \times 10^{-10}$ m) (bullet) and with the ones of the homogeneous unperturbed single strand model (circle) [19]. The effect of the nonlinearity is stronger in the double strand model than in the single strand model. Thus $N_s/N \approx 0.49$, 0.55, 0.45, and 0.31, respectively, at physiological temperature. Results very similar are obtained for chains with prevalence of strong or weak base pairs.

In fig. 2 we show the number of open states of the hydrogen bond as function of temperature, for parameter values of case (a). Case (b) is considered in [26]. Open symbols refer to life-times 5–20 psec, while full symbols correspond to lifetimes longer than 20 psec. It is seen that the presence of inhomogeneities clearly enhances the hydrogen bond breaking. When bonds of weak type, W, prevail most open states are found. Thus at physiological temperatures the thermal fluctuations of the surrounding medium may cause the hydrogen bond breaking with a

significant lifetime. However, this does not imply that all bonds are permanently broken. Such a state of full denaturation may be caused by other effects such as bending of DNA [27,28].

## References

[1] V. Muto, A.C. Scott and P.L. Christiansen, Physica D 44 (1990) 75.

[2] M. Peyrard and A.R. Bishop, Phys. Rev. Lett. 62 (1989) 2755.

[3] M. Techera, L.L. Daemen and E.W. Prohofsky, Phys. Rev. A 40 (1989) 6636.

[4] M. Techera, L.L. Daemen and E.W. Prohofsky, Phys. Rev. A 41 (1990) 4543.

[5] M. Techera, L.L. Daemen and E.W. Prohofsky, Phys. Rev. A 42 (1990) 1008.

[6] V. Muto, A.C. Scott and P.L. Christiansen, Phys. Lett. A 136 (1989) 33.

[7] T. Schneider and E. Stoll, Classical Statistical Mechanics of the Toda Lattice: Static and Dynamics Properties, private unpublished communication (1980).

[8] F. Mertens and H. Büttner, Phys. Lett. A 84 (1981) 155.

[9] F. Mertens and H. Büttner, J. Phys. A 15 (1982) 1831.

[10] F. Yoshida and T. Sakuma, Phys. Rev. A 25 (1982) 2750.

[11] N. Theodorakopoulos, Counting Solitons and Phonons in the Toda Lattics, in: Statics and Dynamics for Nonlinear Systems, eds. G. Benedek et al. (Springer, 1983) p. 271.

[12] N. Theodorakopoulos, Phys. Rev. Lett. 53 (1984) 871.

[13] M. Opper, Phys. Lett. A 112 (1985) 201.

[14] K. Sasaki, Phys. Rev. B 35 (1987) 3445.

[15] F. Marchesoni and C. Lucheroni, Phys. Rev. B 44 (1991) 5303.

[16] R.K. Bullough, Y. Chen and J.T. Timonen, Physica D (1993).

[17] P. Perez and N. Theodorakopoulos, Phys. Lett. A 124 (1987) 267.

[18] J. Halding and P.S. Lomdahl, Phys. Lett. A 124 (1987) 37.

[19] V. Muto, P.S. Lomdahl and P.L. Christiansen, Phys. Rev. A 42 (1990) 7452.

[20] P.L. Christiansen, P.S. Lomdahl and V. Muto, Nonlinearity 4 (1990) 477.

[21] P.L. Christiansen, S. Rionero, P.L. Christiansen and V. Muto, Stability of a Solitary Wave Solution to a Set of Coupled Boussinesq-like Equations, in: Waves and Stability in Continuous Media, ed. S. Rionero (World Scientific, 1991) p. 49.

[22] P.L. Christiansen, V. Muto and S. Rionero, Chaos, Solitons and Fractals 2 (1992) 45.

[23] S.K. Turitsyn, Phys. Rev. E 47 (1993).

[24] M. Salerno, Phys. Rev. A 44 (1991) 5292.

[25] V. Muto, Nanobiology 1 (1992) 325.

[26] V. Muto, F. Vadillo and M. Lezaun, to appear.

[27] S. Swaminattan, G. Ravishanker and D.L. Beveridge, J. Am. Chem. Soc. 113 (1991) 5027.

[28] J.M. Withka, S. Swaminattan, J. Srinivasan, D.L. Beveridge and P.H. Bolton, Science 255 (1992) 597.

PHYSICA D

# Nonlinear science: toward the next frontiers

James A. Krumhansl[1]

*Physics Department, University of Massachusetts, Amherst, MA 01002, USA*

Nonlinear science has erupted in many directions over recent years, with many successes. Two main themes have been found in many different settings – chaos and solitary phenomena. While found initially in simple models, useful for establishing mathematical methods and the behavior of exact limiting cases, their robust features give us assurance that they are generic. The next frontier then is not so much in formalism, but in how to achieve a realistic, eventually utilitarian, representation of an actual biological, material, physical, or chemical system that connects the experimentally significant behavior with its nonlinear properties. An approach is proposed.

## 1. Introduction

The thoughts presented here are an outgrowth of a workshop on Future Directions of Nonlinear Dynamics in Physical and Biophysical Systems at Lyngby, Denmark, July 1992. There is no doubt in my mind that there is an important place for the widespread application of the great amount that has been learned about nonlinearity, but I have developed concerns whether the research as it is now going will succeed in the broad sense hoped for.

Over the past twenty five years nonlinear science has flowered, in mathematics, the physical sciences, and engineering. Bolstered by the great power of computer simulation, formal theories and the experimental identification of nonlinear phenomena have made great progress. The generic character of chaos and solitary excitations have been established; general mathematical methods, ever increasing sophistication of computer simulations, and powerful new experimental techniques, particularly in traditional areas of physics (e.g. fluid dynamics) have achieved marked success.

In fields not so traditional, such as biological science, metallurgy, and others, particularly where mathematics has been less used, the recognition and incorporation of modern nonlinear concepts and methods have not progressed very far. This is not only unfortunate, but unnecessary. The challenge of the next era in nonlinear science is to find what is necessary to close the gap that exists.

The root of the problem as I see it, exaggerated perhaps, is that more and more nonlinear research is becoming either marginal or irrelevant, aided and abetted by the wide availability of larger and larger computers, and the ease of formulating variations on a basic mathematical theme and doing one more case. Indeed, many of these incremental explorations yield fascinating special features. However, the important questions are: first, do they extend our general understanding; or, alternatively, do the special features really provide new, quantitative insight to some particular experimental observation? It's not clear in many instances of published research today whether the answer to either is in the affirmative. Nor do I mean to direct my skepticism only to nonlinear science, though the issue seems to me to be of especial importance to a developing science looking for opportunities on the widest possible front.

[1] Horace M. White Professor of Physics Emeritus, Cornell University; Adjunct Professor of Physics, University of Massachusetts Amherst.

## 2. Criticism

Let me illustrate two sources of my concerns, which I will call marginal modeling, and science by advertisement. To convey what I mean by each I would like to review some history in my own field, condensed matter physics.

One of the great revelations following the discovery of quantum mechanics was the application of the theory to periodic crystals and the discovery of electronic energy bands, developed especially by Bloch, Bethe, Slater and others in the 1920-30s. For a number of years thereafter there were explorations of all sorts of models, periodic square well potentials, multiple sine wave potentials, muffin tin potentials, etc., but looked at objectively it cannot be said that any of these studies added very much substantively new physics. These were "marginal models", albeit some quite interesting as puzzles at the time. There was a resurgence of this topic in the 1950-60s, but with new justification as will be discussed below. A similar situation is developing in certain areas of nonlinear science today, particularly as new supercomputers allow the exploration of more and more complex model problems; but only models they are, for the most part. There are notable exceptions, as in hydrodynamics where exploration of singular properties, local structures, and turbulence has maintained close and faithful contact with the physics.

Certainly, there is a time in the development of a new topic when simple, testable models are of great importance in reconnoitering general behavior. However, as progress provides a generic understanding, models should make way for reality; the statistical mechanics of one more elaboration of a nonlinear potential on a one dimensional chain ought to be explored only if there is a specific suggestion from experiment or other physical reason to test that particular variation on the general theme. Indeed, there are in the literature many extremely sophisticated nonlinear simulations which have been carried out, that show interesting behaviors; but little effort

has been expended to seek out substantively their presence or absence in situations they allegedly represented. This is what I call science by advertisement.

Again, in my own field, to take an example, there have been a series of recent, admittedly titillating, reports on stochastic lattice models of the spread of forest fires. Indeed the dynamics of the models presents neat problems and behavior, which might also apply to corrosion, spread of surface infection, and so forth - might, that is. It is curious, however, that in the bibliographies of the papers I have seen so far the only references are to the physics or mathematics literature, not to forestry publications. Whether or not the models are indeed applicable to real forest fires, it is all too easy to entitle the research as such. The same criticism can be applied to many of the current nonlinear dynamical models of the conformations of biomolecules; they are biology by advertisement.

If the only message I had were negative, it would be unfortunate. That is not the case, however. The viewpoint to take is that nonlinear science has come a long way, and likely can make many further, important contributions to other fields, but it will not unless the gap between formal theories or simulations and the real scientific problems in those other fields is narrowed. Here are some suggestions, with guidance from history.

## 3. New frontiers; basic rules

Many challenging nonlinear problems arose originally from the observation of some natural phenomenon: limit cycles in nonlinear coupled differential equations from modeling competing fish populations in the Baltic Sea, chaos from modeling weather dynamics. In the simplest cases the modeling format may be apparent from the nature of the phenomenon; that, however, is usually not the case. Assuming that the objective is to have nonlinear dynamics describe reality, I

)pose the following check list for developing
)dels:

(1) Identify the phenomenon to be studied.
) so only after consulting with experimentalists
find out what they believe to be important.
itially, what *we* think is important may not be
evant; try to do their problem first. Reduc-
·nism from "first principles" only hardly ever
irks.

(2) Use physically realistic variables, i.e.
rder parameters". The endemic difficulty here
that formal mathematical or theoretical sci-
tists expect experimental scientists to pull out
general models the important results that
ate to their need. This is unrealistic; the inter-
ion must be a two way street. The nonlinear
entist must make the effort to learn enough
out biology, metallurgy, quantum electronics,
whatever the topic may be both to talk to the
perimentalist and to use common sense in
)deling. Clearly, it is important to find the
nimum number of functionally important
ordinates, but not more – a clever trick.

(3) Check frequently in the carrying out of
nulations whether it is better to reconsider the
)deling of the science or to increase the detail
e. size) of simulation as the research goes
·ward.

Let me cite a few instances by way of general
istration, then speak more specifically to some
)molecular examples. Earlier I noted the his-
·y of electronic energy band physics, and that
er an early period developing the foundations
·re was a hyatis of sorts, until the 1950–60s.
hat then happened was the sudden entry of
niconductor physics and its application to elec-
·nic devices; the science then became focussed
a few very specific questions, i.e. (1) above,
d energy band theory had a renaissance.
ose in the field we recall that then, instead of
possible aspects of electron dynamics being
)deled by theory, process (2) took place; elec-
:al properties were well described by two
uids", electrons and holes, with a few aniso-
·pic effective mass parameters. Meanwhile

process (3) took place continually, refining the
modeling, the questions asked of the simula-
tions, and the interpretation of results.

What about nonlinear dynamics these days?
There is a spectrum of answers, depending on
the area of application. I have the impression,
that in the field of hydrodynamics and plasma
physics, at least, that the three rules above are
employed very well; this may not be too surpris-
ing since the basic physical variables are easily
identified, and the elementary laws governing
them are well known. In this case reproducible,
physically significant phenomena, including col-
lective and local nonlinear behaviors, are found
and correlated to theory. An excellent overview
of this area of nonlinear science is given recently
by Rabinovich, Fabrikant, and Tsimring [1].

The situation in other areas is not nearly as
well defined; one close to my own interests for
the past decade has been that of topological
solitons to describe transitional behavior in dis-
placive structural transformations of solid ma-
terials, of which martensitic materials (hardened
steels, shape memory alloys) are a particularly
interesting subclass. The early simulations were
done with ball and spring models, but the subject
has progressed much further in quantitative rele-
vance to reality, with the participation of metal-
lurgists, and through the efforts of a number of
experimental and theoretical researchers focus-
sed on *common* problems [2,3]. The rules stated
above now apply there. Nonetheless, *in spite of*
this progress toward joining nonlinear and ma-
terials science, much work continues in the ball
and spring vein, involving increasingly extensive
simulations. Let me move on, however, to the
much less clear case of biomolecular nonlinear
science.

## 4. Biomolecular conformations; proteins and DNA

Biology is concerned with structure at several
different levels; primary, the local molecular

bonding; secondary, the extended macro-molecular geome ry, as in proteins and DNA; tertiary, as in functional units such as hemoglobin; and quaternary, such as chromosomes and larger units. Of course beyond that is cellular biology, and so forth. Pause for a moment to compare this with the situation in fluid dynamics, which does not have these distinct levels of mesostructure. There is plenty of evidence suggesting nonlinear behavior at most of these levels, but it must be obvious that the issues and the physical phenomena involved are different between each; biology is far more complex than fluid dynamics. It is of the essence, then, to select problems carefully.

For example, the subject of energy transport in proteins is important biologically. Alwyn Scott [4], Davydov [5], and others have in effect followed rule (1) by choosing the particular problem of energy transport in alpha-helical proteins, and rule (2) by choosing the amide bond vibronic excitation as the primary physical variables. Rule (3) has fallen into place rather naturally as a number of experiments stimulated by the proposition of vibronic solitons were carried out over the past decade. Although open questions remain, even about the necessary quantum formulation of the nonlinear dynamics, much contact has been made with biologists.

The question of protein folding conformations is an entirely different matter, of paramount importance in molecular biology, and many aspects remain open. Clearly the physical variables involved, the equilibrium criteria, and the dynamics are fundamentally different from those for the energy transport problem; but again there are clear signs of important nonlinear behavior to be explained. A very useful introduction to the structure and dynamics of biological macromolecules is the book by McCammon and Harvey [6]; a very recent quantitative study of protein dynamics, and the development of optimally descriptive coordinates, has been reported by Garcia [7] on the hydrated protein, crambin.

Beyond proteins, however, the beauty and the challenge of the concepts involved in genetic expression by the polynucleic acids DNA and RNA have captivated our scientific interests, no matter what field we are in. It is only natural that we ask the question whether nonlinear dynamics plays a role in the expression process. A priori there is no particular reason that this be the case. In the context of biochemistry the triumph of Watson, Crick and Wilkins [8] was to combine necessary capabilities from chemistry, physics, and biology in the double-helical (phosphate-ribose sugar) ladder with (purine-pyrimidine) base pair rungs, appropriately hydrated. All of the pieces were necessary to simultaneously combine the stability, the ability to store the genetic code, and the capability with the proper opening of the helix tr transfer a chemical copy of the code to produce selected proteins, i.e. expression. The combined biological and biochemical behavior has been established in great detail. The issue here is what the kinetics or dynamics of the expression process might be, for that is where nonlinear dynamics might play a role. Now, returning to the earlier discussion, we must face the issue of how and what to model realistically.

There are two aspects of the expression process that one might focus upon: first, that base pair breaking and opening of the helix to expose the code sequence occurs in a localized region, suggesting a solitary localized concentration of strain energy; second, because it has been noted experimentally that there are enzymes which seem to promote at one location on the helix to induce transcription at another location, a localized excitation of some sort might propagate from one region to the other. In either case some sort of solitary excitation might exist.

On the other hand the entire process may be a series of chemical steps, localized chemical reactions together with diffusion along the helix between steps. Clearly, given the complexity of the DNA helix as a dynamical system it is quite beyond reason to think of doing any formal

neral analysis of the dynamics seeking some ʒcial excitations which play a key role in bioʒical function. There are only two choices: to ʒort to experiment for guidance toward imʒrtant motions, or to attempt complete molecuʒdynamic simulations starting with molecular/ ʒmic interaction potentials – or, better, to use ʒormation from both. But to proceed it is ʒential to see the molecular structure of a ʒA helix.

However, before going directly to a possible ʒdeling it is necessary again to aquaint matheʒticians and physicists with some parts of the ʒg history of chemical molecular dynamics. ʒpically, in simple solids the atoms, being ʒsely packed, can only change their interaction ʒergy by small motions which change the disʒce between them. Thus, potential energy ʒctions in the form of two-center Hooke's law ʒings are not a bad approximation. Literally ʒusands of simulations have been done with ʒss and spring models *in this vein.* Unforʒately these are almost completely irrelevant ʒmany situations in real materials, and parʒularly to biological systems; why? Almost all ʒlogical material is polymeric, covalently (diʒtionally) bonded, and with open structures. ʒis now allows a number of additional degrees ʒfreedom, generally of much lower excitation ʒrgy than bond stretching, to come into play, ʒh as bond bending and twisting about coʒent bonds. Such distortions are far easier to ʒite than bond stretching, typically 1–2 orders ʒmagnitude lower in energy. In fact, the potenʒls needed for biomolecules are of the form [6]:

$$= \frac{1}{2} \sum_{\text{bonds}} K_b (b - b_0)^2$$

$$+ \sum_{\text{bond angles}} K_\theta (\theta - \theta_0)^2$$

$$+ \frac{1}{2} \sum_{\text{dihedral angles}} K_\phi [1 + \cos(n\phi - \delta)]$$

$$+ \sum_{\substack{\text{nonbonded} \\ \text{pairs}}} \left( \frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_1 q_2}{Dr} \right).$$

In order, the terms are, respectively, bond streching (Hooke), bond bending, bond twisting, core repulsion, Van der Waals, and Coulombic, since many ions are charged. Biological molecules must operate at temperatures of order 300 K, for which thermal energy is nowhere near enough to break bonds (therefore base pair breaking simply by dynamical excitations is hardly likely), but bending and twisting of biological macromolecules can easily be excited. In fact, today it has become clear from both experiment and large scale simulations [9] that the key excitations which participate in DNA expression are the bending and twisting distortions, which allow the molecule to accomodate attachment of a suitable enzyme, that in turn either locally or in coordination with other parts of the molecule provides sufficient chemical energy to open the coded base pair sequence.

What then is the current thinking on the dynamical degrees of freedom of a DNA macromolecule? Reference to the book by McCammon and Harvey [6] is a good place to start, but to provide some image at least of what is involved, in fig. 1 is shown a section of B-DNA, together with a schematic of the significant torsional degrees of freedom along a representative section of one strand. The torsional i.e. dihedral motions *cannot be represented by pair potentials*, since for purely geometric reasons it takes 4-atom positions to define twisting about one bond. The main components are: a phosphate unit ($PO_4$) of which the P and two O are part of the helical ladder, a 5-membered ribose sugar ring (ribose; with or without an attached oxygen i.e. oxy or deoxy) one of whose edges is part of the helix while opposite corner ties to the base which in turn connects to the base on the complementary helix. The main degrees of freedom (per unit) are bending of PO bonds, twist of the ribose ring, and thereby twist of the attached base. These perform the dominant motions; base pair opening is not likely to take place as a result of the intrinsic dynamics.

It is far beyond the scope of this short report
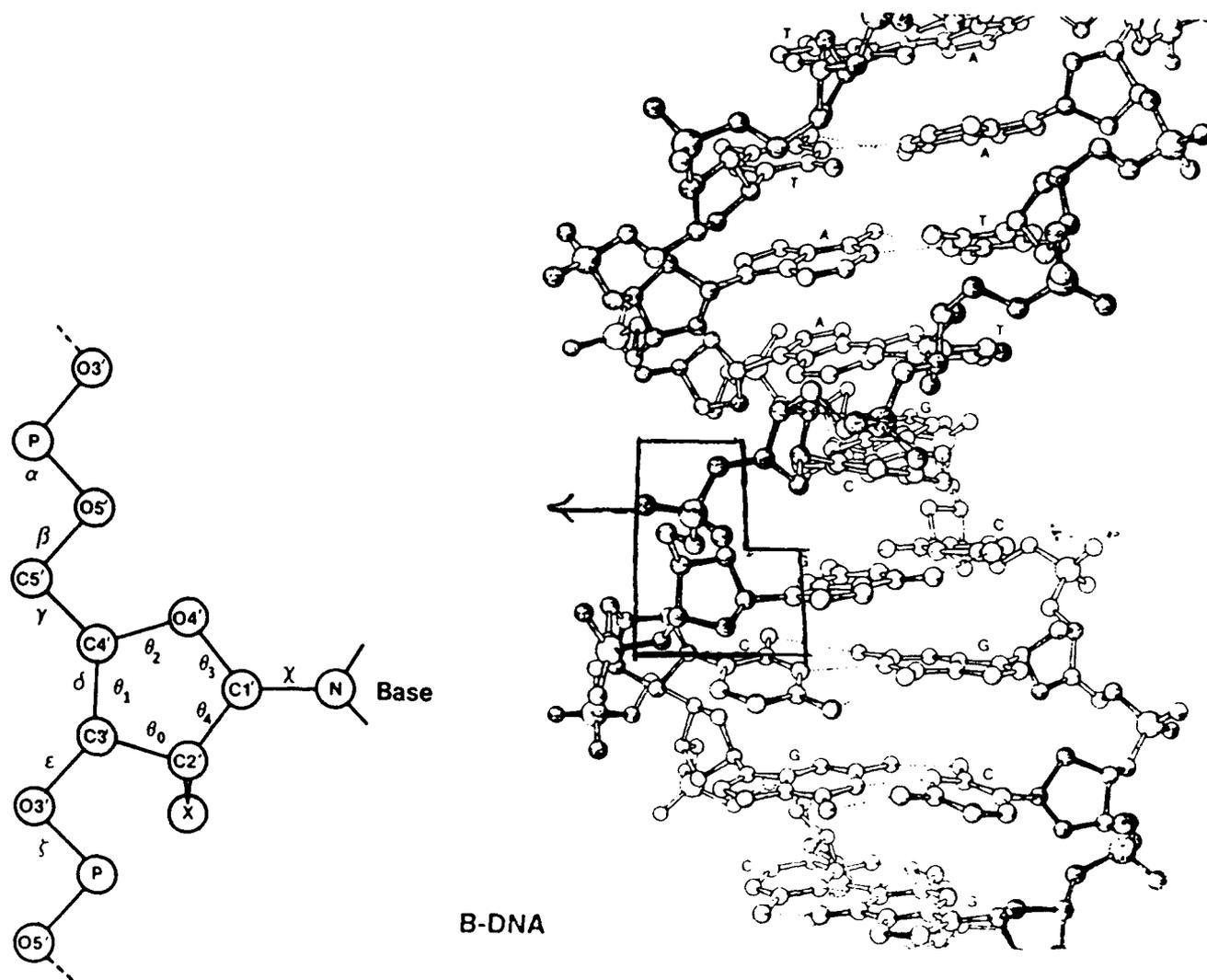
Fig. 1. IUPAC nomenclature for backbone torsion angels, ribose endocyclic torsions, and the glycosidic torsion. The backbone torsion $\delta$ is defined by the four atoms O3′–C3′–C4′–C5′, while the endocyclic ribose torsion about the same bond, $\theta_1$, is defined by C2′–C3′–C4′–O4′. The proper definition of $\chi$ is O4′–C1′–N1–C2 for pyrimidines and O4′–C1′–N9′–C4 for purines.

to discourse on the many fascinating types of dynamical and static distortions, general conformations, that this beautiful molecular system can sustain. To data, almost all of what we known comes from "brute force" massive simulations[*1]. Are there possibly generic nonlinear modes? The answer I would guess would be a

*1 See [10], this reference provides not only detailed results of large, realistic simulations, but an extensive bibliography as well.

cautious affirmative; the plane of the 5-membered sugar ring is puckered and seems to have two possible minimum energy configurations, describable by a 2-well potential function. These, coupled by the phosphate ions along the backbones might give solitary kinked bending modes [11]. The most difficult challenge facing nonlinear science at this time in biomolecular dynamics is how, if indeed it is possible, to model the significant biological excitations with a

few important modes of excitation[*2]. The time is near when we may be able to make progress toward this goal, as long as we now learn enough of the up the date molecular biology literature to take a quantum jump toward biologically realistic current nonlinear models.

## Acknowledgements

[*2]See the editorial comment in [12]: "... it is remarkable, in this day and age, how simple are the systems that form the substrate of people's elaborate calculations. That is one measure of how far there is to go to the treatment of the real world ...".

## References

[1] M.I. Rabinovich, A.L. Fabrikant and L.Sh. Tsimtring, Usp. Phys. Nauk (Soviet Phys. Uspekhi) August 1992.

[2] L.E. Tanner and M. Wuttig, eds., First order displacive phase transitions, Mat. Sci. Eng. A 127 (1990) 137–276.

[3] G.R. Barsch and J.A. Krumhansl, in: Nonlinear Physics in Martensitic Transformations, Martensite, eds. G.B Olson and W.S. Owen (ASM Int., 1992)[ISBN:0 87170-434-X].

[4] A.C. Scott, Phys. Rev. A 26 (1982) 578.

[5] A.S. Davydov, Solitons in Biological Systems (Reidel, Dordrecht, 1991).

[6] J.A. McCammon and S.C. Harvey, Dynamics of Proteins and Nucleic Acids (Cambridge Univ. Press, 1987).

[7] A.E. Garcia, Phys. Rev. Lett. 68 (1992) 2696.

[8] See the review by R.E. Dickerson et al., Science 216 (1982) 475.

[9] N. Angier, New York Times, August 4, 1992, Science News section.

[10] S. Swaminathan, G. Ravishanker and D.L. Beveridge, J. Am. Chem. Soc. 113 (1991) 5027.

[11] J.A. Krumhansl et al., in: Structure and Motion, Nucleic Acids and Proteins, eds. E. Clementi, G. Corongiu, M.H. Sarma and R.H. Sarma (Adenine Press, 1985).

[12] John Maddox, Nature 359 (1992) 771.

**PHYSICA** D

# Biomolecular dynamics of DNA: statistical mechanics and dynamical models

M. Peyrard[a], T. Dauxois[a], H. Hoyet[b] and C.R. Willis[c]

[a]*Laboratoire de Physique, CNRS URA 1325, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France*
[b]*Physique non linéaire: Ondes et Structures Cohérentes, Faculté des Sciences, 6 blvd Gabriel, 21000 Dijon, France*
[c]*Department of Physics, Boston University, 590 Commonwealth Ave, Boston, MA 02215, USA*

There is a growing feeling that biomolecular structure is not sufficient to determine biological activity which is also governed by large amplitude dynamics of the molecules. The transcription of DNA or its thermal denaturation are typical examples. Traditional approaches use Ising models to describe the denaturation transition of DNA. They have to introduce phenomenological "cooperativity factors" to explain the rather sharp "melting" of this quasi one-dimensional system. We present models which describe the full dynamics of the melting. Using molecular dynamics simulations and statistical analysis, we discuss the mechanism of the denaturation, including precursor effects that can be related to large amplitude localized nonlinear excitations of the molecule in which discreteness effects play a large role. We also show the microscopic origin of the cooperativity factors.

## 1. Introduction

Nucleic acids are the repository of genetic information and each of the units that compose DNA or RNA molecules plays an essential role in the biological functions. The famous discovery of the double helix has emphasized a strong relationship between *structure* and function in molecular biology. However this structure, which is so well designed to include the genetic code in two complementary strands and protects it against external perturbations, would also prevent the expression of the code if the molecule were static because the coding bases are not directly accessible to chemical reaction. There are however many indications that DNA is a very *dynamical* entity, undergoing very large deformations and should not be viewed merely as a solid with a particular structure.

A typical example in which the dynamics of the molecule is essential for its function is DNA transcription during which a segment of the genetic code is copied into RNA. In order to expose the coding bases to chemical reaction, the double helix unwinds locally and forms a "bubble" which is about 20 base-pair long and moves along the molecule as the transcription proceeds. This complex process, which is activated by an enzyme, is still beyond a physical analysis but it has strong similarities with the early stage of the termal denaturation, or "melting" of the double helix. The melting, which is the separation of the two complementary strands, starts locally by the formation of small denaturated regions very similar to the transcription bubble. Another important motion of the DNA molecule is its "breathing" or fluctuational opening. In these very large fluctuations, base-pairs are temporarily broken and the two bases are exposed for chemical reaction for a very short time ($10^{-7}$ s). These fluctuational openings can be considered as intrinsic precursors for the denaturation and they could play a role in carcinogenesis by external molecules [1].

The molecular deformation involved in melting or in fluctuational openings are so large that

they cannot be described by linear approximations. Therefore biomolecular dynamics is a fascinating topic for nonlinear science because it is related to basic phenomena of life and we *know* that it has to be fundamentally nonlinear. We discuss here some aspects of DNA dynamics and we show that simple nonlinear models can provide a good description of the large amplitude distortions of the molecule which are observed experimentally. Our basic approach can be viewed as an extension of the Ising models, which have been widely used to study the statistical mechanics of the melting, in which we treat completely the dynamics of the bases. In section 2 we present our basic model and its statistical mechanics. This section makes the connection with the usual Ising models for DNA. Section 3 studies the dynamics of the model. Molecular dynamics is used to detect the main types of large amplitude motions and connect them with the experimental observations. Then we propose two analytical investigations adapted to the description of the fluctuational openings and denaturation bubbles. Section 4 discusses more precisely the thermodynamics of the melting of this one dimensional object, the DNA molecule. We show how the introduction of nonlinear coupling terms to describe the base stacking interactions is essential to explain the sharp melting transition observed experimentally.

## 2. A simple model for DNA melting and its statistical mechanics

The simplest description of DNA melting represents a base-pair by an Ising-like variable which takes only two values, 0 and 1, i.e., closed and open. The denaturation transition is then analyzed by treating the statistical mechanics of this one-dimensional Ising-spin chain [2]. The structure of DNA appears in the calculation of the statistical weight of each state of the molecule which is expressed as the product of a stability parameter for each base-pair, a

cooperativity parameter taking into account the fact that breaking a base-pair destroys two stacking interactions unless the pair is the terminal pair of an open region, and an entropy parameter for each loop measuring the "stiffness" of the DNA strands. Many varieties of these Ising-models have been presented. For instance the stability parameter can be assigned to base-pair doublets rather than to a single base-pair because, as we discuss in the last section, there is a strong relationship between the stacking interaction of adjacent bases and the stability of the pairs. The model parameters are determined phenomenologically in order to get the best possible agreement between the theoretical predictions and the experimental melting curves. Once a particular model has been calibrated it may be used to predict the melting curve of another DNA segment. The success of this approach to reproduce experimental melting curves is impressive, but it involves a large number of adjustable parameters. For instance, 10 parameters are used to represent the 10 possible types of base-pair doublets along the molecule. Moreover, using an Ising-variable prevents any attempt to describe the dynamics of the fluctuational openings since states intermediate between closed and open cannot be represented.

Our approach goes further but still keeps the model as simple as possible in an attempt to determine the fundamental mechanism of the melting. Therefore we consider a simplified geometry for the DNA chain in which we have neglected the asymmetry of the molecule and we represent each strand by a set of point masses that correspond to the nucleotides. The characteristics of the model are the following:

(i) The longitudinal displacements are not considered because their typical amplitudes are significantly smaller than the amplitudes of the smaller transverse ones [3]. The stretching of a base-pair in the transverse direction is represented by a real variable $y_n$ which can therefore describe all the states of the pair from closed ($y_n = 0$) to completely broken.

(ii) Two neighboring nucleotides of the same strands are connected by an harmonic potential to keep the model as simple as possible. On the other hand, the bonds connecting the two bases belonging to different strands are extremely stretched when the double helix opens locally so that their nonlinearity must not be ignored. We use a Morse potential to represent the transverse interaction of the bases in a pair. It describes not only the hydrogen bonds but the repulsive interactions of the phosphate groups, partly screened by the surrounding solvent as well. The Hamiltonian of the model is then the following:

$$H = \sum_n \left[ \tfrac{1}{2} m \dot{y}_n^2 + \tfrac{1}{2} K (y_n - y_{n-1})^2 \right.$$
$$\left. + D(e^{-a y_n} - 1)^2 \right]. \tag{1}$$

Since we are interested in the thermal denaturation transition of the molecule, the natural approach is to investigate the statistical mechanics of the model. Due to the one-dimensional character of the system, and because the interactions are restricted to nearest neighbor interactions, it can be treated exactly, including fully the nonlinearities, with the transfer operator method [4].

For a chain containing $N$ units with nearest neighbor coupling, the classical partition function, given in terms of the Hamiltonian (1), can be expressed as

$$\mathscr{Z} = \int_{-\infty}^{+\infty} \prod_{n=1}^{N} dy_n \, dp_n \, e^{-\beta H}$$

$$= \mathscr{Z}_p \times \int_{-\infty}^{+\infty} dy_n \, e^{-\beta f(y_n, y_{n-1})}$$

$$= \mathscr{Z}_p \mathscr{Z}_y , \tag{2}$$

where $f(y_n, y_{n-1})$ is the potential part of the Hamiltonian. The momentum part are readily integrated to give the usual kinetic factor for $N$ particles $\mathscr{Z}_p = (2\pi m k_B T)^{N/2}$. The potential part can be evaluated exactly [5–7] in the thermodynamic limit of a large system ($N \to \infty$) using

the eigenfunctions and eigenvalues of the transfer integral operator

$$\int dy_{n-1} \, e^{-\beta f(y_n, y_{n-1})} \phi_i(y_{n-1})$$

$$= e^{-\beta \epsilon_i} \phi_i(y_n) . \tag{3}$$

The calculation is similar to the one performed by Krumhansl and Schrieffer [6] for the statistical mechanics of the $\phi^4$ field. It yields $\mathscr{Z}_y = \exp(-N\beta\epsilon_0)$, where $\epsilon_0$ is the lowest eigenvalue of the operator. We can then compute the free energy of the model as $\mathscr{F} = -k_B T \ln \mathscr{Z} = -(\tfrac{1}{2} N k_B T) \ln(2\pi m k_B T) + N\epsilon_0$ and the specific heat $C_v = -T(\partial^2 \mathscr{F}/\partial T^2)$. The quantity which gives a measure of the extent of the denaturation of the molecule is the mean stretching $\langle y_m \rangle$ of the hydrogen bonds, which can also be calculated with the transfer integral method [5] and yields

$$\langle y \rangle = \langle y_m \rangle = \frac{\sum_{i=1}^{N} \langle \phi_i(y)|y|\phi_i(y)\rangle \, e^{-N\beta\epsilon_i}}{\sum_{i=1}^{N} \langle \phi_i(y)|\phi_i(y)\rangle \, e^{-N\beta\epsilon_i}}$$

$$= \langle \phi_0(y)|y|\phi_0(y)\rangle$$

$$= \int \phi_0^2(y) y \, dy , \tag{4}$$

since in the limit of large $N$ the result is again dominated by the lowest eigenvalue $\epsilon_0$ associated with the normalized eigenfunction $\phi_0(y)$.

In the continuum limit approximation, the TI eigenvalue problem can be solved exactly, but experiments on proton exchange in DNA [8] show some evidence of exchange limited to a single base pair which suggests that discreteness effects can be extremely large in DNA. Therefore we have solved numerically the eigenvalue equation of the transfer operator [9] without approximations. The TI operator is symmetrized and the integral is replaced by sums of discrete increments, using summation formulas at different orders. The problem is then equivalent to finding the eigenvalues and eigenvectors of a symmetric matrix.
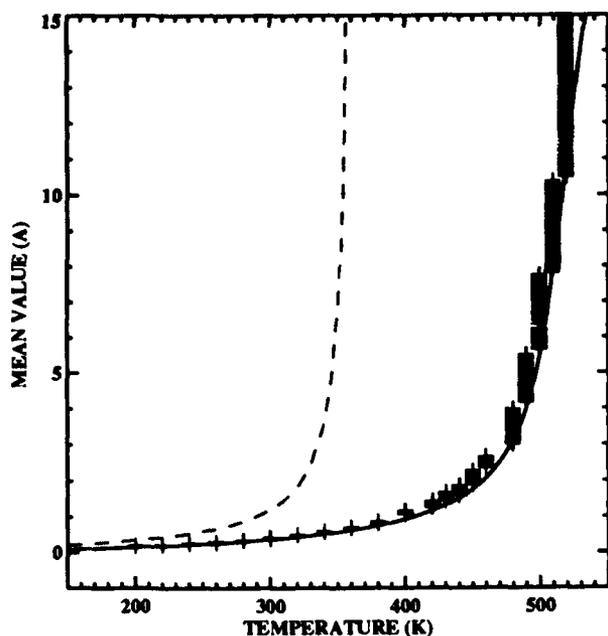
Fig. 1. Variation of $\langle y \rangle$ versus temperature: the dash line corresponds to the TI results in the continuum limit, the solid line gives the exact TI results obtained by numerical solution of the TI operator, and the plus signs correspond to molecular dynamics simulations.

Figure 1 compares the thermal evolution of $\langle y_m \rangle$ obtained with the continuum approximation and the exact numerical calculation, for the model parameters discussed in the next section. Both methods show a divergence of the hydrogen bond stretching over a given temperature, but the melting temperature given by the numerical treatment is significantly higher, pointing out the large role of discreteness in DNA dynamics if one uses realistic parameters for the model. The TI calculation shows that the specific heat has a broad maximum around the denaturation temperature.

## 3. Dynamics of the DNA molecule

The thermodynamics of our DNA model shows that it exhibits a thermal evolution that is qualitatively similar to the denaturation of the molecule observed experimentally. But this statistical approach does not give information on the mechanism of the denaturation, and in par-

ticular, does it start locally by the formation of denaturation bubbles in agreement with the experiments. In order to study this aspect, we have investigated the dynamics of the model in contact with a thermal bath by molecular dynamics simulation with the Nose method [10–11].

Beginning with the Hamiltonian $H$ and the $2N$-dimensional phase space of a chain of $N$ base-pairs with periodic boundary conditions, the fixed temperature canonical ensemble can be simulated by the addition of a single variable $s$, which regulates the energy flows, and an additional parameter $M$, which fixes the scale of the temperature fluctuations. Nose demonstrated that in this phase space of the extended Hamiltonian $H'$, the microcanonical ensemble of $H'$ is precisely the canonical ensemble of $H$ at temperature $T$. This property is only exact for equilibrium properties, but investigations currently in progress [12] show that, provided that the characteristic time of the Nose thermostat, controlled by $M$, is properly chosen, it can also give reliable results for the dynamical properties. Most of the simulations have been performed with a chain of 256 base-pairs with periodic boundary conditions, but in order to achieve better statistics, some simulations have been performed on a Connection Machine-200 with 16384 base-pairs. Equations are integrated with a 4th order Runge–Kutta scheme with a time-step chosen to conserve $H'$ to an accuracy better than 0.001% during a run.

We have chosen a system of units adapted to the energy and time scales of the problem. Energies are expressed in eV, masses in atomic mass unit (a.m.u.) and length in Å. The resulting time unit is $1$ t.u. $= 1.0214 \times 10^{-14}$ s. The choice of appropriate model parameters is a very controversial topic, as attested by the debate in the literature [13]. There are well established force fields for molecular dynamics of biological molecules, but they have been designed to provide a good description of the *small* amplitude motions of the molecule and are not reliable for the very *large* amplitude motions involved in the denaturation. In our model, the Morse

potential is an effective potential which links the two strands. It results from a combination of an attractive part due to the hydrogen bonds between two bases in a pair and the repulsive interaction between the charged phosphate groups on the two strands. The potential for the hydrogen bonds can be rather well estimated [14] but the repulsive part is harder to determine because the repulsion is partly screened by ions of the solvent. Consequently we had to rely on estimations. The parameters that we use have been chosen to give realistic properties for the model in terms of vibrational frequencies, size of the open regions, etc., but future work will be needed to confirm our choice. We do not expect however that a better choice would change *qualitatively* the results presented here. The parameters that we have chosen are: a dissociation energy $D = 0.04$ eV, a spatial scale factor of the Morse potential $a = 4.45$ Å$^{-1}$, a coupling constant $K = 0.06$ eV/Å, a mass $m = 300$ a.m.u. The constant of the Nose thermostat has been set to $M = 1000$.

A first scan of the dynamics of the model is obtained by imposing a slow temperature ramp (200–540 K) that generates sets of states which are approximately equilibrated and are used then as initial states for simulations at constant temperature. Figure 2 shows a time evolution of the dynamics of the model at three temperatures. The stretching of the base-pairs is indicated by a grey scale, darker dots corresponding to larger stretching. Looking at this figure, one notices immediately two major features. First, as one moves along an horizontal direction, i.e., along the molecule for a given time, the amplitude of the stretching varies very much from site to site. This is especially true at high temperature, but it is still noticeable at 150 K, well below the melting temperature. This shows that there is no equipartition of energy in this nonlinear system, but on the contrary a tendency for the energy to localize at some points which is more and more pronounced as temperature increases. At high temperature, the figure shows large black regions which correspond to denaturated regions of the

molecule. These black areas are the denaturation bubbles observed experimentally. At the highest temperature shown here (fig. 2c) they extend over 20 to 50 base pairs and their boundaries are sharp.

If the temperature is raised slightly above 540 K, the bubbles grow even more and finally extend over the whole chain: the molecule is completely denaturated. The second remarkable feature on fig. 2 can be observed by moving along a vertical line on the figure, i.e., following the time evolution of a given base pair. If one choses one region of the molecule in which the energy is concentrated, one can see alternating black and light-grey dots. This is due to an *internal breathing* of the localized excitations that oscillate between a large amplitude (black dots in the figure) and a small amplitude state (light dots) in a regular manner. These motions are the fluctuational openings of DNA. They exist even well below the denaturation temperature and coexist with denaturated bubbles in the high temperature range. Figure 2b shows that they play the role of precursor motions for the formation of the bubbles.

The calculation of the dynamical structure factor from the molecular dynamics results exhibits two types of excitations. In the high frequency range, one recognizes the phonon modes corresponding to linear motions of the chain. At low temperature their dispersion curve is well described by the linear dispersion curve resulting from the equations of motions of the model. Close to melting, on the contrary, most of the chain is on the plateau of the Morse potential and therefore experiences almost no restoring force that brings it back to $y = 0$. The dispersion curve is then the dispersion relation of a chain of harmonically coupled particles, without a substrate potential, i.e., a dispersion relation without gap. The variation versus temperature of the frequency of main phonon peak at wavevector $q = \frac{1}{2}\pi$ plotted on fig. 3 shows clearly the transition between a frequency belonging to the original dispersion curve at low $T$ toward that of a gapless dispersion curve. This phonon softening

(a)     (b)     (c)



Fig. 2. Results of molecular dynamics simulations at three different temperatures (a) $T = 150$ K. (b) $T = 340$ K. (c) $T = 450$ K. The horizontal axis indicates the position along the 256 cells of the molecule and the vertical axis indicates time. The stretching $y_n$ of the base-pairs along the molecule is indicated by a grey scale, the lighter grey corresponding to $y \leq -0.1$ Å and black indicating $y \geq 1$ Å. Therefore black regions show broken base-pairs.

should be observable experimentally in the vicinity of DNA melting transition. The second characteristic feature of the dynamical structure factor is a low frequency peak, associated to the fluctuational opening, which shifts to zero frequency as the denaturation bubbles form near the melting point.

Since the molecular dynamics simulations have found the two types of motions observed experimentally, fluctuational openings and bub-
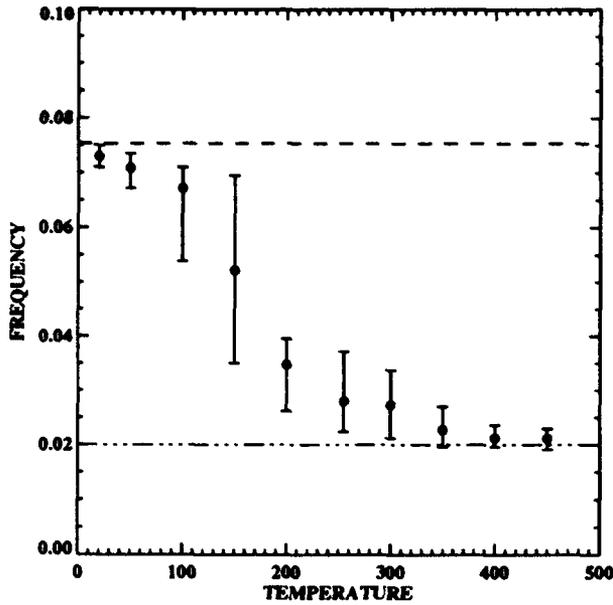
Fig. 3. Frequency of the phonon modes versus temperature for $q = \frac{1}{2}\pi$. The error bars indicate an interval where the frequency lies, at those temperatures for which a single mode cannot be identified. The horizontal lines correspond to the phonon frequency at the bottom of the well (dashed) and on the Morse plateau (dash-dot-dot-dot).

bles, it is interesting to see whether an analysis in terms of nonlinear excitations can explain these motions.

The fluctuational openings correspond to large amplitude breathing modes which are localized by nonlinear effects. As they extend only over a few base pairs, they are intrinsically discrete. Their existence and long term stability poses the general question of the existence of breathers in discrete Klein–Gordon models which has already attracted a great deal of attention without receiving a definite answer [15]. An analytical investigation of the DNA model presented above is difficult due to the Morse potential, but, since the fluctuational openings are intermediate amplitude oscillations, their study can be performed with a simpler potential that has qualitatively the shape of the Morse potential for small and intermediate amplitude but is more suitable analytically. The potential

$$V(y_n) = D(\tfrac{1}{2} y_n^2 - \alpha \cdot \tfrac{1}{3} y_n^3) ,$$  (5)

for $y_n \leq 1/\alpha$ is convenient. Introducing the variable $w_n = \alpha y_n$ and the scaled time $\tau = \sqrt{K/m}\, t$, and defining the parameter $\omega_d^2 = D/K$, which measures the discreteness, the Hamiltonian of this simplified models becomes

$$H = \frac{K}{\alpha^2} \sum_n [\tfrac{1}{2} \dot{w}_n^2 + \tfrac{1}{2}(w_n - w_{n-1})^2$$
$$+ \omega_d^2 (\tfrac{1}{2} w_n^2 - \tfrac{1}{3} w_n^3)] .$$  (6)

Large amplitude localized breathers can be obtained for this model [16] by the lattice-Green functions method [17] developed recently to study local modes in nonlinear lattices. We seek long-lived oscillatory solutions under the form

$$w_n = \sum_{i=0}^{\infty} \phi_n^i \cos(i\omega_b t)$$  (7)

where $\omega_b$ is the eigenfrequency of the fundamental mode ($i = 1$) and $\phi_n^i$ is the time independent amplitude of the $i$th mode (the ansatz has a dc part simply because of the asymmetry of the potential). Inserting this ansatz in the equations of motion and setting the coefficients of $\cos(i\omega_b t)$ equal to one another and retaining only the first three terms, which gives already a good approximation, we obtain:

$$\omega_d^2 \phi_n^0 - (\phi_{n+1}^0 + \phi_{n-1}^0 - 2\phi_n^0)$$
$$= \omega_d^2 [\phi_n^{0^2} + \tfrac{1}{2}(\phi_n^{1^2} + \phi_n^{2^2})] ,$$  (8)

$$(\omega_d^2 - \omega_b^2)\phi_n^1 - (\phi_{n+1}^1 + \phi_{n-1}^1 - 2\phi_n^1)$$
$$= \omega_d^2 (2\phi_n^0 + \phi_n^2)\phi_n^1 ,$$  (9)

$$(\omega_d^2 - 4\omega_b^2)\phi_n^2 - (\phi_{n+1}^2 + \phi_{n-1}^2 - 2\phi_n^2)$$
$$= \omega_d^2 (2\phi_n^0 \phi_n^2 + \tfrac{1}{2}\phi_n^{1^2}) .$$  (10)

Using the lattice Green's function of the linear part of the above equations,

$$G(n - m, \omega_b) = \omega_d^2 \frac{1}{N} \sum_q \frac{e^{iq(n-m)}}{\omega^2(q) - \omega_b^2} ,$$  (11)

where $q$ is the wave vector inside the first Brillouin zone, $\phi_n^0$, $\phi_n^1$, $\phi_n^2$ can be expressed in terms of the r.h.s. of eqs. (8)–(10), which gives a set of simultaneous nonlinear eigenvalue equations determining the eigenfrequency $\omega_b$ and the eigenfunctions $\phi_n^i$. They can be solved numerically by iterations, starting with appropriate initial conditions. We fix the dc term, which amounts to choosing a particular amplitude for the solution. Only 15 iterations are necessary to determine the values with an accuracy of $10^{-4}$. A numerical simulation of the dynamics of the chain with the solution determined as above as an initial condition shows that a large amplitude breather localized on very few lattice sites is extremely long lived, in agreement with the results of the molecular dynamics simulations of our DNA model. Consequently the very narrow fluctuational openings observed experimentally in DNA [8] and in our molecular dynamics investigations could well be discrete breathers stabilized by nonlinearity.

The study of the denaturation bubbles turns out to be much more difficult because they cannot be studied independently of the thermal effects. Due to the shape of the Morse potential, at $T = 0$, any large amplitude opening bringing a set of neighboring base-pairs on the plateau of the Morse potential is unstable. This initial state would oscillate at very low frequency, but cannot stay in an open state since the bases in the open region are called back to their closing state by the bases which are still closed in the molecule and by the small downward slope of the Morse plateau. Therefore the existence of long lived open bubbles in our model, as they are seen in the molecular dynamics simulations, is fundamentally a thermal effect which cannot be studied by Hamiltonian mechanics, as we did for the fluctuational openings. One way around this difficulty is to *include temperature effects in the potential itself*. This is exactly what the self-consistent phonon (SCP) method does using a trial Hamiltonian for the calculation of the free energy of the system at temperature $T$ [18].

Introducing $u_n = y_n - \langle y \rangle = y_n - \eta$ and two parameters $\Omega^2$ and $\phi$, we apply the SCP method [18] by considering the trial harmonic Hamiltonian

$$H_o = \sum_n [\tfrac{1}{2}m\dot{u}_n^2 + \tfrac{1}{2}\phi(u_n - u_{n+1})^2 + \tfrac{1}{2}\Omega^2 u_n^2] .$$
(12)

The canonical partition function $\mathcal{Z}$ can be expressed as the product of the unperturbed partition function $\mathcal{Z}_0$ and a perturbation factor $\mathcal{Z}_1$:

$$\mathcal{Z} = \int \prod_i du_i \, e^{-\beta V}$$
$$= \left( \int \prod_i du_i \, e^{-\beta H_o} \right) \langle e^{-\beta(H-H_o)} \rangle_0$$
$$= \mathcal{Z}_0 \mathcal{Z}_1 .$$

The perturbation series for $\mathcal{Z}_1$ is calculated by expanding the exponential, and the logarithmic function; the coefficient of $(-\beta)^n/n!$ in the expansion of $\ln \mathcal{Z}_1$ is called the $n$th cumulant and is written [19] $\langle (H - H_o)^n \rangle_{0,c}$. Thus

$$\mathcal{F} = -k_B T \ln \mathcal{Z}$$
$$= -k_B T \ln \mathcal{Z}_0 - k_B T \ln \langle e^{-\beta(H-H_o)} \rangle_0$$
$$= \mathcal{F}_0 - k_B T \sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \langle (H - H_o)^n \rangle_{0,c}$$
$$= \mathcal{F}_0 + \mathcal{F}_1 + \mathcal{F}_2 + \cdots .$$

The first contribution $\mathcal{F}_0$ is the contribution of $N$ harmonic oscillators,

$$\mathcal{F}_0 = -\tfrac{1}{2}k_B T \sum_{q=0}^{N-1} \ln \frac{2\pi k_B T}{\omega^2(q)} .$$
(13)

It can be shown [20] that the variational free energy $\mathcal{F} \le \mathcal{F}_0 + \mathcal{F}_1 = \mathcal{F}_v$ gives an upper bound for the actual free energy. The only difficulty in evaluating $\mathcal{F}_v$ is the self-consistent substrate potential. It can be evaluated and gives an "effective" potential that keeps the shape of the Morse potential, but with a minimum which is

temperature dependent. The expression for the first order correction for the free energy is then:

$$\mathcal{F}_1 = (K - \phi)(\langle u^2 \rangle - \langle v^2 \rangle) - \tfrac{1}{2}\Omega^2 \langle u^2 \rangle$$

$$+ D(1 + e^{-2a\eta + 2a^2\langle u^2 \rangle} - 2\,e^{-a\eta + (a^2/2)\langle u^2 \rangle})\,,$$

$$(14)$$

with $\langle u^2 \rangle = \langle u_n^2 \rangle$ and $\langle v^2 \rangle = \langle u_n u_{n+1} \rangle$. Considering $\eta$, $\langle u^2 \rangle$ and $\langle v^2 \rangle$ as *variational* parameters, the conditions for $\mathcal{F}_v$ to be stationary give $\eta = \tfrac{3}{2}a\langle u^2 \rangle$, $\phi = K$ and $\Omega^2 = 2a^2 D\exp(-\tfrac{2}{3}a\eta)$. As $\eta$ is an increasing function of temperature, $\Omega^2$ decreases with $T$, which corresponds to the mode softening observed in the molecular dynamics simulations. Since $\langle u^2 \rangle$ and $\eta$ are related, the minimization of $\mathcal{F}_1$ amounts simply to solving the equation

$$\eta = \frac{3ak_B T}{2N} \sum_p [2a^2 D\,e^{-2a\eta/3}$$

$$+ 4K\sin^2(\pi p/N)]^{-1}\,.$$

$$(15)$$

In practice, eq. (15) is solved by a simple bisection method and fig. 4 shows the free energy $\mathcal{F}_v$ versus $\eta$ at different temperatures. We see
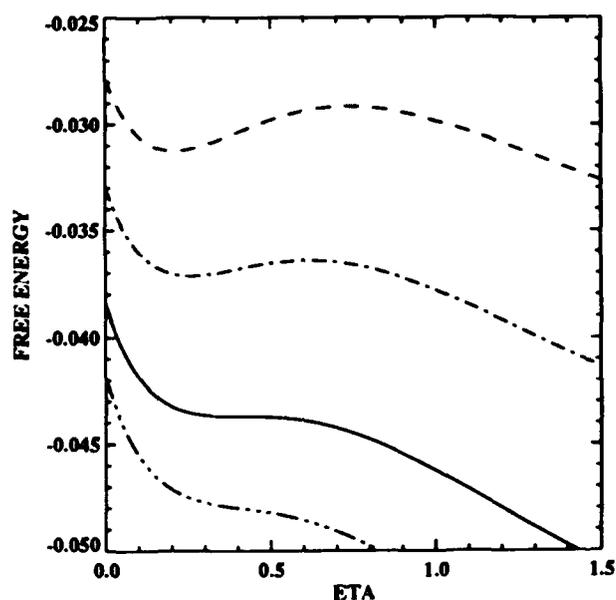


Fig. 4. Free energy $\mathcal{F}_v$ versus $\eta$ at different temperatures: $T = 350$ (dashed), 380 (dash-dotted), 411 (solid), 430 (dash-dot-dot-dot) K.

clearly that the self-consistent solution, which corresponds to the metastable minimum of the dashed and dash-dotted curves, disappears at $T_c = 411$ K to give a strictly decreasing function of $\eta$. Over $T_c$ the only minimum is obtained for an infinite value of $\eta$, which does not correspond to a self-consistent solution of the problem. Consequently $T_c$ can be identified as the denaturation temperature given by the SCP. Moreover, assuming that the free energy can play the role of an effective potential at temperature $T$, one can see that, as $T$ approaches the melting temperature, the free energy exhibits a maximum followed by a decreasing part. In a "mechanical" equation of motion, such a potential would give stable open bubbles.

## 4. Is DNA melting a one-dimensional phase transition?

The model discussed above has been able to describe some of the main features of DNA melting as it is observed experimentally. However there is a crucial point in which this model gives incorrect results, it is the *sharpness* of the phase transition. For an homopolymer, the experiments show that the melting occurs very abruptly over a temperature interval which is only a few K or even less. This poses a very fundamental question since DNA is basically a one-dimensional system, which is not expected to have a phase transition. We would like to conclude our paper by showing that, *within a one-dimensional model with short range interactions*, a sharp transition is possible if one takes into account properly the nonlinearity of the base stacking interaction [21]. The possibility of a phase transition in one-dimensional DNA was already examined within the Ising-model approach by Poland and Scheraga [22] and Azbel [23] who concluded that it can be attributed to cooperativity effects and to the role of the winding entropy released when the two strands separate. A simple extension of our DNA model (1)

can describe the dynamics of these effects and gives a very sharp transition in agreement with the experiments.

The stacking energy between two neighboring base pairs is described by the anharmonic potential:

$$W(y_n, y_{n-1}) = \tfrac{1}{2}K(1 + \rho\,e^{-\alpha(y_n + y_{n-1})})$$

$$\times (y_n - y_{n-1})^2 . \tag{16}$$

This new intersite coupling, replacing the simple harmonic coupling of our previous approach, is responsible for qualitatively different properties. The choice of this potential has been motivated by the observation that the stacking energy is not a property of individual bases, but a character of the base pairs themselves [24]. When the hydrogen bonds connecting the bases break, the electronic distribution on bases are modified, causing the stacking interaction with adjacent bases to decrease. In eq. (16), this effect is enforced by the prefactor of the usual quadratic term ($y_n - y_{n-1})^2$. This prefactor depends on the sum of the stretchings of the two interacting base-pairs and decreases from $\tfrac{1}{2}K(1 + \rho)$ to $\tfrac{1}{2}K$ when either one (or both) base-pair is stretched. Although its form was chosen for analytical convenience, the qualitative features of potential (16) are in agreement with the properties of chemical bonds in DNA. They also provide the cooperativity effects that were introduced phenomenologically in the Ising models. A base pair that is in the vicinity of an open site has lower vibrational frequencies, which reduces its contribution to the free energy. Simultaneously a lower coupling along the strands gives the bases more freedom to move independently from each other, causing an entropy increase which drives a sharp transition. Our approach can be compared to recent views on structural phase transition in elastic media which stress that intrinsic nonlinear features characterize the physics of these transformations, and extend the standard soft mode picture [25,26]. It is important to notice that,

although cooperativity in introduced through purely nearest neighbor coupling terms, it has a remarkable effect on the 1D transition.

Figure 5 shows the drastic change introduced by the anharmonic coupling on the specific heat of the model calculated by the transfer integral method [21]. The full curve corresponding to the anharmonic stacking interaction shows a sharp peak very similar to that one would expect from a first-order phase transition, whereas the harmonic coupling investigated before gives only a smooth maximum. This result suggests that, although DNA structure is very complicated, a simple nonlinear model is able to reproduce with a good agreement the main experimental features of its dynamics for the fluctuational openings as well as the melting curves. Indeed such a model does not pretend to give exact quantitative results that would fit exactly the experimental melting curve of an heteropolymer with all its small structure. Our feeling is that it is more important to get a basic understanding than to
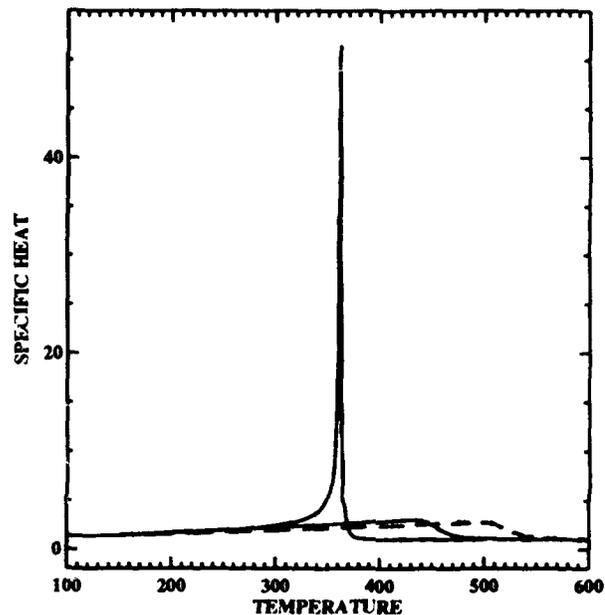


Fig. 5. Variation of the specific heat versus temperature. The very narrow peak corresponds to the anharmonic coupling case ($\alpha = 0.35$, $\rho = 0.5$), the dotted curve and the solid broad peak to harmonic coupling ($k' = 1.5k$ and $k'' = k$, respectively).

try to explain the very small experimental result. This does not mean that we think that the simple model that we discussed here is complete and should not be extended. But any extension will have to be measured for the new fundamental feature it brings against the unnecessary complications it introduces. Finally we would like to point out that, although we have discussed nonlinear dynamics of DNA, we have not introduced solitons in our picture (although the discrete breathers are probably close to being solitons). It is simply because they do not seem to be necessary to explain the dynamical features of DNA we are interested in. Solitons are marvellous objects but nonlinear science can also live without them.

## Acknowledgements

## References

[1] J. Ladik and J. Cizek, Probable physical mechanisms of the activation of oncogenes through carcinogens, Int. J. Quantum Chem. 26 (1984) 67.

[2] R.M. Wartell and A.S. Benight, Thermal denaturation of DNA molecules: a comparison of theory with experiment, Phys. Rep. 126 (1985) 67.

[3] J.A. MacCammon and S.C. Harvey, in: Dynamics of Proteins and Nucleic Acids (Cambridge Univ. Press, Cambridge, 1988).

[4] M. Peyrard and A. Bishop, Statistical mechanics of a

[5] D.J. Scalapino, M. Sears and R.A. Ferrel, Statistical mechanics of one-dimensional Ginzburg–Landau fields, Phys. Rev. B 6 (1972) 3409.

[6] J.A. Krumhansl and J.R. Schrieffer, Dynamics and statistical mechanics of a one-dimensional model Hamiltonian for structural phase transitions, Phys. Rev. B 11 (1975) 3535.

[7] J.F. Currie, J.A. Krumhansl, A.R. Bishop and S.E. Trullinger, Statistical mechanics of one-dimensional solitary-wave-bearing scalar fields: exact results and ideal-gas phenomenology, Phys. Rev. B 22 (1980) 477.

[8] J.L. Leroy, M. Kochoyan, T. Huynh-Dinh and M. Gueron, Characterization of base-pair opening in deoxynucleotide duplex using catalyzed exchange of the imino proton, Mol. Biol. 200 (1988) 223.

[9] T. Schneider and E. Stoll, Classical statistical mechanics of the sine-Gordon and $\phi^4$ chains. Static properties, Phys. Rev. B 22 (1980) 5317.

[10] S. Nose, A molecular dynamics method for simulations in the canonical ensemble, Mol. Phys. 52 (1984) 255.

[11] S. Nose, A unified formulation of the constant temperature molecular dynamics methods, J. Chem. Phys. 81 (1984) 511.

[12] N. Theodorakopulos and M. Peyrard, unpublished.

[13] L.L. Van Zandt, DNA solitons with realistic parameters, Phys. Rev. A 40 (1989) 6134;
M.A Techera, L.L. Daemen and E.W. Prohofsky, Comment on "DNA solitons with realistic parameters", Phys. Rev. A 42 (1990) 5033;
L.L. Van Zandt, Reply to "Comment on DNA solitons with realistic parameters", Phys. Rev. A 42 (1990) 5036, and references therein.

[14] Y. Gao, K.V. Devi Prasad and E.W. Prohofsky, A self-consistent microscopic theory of hydrogen bond melting with application to poly(dG)–poly(dC), J. Chem. Phys. 80 (1984) 6291.

[15] D.K. Campbell and M. Peyrard, Chaos and order in nonintegrable field theories, in: Chaos, eds. D.K. Campbell (AIP, New York, 1990), and references therein.

[16] T. Dauxois, M. Peyrard and C.R. Willis, Localized breather-like solution in a discrete Klein–Gordon model and application to DNA, Physica D 57 (1992) 267.

[17] S. Takeno, K. Kisoda and A.J. Sievers, Intrinsic localized vibrational modes in anharmonic crystals, Prog. Theor. Phys. Suppl. 94 (1988) 242.

[18] N. Bocarra and G. Sarma, Théorie microscopique des transitions s'accompagnant d'une modification de la structure cristalline, Physics (Long Island City) 1 (1965) 219.

[19] R. Kubo, The fluctuation–dissipation theorem, J. Phys. Soc. Japan 17 (1962) 1100.

[20] R.P. Feynman, Statistical Mechanics (Benjamin, 1972).

[21] T. Dauxois, M. Peyrard and A.R. Bishop, Entropy driven DNA denaturation, Phys. Rev. E 47 (1993) R44.

[22] D. Poland and H. Scheraga, Phase transitions in one-

dimension and the helix-coil transition in polyamino acids, J. Chem. Phys. 45 (1966) 1456.

[23] M. Ya. Azbel, Generalized one-dimensional Ising model for polymer thermodynamics, J. Chem. Phys. 62 (1975) 3635.

[24] C. Reiss, private communication (1991).

[25] J.A. Krumhansl and D.M. Alexander, in: Nucleic Acids and Proteins, eds. E. Clementi and R.H. Sarma (Academic Press, 1983).

[26] W.C. Kerr, A.M. Hawthorne, R.J. Gooding, A.R. Bishop and J.A. Krumhansl, First-order displacive structural phase transition studied by computer simulation, Phys. Rev. B 45 (1992) 7036.

PHYSICA D

# Light interacting with liquid crystals

David W. McLaughlin [a,1] , David J. Muraki [a,2] and Michael J. Shelley [b,3]

[a] *Department of Mathematics, Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA*

[b] *The Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA*

In this paper we describe laser light interacting with nematic liquid crystals. The paper begins with a summary of recent experimental results of E. Braun, L. Faucheux, and A. Libchaber in which the liquid crystal sample is studied in three geometries – film, pipe, and droplet. Then, after a very brief glimpse at the history of liquid crystals, a theoretical model of the interacting system is described. In a one transverse dimensional idealization, we investigate the pipe and film configurations. In these cases the model reduces to a coupled system of nonlinear pde's – an *elliptic sine-Gordon* equation for the director field coupled to a Schroedinger equation for the electromagnetic field. Properties and qualitative behavior of this coupled system are described, both numerically and theoretically. As an illustrative example of boundary layer analysis of such coupled light–nematic systems, we describe calculations in the film geometry in some detail. Results of this analysis include: (i) an extension of the Frederiks bifurcation analysis to electric fields with spatial variation; (ii) the determination of the transverse scale at which self-focusing saturates in this nematic; (iii) the derivation of a nonlocal nonlinear Schroedinger equation which governs the inner structure of the laser beam. We conclude the paper with a summary of similar boundary layer calculations for light–nematic systems in other geometries.

## 1. Introduction

Liquid crystals provide an optical medium with *very strong* nonlinearity. For example, the coefficient of nonlinearity in liquid crystals can easily be $10^6$–$10^{10}$ times greater than in "typical" nonlinear optical media such as carbon bisulfide ($CS_2$). The extreme nonlinearity of liquid crystals permits the investigation of nonlinear optical effects such as self-focusing and filamentation with low power cw lasers. For this reason liquid crystals are excellent materials for the investigation of laser radiation interacting nonlinearly with matter, and thus they provide an optical arena for the study of the physics of complex nonlinear patterns.

In recent experiments of E. Braun, A. Libchaber, and L. Faucheux [1], nematic liquid crystals (MBBA, 6CB, and E209) were irradiated with light from a cw argon laser with 1 watt of power. A linearly polarized argon laser beam in the $TEM_{00}$ gaussian mode (the pump beam) was applied to the sample of liquid crystal which sat on the stage of an inverted microscope. The pump beam had a waist of $\simeq 50\ \mu$m at the bottom plate of the sample. Thermal effects were negligible, while scattering losses were more substantial but controlled.

A detailed description of the experimental results may be found in [2,1]. Here we restrict ourselves to a few remarks about them, focusing attention upon two representative experiments. Studies were performed in three geometries – film, droplet, and capillary tube. Self-focusing effects were observed in all three geometries. In

Fig. 1. Self-focusing and filamentation in a droplet config-
uration of the nematic. Note the boundary of the droplet
is observable. Also note that polarization effects produce
"ordinary" and "extraordinary" rays. The latter coherently
self-focus.



Fig. 2. Longitudinal view of self-focused filaments in the
capillary tube configuration of the nematic.

the thin film case, the observations were made
outside the liquid crystal "after" the light beam
had passed through the thin sample; thus, self-
focusing was observed only indirectly. However,
in the droplet and tube cases, filamentation was
directly observed. Figure 1 shows a beautiful
example of focusing and filamentation in the
droplet geometry.

In order to achieve better control of thermal
effects, Braun, et al. [2,1] changed to the geom-
etry of a cylindrical capillary tube, together with
a different nematic liquid crystal, E209, which
offers a wide range in temperature in the ne-
matic phase. A cylindrical capillary tube, of in-

side diameter $\simeq$ 1.5 mm, was coated with the
polymer MAP to enforce tangential alignment of
the director at the glass surface, and thus to en-
sure uniform bounday conditions along the cap-
illary tube. The liquid crystal was then injected
in the nematic phase. One end of the capillary
was sealed with fast epoxy to avoid any motion
of the nematic. Cooling was achieved by a flow
of nitrogene around the capillary. The strongest
lens in this system is a diverging one due to the
meniscus at the air–liquid crystal interface. With
this setup, the argon laser beam was focused in-
side the sample, far from the entrance meniscus,
and the scattered light was observed from below.

Figure 2 shows a striking longitudinal view
of the light field along the capillary tube. Many

other observations, for several increasing values of power in the pump beam, can be found in [2,1]. There one observes the presence of a focal spot, and, as the pump intensity is increased, its "movement" along the tube toward the meniscus at the front entry. As the intensity is further increased, this movement stops and a second spot forms and moves toward the front of the tube. Three coexisting focal spots have been observed in the experiments [2,1]. Also one observes the onset of transverse undulations of the filament. The wavelength of these undulations at first decreases, and then increases with increasing pump intensity [1]. Finally, at still higher values of pump intensity, two filaments form and interact. As far as we know, these are the first observation of the actual longitudinal behavior of filamentation of a low power cw laser beam in a nonlinear medium. In any case, it is certainly a striking visualization.

## 2. Brief overview of the history of liquid crystals

The twelfth International Liquid Crystal Conference was held in Germany in 1988. One aspect of that conference was the celebration of the one hundredth anniversary of the "discovery" of liquid crystals. At that conference, H. Kelker described that history. Our account is taken from the second chapter of a recent book by Peter Collings [3].

During the time period from 1850–1888, just before the "discovery" of liquid crystals, several European scientists were observing striking visual effects such as cloudiness and color changes which would later be attributed to the liquid crystal phase of their samples. For the most part these scientists were biologists who were investigating under microscopes the effect of temperature on different biological materials including the outer covering of nerve fibers, natural fats, and cholesterol. An Austrian botanist, Friedrich Reinitzer, is generally credited with the actual discovery of liquid crystals in his studies of

cholesterol in plants. In 1888 he was observing the melting behavior of an organic substance related to cholesterol and noted that it possessed two melting points – at 145.5°C it melted to a cloudy liquid and at 178.5°C this cloudy liquid turned clear. He also observed the appearence of blue colors at the phase transitions. The substance that Reinitzer was observing was in fact cholesteryl benzoate, a chiral nematic liquid crystal.

Reinitzer sent some of his samples to a German physicist, Otto Lehmann, who had just developed an important microscope. Lehmann's microscope was a significant technical advancement for that time period in that it had a heating stage that allowed him to observe crystallization as temperature was slowly lowered, as well as polarizers which allowed him to view polarization phenomena. Through his observations of Reinitzer's samples, Lehmann became convinced that the cloudy liquid was a liquid phase, but a phase that affected polarized light in a manner typical of solid crystals. The substance had the flow properties of a liquid, but the optical properties of a solid, and led Lehmann to call such substances "liquid crystals".

As described by Collings [3], European chemists, biologists, and physicists continued to study liquid crystals, both experimentally and theoretically, during the first half of the twentieth century. The theory culminated with the continuum elastic theory of English physicist F. C. Frank in the 1950's. Little was done in liquid crystal research from the end of World War II through the early 1960's, perhaps because no important applications had emerged. The discovery of liquid crystal displays (LCD's) at the end of the 1960's certainly changed the status of research in the area! Today liquid crystal physics is alive, well, and dynamic – with a great many scientists actively studying nonlinear properties of this liquid crystal phase of matter.

While this "history of liquid crystals" is certainly interesting, we elected to describe it in such detail in our paper for another reason. Our

conference here at the Technical University of Denmark has as one of its goals the celebration of the sixtieth birthday of Professor Alwyn C. Scott. As we learned about the history of liquid crystals, we were struck by the similarities between the development of the science of liquid crystals and the manner with which Al does, and advocates doing, science: – The development of liquid crystals was *interdisciplinary* – between the biological and physical sciences; *international*; *nonlinear*; and driven by experimental, theoretical, and practical developments. A brief consideration of such historical perspectives certainly seems appropriate at our conference on Future Directions of Nonlinear Dynamics in Physical and Biological Systems.

We conclude this historical section with two additional references: the book of de Gennes [4], and the survey [5].

## 3. Theoretical considerations

We are interested in a coupled field description of the interaction of laser light with a nematic liquid crystal. A nematic liquid crystal may be thought of as consisting of cigar shaped "objects" which are either polarized or easily polarizable. These objects have both translational and orientational degrees of freedom. In the liquid crystal phase, the material behaves as a fluid with respect to its translational degrees of freedom; however, with respect to its orientational degrees of freedom, long range order is maintained as if the material were in a crystalline phase. In this work we will ignore fluid motion and concentrate upon the orientational behavior of the liquid crystal.

This orientation is described by a *director field*,

$$n(x, t), \qquad |n| = 1.$$

This unit vector $n(x, t)$ describes the density of "rods", at position $x$ at time $t$, oriented in the direction $n$.
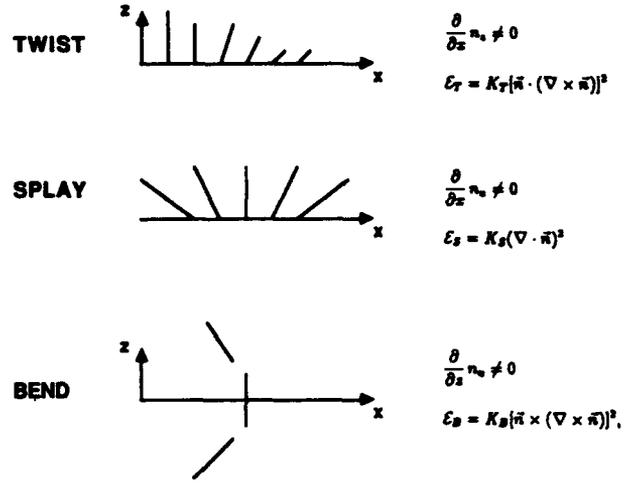


Fig. 3. Schematic of the "twist", "splay", and "bend" elastic distortions of the nematic.

Rotation of these rods is opposed by elastic, nearest neighbor, forces. Three independent orientational distortions exist for the vector field $n(x, t)$. These are called "twist", "splay", and "bend"; they are displayed in the cartoons of Fig. 3 [6]; the energy densities associated with each are given by

$$\mathcal{E}_T = K_T [n \cdot (\nabla \times n)]^2,$$

$$\mathcal{E}_S = K_S (\nabla \cdot n)^2,$$

$$\mathcal{E}_B = K_B [n \times (\nabla \times n)]^2,$$

where $K_T$, $K_S$, and $K_B$ are the elastic constants associated with twist, splay, and bend respectively. The total elastic energy is then given by

$$E_{el} = \tfrac{1}{2} \int (\mathcal{E}_T + \mathcal{E}_S + \mathcal{E}_B) \, d^3x.$$

Infinitesimal variations (subject to the unit director constraint) of this elastic energy produce an elliptic operator which, in the "one constant" approximation ($K \equiv K_S = K_B = K_T$) reduces to the Laplacian. (Typical nematic liquid crystals have $K_S < K_B \sim K_T$.)

We seek to describe the interaction of this director field with laser light. First, note that the

director $n$ is very sluggish in $t$ and can only respond to the time averaged electromagnetic field $E$. A phenomenological description of this interaction is as follows: The liquid crystal has its director field $n$ specified on the boundary of the sample. When the laser shines on this sample, energy minimization favors the alignment of $n$ with $E$. During this alignment procedure, $E$ induces a dipolar change in the index of refraction. The index of refraction is maximal when $n$ is parallel to $E$, producing a "self-focusing" lens. The reason behind the large values of the nonlinear dielectric constant in liquid crystals is that it is relatively easy to rotate the director field.

The coupled equations for the electric field interacting with the director field are given by

$$-\nabla \times \nabla \times E + k_0^2 (E + \alpha(n \cdot E)n) = 0,$$
$$K\left(n \times \nabla^2 n + g\Re(n \times (n \cdot E)E^*)\right) = 0,$$

$$(3.1)$$

where $g$ denotes a coupling parameter. Here we note that $E$ is the time averaged complex electric field amplitude, and that the interaction in the coupled equations (3.1) is invariant to $n \to -n$.

It will be convenient to specialize to a simple two-dimensional geometry which can apply in either the "film" or the "tube" cases:

$$n = \begin{pmatrix} \sin\theta \\ 0 \\ \cos\theta \end{pmatrix}, \qquad (3.2)$$

$$E = \begin{pmatrix} F \\ 0 \\ H \end{pmatrix}, \qquad (3.3)$$

where the angle $\theta = \theta(x, z)$, and the components of the field $(F, H)$ also only depend upon $(x, z)$. In this two dimensional geometry, $z$ is the longitudinal coordinate; $x$ is the transverse coordinate; $F$ and $H$ are the transverse and longitudinal field components, respectively. If, at "leading order", one neglects the longitudinal field $H$,

the equation for the director takes the form of an *elliptic, variable coefficient, sine-gordon equation*:

$$\theta_{xx} + \theta_{zz} + |F|^2 \sin 2\theta. \qquad (3.4)$$

### 3.1. The Frederiks transition

The relevant nonlinear effect for our studies is known as the "Frederiks transition" and can be easily seen from eq. (3.4): One assumes an $F$ field which is constant in both $x$ and $z$, and seeks a director field $\theta$ which is a function of $z$ only. In this situation, one reduces (3.4) to the ode

$$\theta_{zz} + \lambda^{-2} \sin 2\theta = 0,$$
$$\theta(z = 0) = \theta(z = L) = 0.$$

Here we have introduced a parameter $\lambda$ known as the Frederiks transition length,

$$\lambda \equiv |F|^{-1},$$

and have introduced the boundary conditions on the director that $n = \hat{z}$ at $z = 0$ and $z = L$; that is, $\theta$ vanishes at the front and rear faces of the sample. $\theta = 0$ is one solution of this two-point boundary problem corresponding to alignment with the boundary conditions of the director throughout the sample. On the other hand, as the intensity of the constant $F$ field is increased, solutions of the two-point boundary experience a "pitchfork bifurcation" and a second solution appears with $\theta = \theta(z)$. This second solution represents a $z$ dependent rotation of the director field within the sample, away from its boundary orientation toward the orientation of the $\hat{x}$ component of the electric field. This bifurcation is the primary nonlinear effect behind our study. We will investigate it numerically and analytically, in the more general situation of coupled fields with both $x$ and $z$ dependence.

## 3.2. A model problem

By scaling all lengths on the Frederiks transition length of the incoming pump beam, the experimental value of the optical wavenumber becomes $k \simeq 189$. (Other natural scalings produce much larger nondimensional wave numbers.) With such a large wavenumber, the electric field is certain to obey a *slowly varying envelope approximation*. Here we study the simplest such (two dimensional) envelope model – one which incorporates a coupling between paraxial optics and nematic deformation:

$$2ikF_z + F_{xx} + k^2\alpha \sin^2\theta \, F = 0, \qquad (3.5)$$

$$\theta_{zz} + \theta_{xx} + |F|^2 \sin 2\theta \qquad = 0. \qquad (3.6)$$

Here the refractive anisotropy is represented by the parameter $\alpha$, which is approximately 0.25 for the experiments [2,1]. Depending upon boundary conditions, this model can be used to study either the thin film or the tube geometries. In the thin film case, the boundary conditions are

$$F(x, z = 0) = F_{\text{in}}(x)$$

$$\theta(x,0) \qquad = \theta(x,L) = 0,$$

$$\theta(x,z) \qquad \to 0 \quad \text{as } x \to \pm\infty. \qquad (3.7)$$

(In the tube geometry, the latter boundary condition is replaced by $\theta(\pm W, z) = 0$, where $W$ is the width (radius) of the tube.)

Heuristically, eq. (3.5) is a scalar paraxial equation that includes the effects of diffraction and nematic anisotropy, while eq. (3.6) models static nematic distortions. Of course, this two dimensional model neglects birefringence (polarization) effects and scattering losses – which are important for quantitative detail. However, this simple model has proven useful in the development and description of our mathematical approximations.

## 4. An initial numerical study

We begin with a brief summary of our initial numerical study of model (3.5)–(3.6). To set up the numerical algorithm we impose Dirichlet boundary conditions for the $\theta$ field in $z$, Neumann boundary conditions for the $\theta$ field in $x$, as well as periodic boundary conditions with respect to the transverse variable $x$ for the $F$ field. The transverse width models an extremely wide thin film, and in practice is chosen wide enough that it does not effect the results. (The choice of transverse Neumann boundary conditions has been made to improve this insensitivity.)

The parabolic equation for $F$ is integrated with an "integrating factor" method which exactly factors out the Laplacian from the evolution in $z$ by introducing analytically the factor $\exp[(iz/2k)\,\partial_{xx}]$. The elliptic equation for the $\theta$ field is integrated with a "relaxation method". Both the "integrating factor" and "relaxation" methods are implemented through "fast Fourier transforms", which are natural given the boundary conditions.

Sample results from these numerical experiments are displayed in Fig. 4. In these numerical experiments, $k = 100.00, \alpha = 0.10$, and the critical field for the "Frederiks transition" (below which there is no transition) is $|F_{\text{cr}}| \simeq 0.12$. In Fig. 4a, the maximum field intensity is $|F_{\text{max}}| = 0.20$, somewhat above the transition value; in Fig. 4b, the maximum field intensity is larger, $|F_{max}| = 0.50$, well above the transition value. Examination of these figures shows:

(i) The transverse profile of the electric field $|F|$ is much narrower than that of the nematic field $\theta$.

(ii) The contour lines of the nematic field $\theta$ show transition regions (in $z$) near the front and rear faces of the nematic film. These rather sharp transitions are particularly apparent at the larger input field of Fig. 4b.

(iii) As the electric field $F$ "propagates" in $z$, it "self-focuses" in that its intensity increases while its tranverse structure narrows. Again,
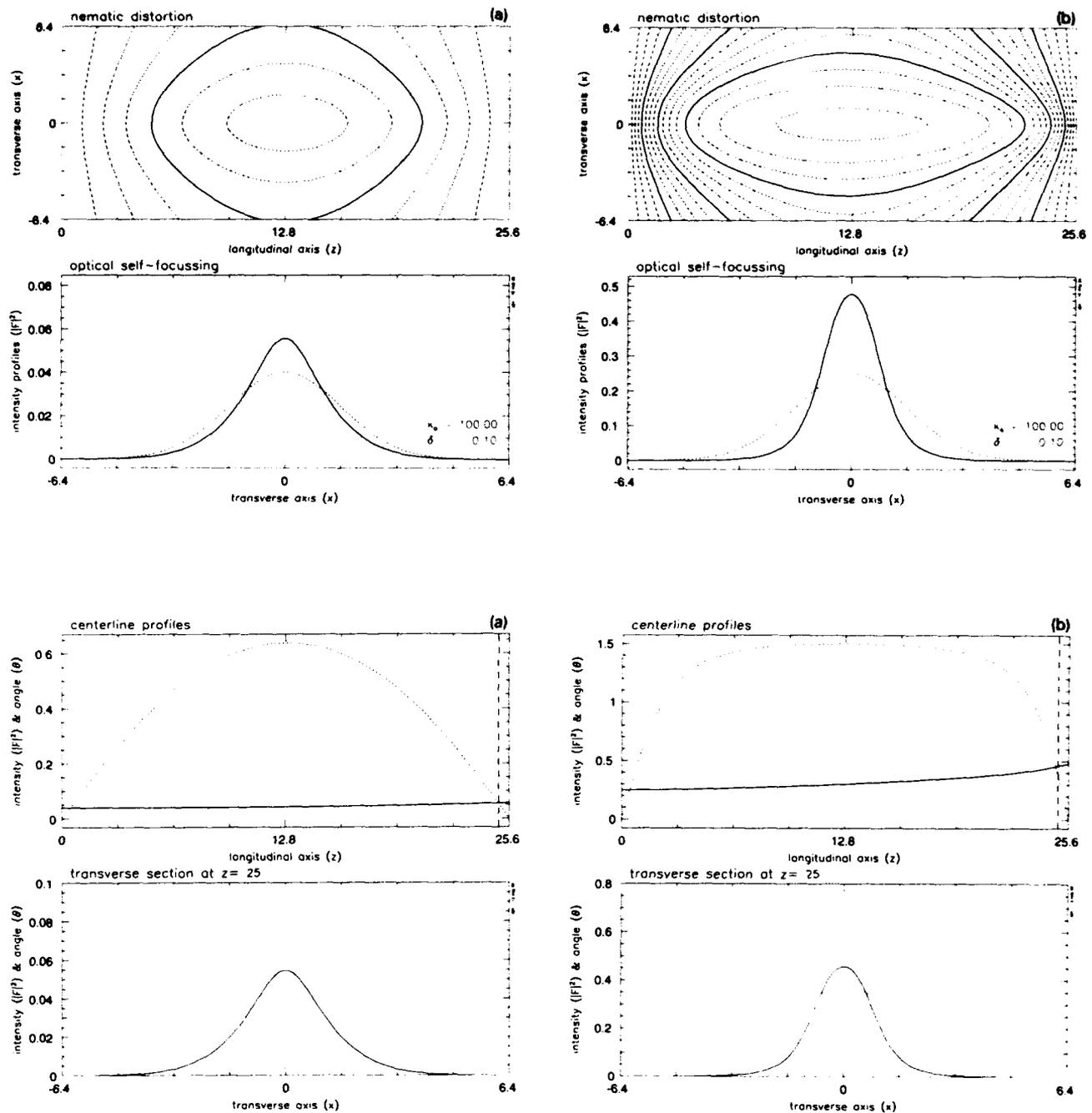
Fig. 4. Sample numerical experiments for the scalar model problem, for $\alpha = 0.10, k = 100.00$. At these parameter values the critical field for the Frederiks transition has $|F_{cr}|^2 \simeq (0.12)^2$. (a) Pump beam: $|F_{max}|^2 = (0.2)^2$; (b) Pump beam: $|F_{max}|^2 = (0.5)^2$. In each case, we display contour lines for the nematic angle $\theta$; the electric field intensity profiles $|F|^2$ vs $x$, at the front (dotted curve) and rear (solid curve) faces; the behavior of the electric field intensity (solid) and of the nematic (dotted) along the center beam line; and a tranverse profile near the rear face for the electric field intensity (solid) and nematic angle (dotted).
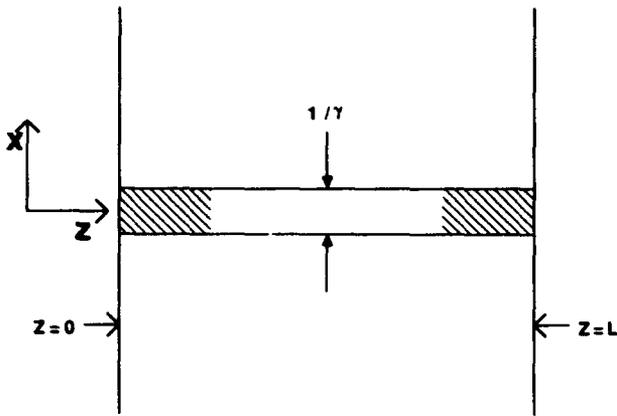
Fig. 5. A sketch of the boundary layer behavior for the thin film case.

this self-focusing is particularly apparent for the larger input field of Fig. 4b. In both cases a and b, the self-focusing continues to the rear face of the film. It is not clear from these experiments what, if anything in the absence of scattering losses, saturates this self-focusing process.

Clearly, for this simple model problem, these numerical discretization effects could be eliminated with a finer resolution of the numerical grid. Furthermore, one could also further stress the model by running the numerical experiments at more realistic parameter values – for example, $k$ could be increased from 100.00 to 189.00 and $\alpha$ from 0.1 to 0.25. However, the "boundary-layer–self-focusing" nature of the phenomena make this solely numerical approach unrealistic for the actual system. Thus, even for the model problem, it is judicious to develop a mathematical analysis based upon a boundary layer reduction.

## 5. A boundary layer analysis for the thin film case

With our thin film model (3.5), (3.6), (3.7), we illustrate a boundary layer approach for the light–nematic system. The geometry is depicted in Fig. 5. We emphasize that the boundary layer used in the analysis is a *transverse* boundary layer located at the beam center. While the nematic field $\theta$ has contour lines (see Fig. 4) reminiscent of longitudinal boundary layer behavior near the front and rear faces of the nematic film, this longitudinal transition scale is much longer than the characteristic width of the self-focused electric field $F$. In the absence of scattering losses, the self-focusing continually narrows the electric field until diffraction saturates this process at a very narrow beam. This transverse self-focusing is also apparent in the numerics of Fig. 4b. Indeed, it is this transverse effect of the electric field, rather than the longitudinal behavior of the nematic, that limits numerical resolution, and introduces the small width scale for our analysis.

Let $\gamma = \gamma(k)$ denote a large parameter which fixes the inner, transverse beam scale:

$$\bar{x} \equiv \gamma x, \quad \gamma \gg 1.$$

Here the scaling function $\gamma(k)$ will be determined later in the analysis; for the moment it is only taken to be large. Next we assume that in the *outer-relaxation* and in the *inner-beam* regions, the fields take the following form:

$$\theta \simeq \begin{cases} \theta^{(r)}(z,x) & \text{(outer)}, \\ \theta^{(r)}(z,0) + \gamma^{-1}\theta^{(b)}(z,\bar{x}) & \text{(inner)}, \end{cases} \quad (5.1)$$

$$F \simeq \begin{cases} 0 \text{ (exp. small)} & \text{(outer)}, \\ \gamma^{1/2} F^{(b)}(z,\bar{x}) & \text{(inner)}. \end{cases} \quad (5.2)$$

When viewed from the outer scale, the $|F|^2$ field appears as a delta function located at the beam center $x = 0$; in turn, from eq. (3.6), the $\theta$ field has a jump discontinuity in its first derivative $\theta_x(x = 0, z)$. Thus,

$$|F|^2 \simeq I\delta(x),$$

where

$$I \equiv \int_{-\infty}^{+\infty} |F|^2 \, dx = \int_{-\infty}^{+\infty} |F^{(b)}|^2 \, d\bar{x},$$

and where the scaling of $F^{(b)}$ insures the last equality.

In the two *outer-relaxation* regions ($x > 0$ and $x < 0$), the electric field is exponentially small and the nematic field satisfies Laplace's equation

$$\nabla^2\theta^{(r)} = 0,$$

together with vanishing boundary conditions at $z = 0, z = L, x = \pm\infty$ and the jump condition across $x = 0$. Thus, $\theta$ has the Fourier series representation

$$\theta^{(r)} = \sum_{n=1}^{\infty} \hat{c}_n \exp\left(-n\frac{\pi}{L}|x|\right) \sin\left(n\frac{\pi}{L}z\right),$$

where the constants $\hat{c}_n$ will be determined by the jump condition. Notice that because $\theta^{(r)}$ solves Laplace's equation, the transverse half-width of the nematic distortion is determined by the thickness $L$ of the thin film.

To derive the jump condition, one integrates eq. (3.6) across the beam zone, paying attention to scales and to the local behavior of the fields. The result is

$$\theta_x^{(r)}\Big|_{0^-}^{0^+} + I\sin 2\theta^{(r)}\Big|_{x=0} = 0.$$

Defining

$$C(z) \equiv \theta^{(r)}(z, x = 0) = \sum_{n=1}^{\infty} \hat{c}_n \sin\left(n\frac{\pi}{L}z\right),$$

one sees that the jump condition yields the expression

$$\frac{I}{2}\sin\left[2\sum_{n=1}^{\infty} \hat{c}_n \sin\left(n\frac{\pi}{L}z\right)\right]$$

$$= \sum_{n=1}^{\infty} \hat{c}_n \sin\left(n\frac{\pi}{L}z\right), \qquad (5.3)$$

which is a nonlinear equation which determines the constants $\hat{c}_n$ as a function of the beam intensity $I$; thus, it determines the nematic distortion at beam center as described by the function
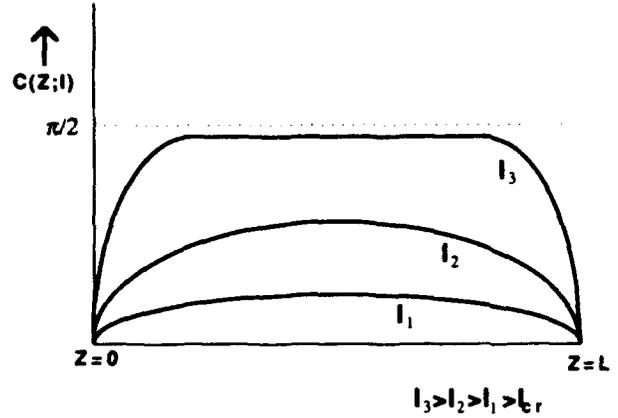


Fig. 6. A sketch of the qualitative behavior of the nematic distortion at beam center as would be obtained by solving the bifurcation equation (5.3) for $C(z; I)$.

$C(z)$. This equation (5.3) should be viewed as an extension of the Frederiks bifurcation equation to the boundary layer situation of variable field strength. It must be solved numerically. Its solution will behave as a pitch-fork bifurcation and, above a critical intensity $I_c$, will have the qualitative behavior sketched in Fig. 6. Note, in particular, this nematic distortion $C(z; I)$ contains the "longitudinal transitions" in $z$.

Turning to the behavior of the electric and nematic fields in the *inner-beam* region, we express the nematic equation (3.6) in the inner coordinates and solve it asymptotically for $\theta^{(b)}$,

$$\theta^{(b)} = -\tfrac{1}{2}\sin(2C)\left(\int_{\bar{x}}^{\infty}\int_{y}^{\infty} |F|^2 \,\mathrm{d}\bar{y}\,\mathrm{d}y\right.$$

$$\left. + \int_{-\infty}^{\bar{x}}\int_{-\infty}^{y} |F|^2 \,\mathrm{d}\bar{y}\,\mathrm{d}y\right). \qquad (5.4)$$

Inserting this expression into the field equation (3.5) written in inner variables yields, after removing an appropriate fast phase $kS(x, z)$,

$$iG_z + \frac{\gamma^2}{2k}G_{\bar{x}\bar{x}} + \frac{k\alpha}{2\gamma}\sin(2C)\,\theta^{(b)}G = 0, \qquad (5.5)$$

where

$$F^{(b)} = G\exp(ikS).$$

Examination of eq. (5.5) shows that the natural balance of diffraction against nematic distortion defines the scaling

$$\gamma(k) = k^{2/3}. \qquad (5.6)$$

With this choice of $\gamma$ scaling, and introducing the propagation scale $\bar{z} \equiv k^{1/3}z$, we obtain the leading order equation which governs the structure of the inner beam:

$$iG_{\bar{z}} + \tfrac{1}{2}G_{\bar{x}\bar{x}} - \tfrac{1}{4}\alpha\sin^2(2C)\left\{ \int_{\bar{x}}^{\infty}\int_{y}^{\infty} |G|^2\,d\bar{y}\,dy \right.$$

$$\left. + \int_{-\infty}^{\bar{x}}\int_{-\infty}^{y} |F|^2\,d\bar{y}\,dy \right\}G = 0. \qquad (5.7)$$

Several remarks should be made about this nonlocal, nonlinear Schroedinger equation (5.7) which governs the local structure of the inner electric field:

(i) This nonlinear Schroedinger equation is a Hamiltonian system, with Hamiltonian

$$H \equiv \int_{-\infty}^{+\infty} \left\{ \tfrac{1}{2}|G_{\bar{x}}|^2 + \tfrac{1}{4}\alpha\sin^2(2C)\left(\int_{\bar{x}}^{\infty}|G|^2\,d\bar{y}\right)\right.$$

$$\left. \times \left(\int_{-\infty}^{\bar{x}}|G|^2\,d\bar{y}\right) \right\}d\bar{x}. \qquad (5.8)$$

(ii) It depends upon $C(z)$, the nematic distortion at beam center, which is determined from the outer problem as matched to the inner beam via bifurcation equation (5.3).

(iii) Even though this NLS equation has coefficients which depend upon $z$, the intensity $I$ is independent of $z$ and can be viewed as a parameter which is controlled by the input laser beam. However, perhaps the most important consequence of this boundary layer analysis follows from the combination of the inner beam equation (5.7) with the scaling law (5.6). Together, these imply that, in the absence of scattering losses, the self-focusing process terminates on the scale $x \simeq O(\gamma^{-1} = k^{-2/3})$.

## 6. Conclusion

In this paper we have described some initial work on the interaction of a laser beam with a nematic liquid crystal. Our interest in this system stems from its very large coefficient of nonlinearity which allows the generation of complex nonlinear optical patterns with low power, CW lasers. Experimentally, Braun, Faucheux, and Libchaber have studied this system in three geometric configurations of film, droplet, and tube [2], [1]. Of particular importance in these experimental studies are the striking longitudinal views of interacting filaments in the cylindrical geometry.

In this paper we have described a theoretical boundary layer analysis of an idealized (two dimensional, scalar) model of this coupled light–nematic system in a thin film geometry. The transverse boundary layer analysis shows that the nematic distortion at beam center is described by a "local Frederiks bifurcation curve" which captures transitions in the nematic near the front and rear faces of the film. The analysis also produces a scaling function which results from "maximal balance" between diffraction and nonlinearity and fixes (in the absence of scattering losses) the spatial extent of the self-focussed beam. It also yields a nonlocal nonlinear Schroedinger equation which governs the inner structure of the beam.

In other theoretical boundary layer analyses of this coupled system, we have studied the two dimensional "tube geometry" for both a model scalar[1] and full vector [7] optical field. In [1] we show that in the tube geometry, the two dimensional scalar model captures the essential physics of both the filamentation and the *transverse undulation* of the laser beam. The latter is a new phenomenon. In order to describe polarization effects, which are certainly absent in these scalar models, we have studied a full (but two dimensional) vector system in the tube geometry [7]. In this study we identify properties of the undulation process that are distinctly polar-

ization effects.

Taken together, these experimental, numerical, and theoretical studies establish that the light–nematic system is an excellent source of complex, nonlinear spatial patterns. Furthermore, our studies show that *coupled nonlinear pde* descriptions of light–nematic interactions are indeed accessible theoretically, through boundary layer asymptotics implemented numerically.

## Acknowledgements

We gratefully acknowledge Erez Braun, Luc Faucheux, and Albert Libchaber, who have introduced us to liquid crystal self-focusing through their elegant experiments, and with whom we have enjoyed many hours of intense collaboration.

## References

[1] E. Braun, L.P. Faucheux, A. Libchaber, D.W. McLaughlin, D.J. Muraki and M.J. Shelley, Filamentation and undulation of self-focused laser beams in liquid crystals, Europhys. Lett. (1993), in press.

[2] E. Braun, L.P. Faucheux and A. Libchaber, Strong self focusing in nematic liquid crystals, Phys. Rev. A 47 (1993).

[3] P. Collings, ed., Liquid Crystals (Princeton Univ. Press, Princeton, 1991).

[4] P.G. de Gennes, The Physics of Liquid Crystals (Clarendon , Oxford, 1974).

[5] N.V. Tabiryan, A.V. Sukhov and B. Ya. Zeldovich, The orientational optical nonlinearity of liquid crystals, Mol. Cryst. Liq. Cryst. 136 (1986) 1.

[6] Paul Manneville, ed., Dissipative Structures and Weak Turbulence (Academic Press, Boston, 1980).

[7] D.J. Muraki, D.W. McLaughlin and M.J. Shelley, Laser light interacting with nematic liquid crystals, in preparation (1992).

**PHYSICA** D

# Space–time complexity in nonlinear optics

J.V. Moloney, P.K. Jakobsen, J. Lega, S.G. Wenden and A.C. Newell
*Arizona Center for Mathematical Sciences, University of Arizona, Tucson, AZ 85721, USA*

Travelling wave solutions are found to be the natural nonlinear modes of wide aperture two-level and Raman lasers for frequency detunings to the positive side of the gain peak.

## 1. Introduction

Complex pattern formation is commonly observed in spatially extended, continuous, dissipative systems which are driven far from equilibrium by an external stress. Under the influence of this stress, the system can undergo a series of symmetry breaking bifurcations or phase transitions and the resulting patterns become more and more complicated, both temporally and spatially, as the stress is increased. Examples abound in ordinary and binary fluids, in liquid crystals and chemically reacting media [1]. Optical systems, both passive and active, are no exception and considerable effort has been expended recently to predict [2, 3] and observe [4, 5] pattern forming instabilities in both passive and active nonlinear optical systems.

The plan of the article is as follows. In the next section we introduce the laser models, compare and contrast their physical characteristics and briefly review their bifurcation behavior. An important conclusion will be that both laser models admit an exact finite amplitude travelling wave solution for positive detuning ($\Omega$) of the laser and a homogeneous solution for negative detuning. A weakly nonlinear analysis near threshold for lasing in the following section yields a coupled set of amplitude equations of the complex Newell–Whitehead–Segel type when $\Omega > 0$. Linear stability analysis of this latter set shows that the travelling wave solution arises from an instability of a standing wave. The very simple form of this exact solution allows us to extend the analysis beyond threshold and derive an equation for the evolution of the phase. This yields analytical expressions which predict the coexistence of stable, Eckhaus and zig-zag unstable regions beyond lasing threshold. We next provide some preliminary numerical results which confirm that the wide aperture Raman laser exhibits weakly turbulent behavior, whereby instabilities on an underlying travelling wave can spontaneously nucleate seas of optical vortices and .oll-like structures. The physical manifestation of the travelling wave solution is a strong off-axis emission of the laser. Finally we conclude with some comments on ongoing and future work.

## 2. Two-level and Raman lasers: background theory

The essential difference between the two lasers lies in the method of pumping employed in order to achieve population inversion. Figure 1 shows a schematic of the energy level schemes and pumping
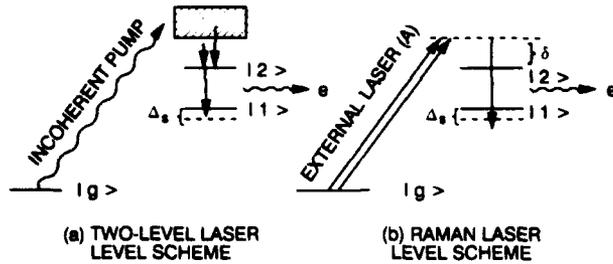
Fig. 1. Energy level diagrams depicting pumping schemes for a (a) two-level and (b) Raman laser.
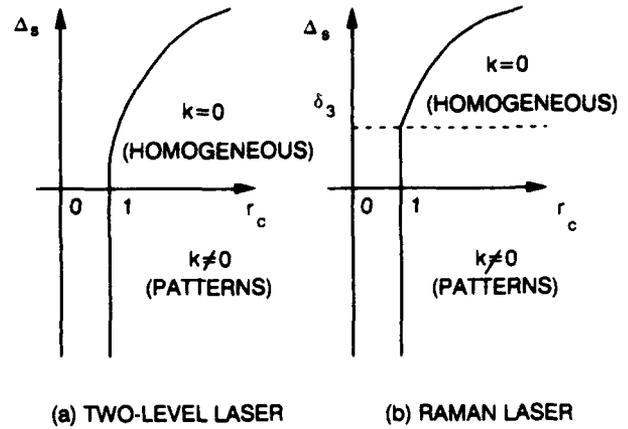
Fig. 2. Stability and instability regions for the nonlasing solution in ($\Delta_s$, $r_c$) parameter space. (a) Two-level and (b) Raman laser.

mechanisms for both lasers. The pump is the principal stress parameter for the problem. Inversion for lasing in a two-level laser is created via incoherent pumping (electrical/flashlamp discharge, collisions, etc.) whereas, in the Raman laser, a classic three-wave interaction involving two optical and one material wave, introduces a strong coherence between the pump wave (amplitude $A$) and the laser emission field (amplitude $e$). Figure 1a shows incoherent pumping of a broad upper manifold of levels with subsequent decay to form an excess of population in the upper lasing level $|2\rangle$. In the Raman laser depicted in fig. 1b, the external pump laser ($A$) can be detuned either above or below ($\delta < 0$ or $> 0$) an intermediate dipole coupled level $|2\rangle$ and the laser emission field ($e$) is generated via the three-wave interaction. Mirror optical feedback in both cases ensures that the finite lasing emission field ($e$) will build up from noise if the external stress parameter $r$ exceeds some critical value $r_c$. The distinction between incoherent and coherent pumping ensures that even the simple single mode Raman laser exhibits much richer nonlinear dynamical behavior than its two-level counterpart [6].

The mathematical description of both lasers derives directly from the Maxwell equations for the optical fields and the appropriate material density matrix equations. When appropriately scaled, the resulting Maxwell–Bloch laser equations can be cast in the form of a generalized set of complex Lorenz equations for each laser.

*Case A. Two-level laser:*

$$e_t - ia\nabla^2 e = -\sigma e + \sigma p, \quad p_t + (1 + i\Omega)p = (r - n)e, \quad n_t + bn = \tfrac{1}{2}(e^* p + ep^*), \tag{1}$$

with $\Omega = -\Delta_s$.

*Case B. Raman laser:*

$$e_t - ia\nabla^2 e = -\sigma e + \sigma p + i\delta_1 en, \quad p_t + (1 + i\Omega)p = (r - n)e + i\delta_2 |e|^2 p,$$

$$n_t + bn = \tfrac{1}{2}(e^* p + ep^*), \tag{2}$$

with $\Omega = -\Delta_s + r\delta_3$.

The complex laser emission field is denoted by $e$, the complex polarization field by $p$ and the real population inversion by $n$. The parameter $\sigma$ is proportional to the cavity loss, $b$ is the ratio of the polarization dephasing to inversion decay rates, $\Delta_s$ measures the laser cavity–atomic transition frequency detuning and $r$, the principal stress parameter of the system, depends on the nature of the pumping. Notice that if the parameters $\delta_i, i = 1, 2, 3$ appearing in eqs. (2) are set to zero, both problems are mathematically identical. The transverse Laplacian operator $\nabla^2 = (\partial^2/\partial x^2 + \partial^2/\partial y^2)$ includes the transverse degree of freedom allowing for pattern formation. The coefficient $a$, which is the inverse of the Fresnel number, measures the characteristic width of the transverse aperture of the laser. The most significant new physics in the Raman laser is the A.C. Stark shift, which appears in the set (2) as a nonlinear detuning term proportional to $\delta_2$. As mentioned earlier, the additional terms proportional to $\delta_i, i = 1, 2, 3$ appearing in the Raman laser model, substantially modify the bifurcation behavior of even the simplest single mode laser [6], making it one of the earliest laser systems showing evidence for chaotic dynamics [7].

Some of the more important simple solutions to eqs. (1) and (2) and their associated bifurcation behavior will now be briefly reviewed. Both models admit a trivial (nonlasing) solution $(e, p, n) = 0$, whose linearization differs only in the dependence of the detuning parameter $\Omega$ on the stress parameter $r$ for the Raman laser (case B). The stability characteristics of this solution are succinctly captured in the $(\Delta_s, r_c)$ plane, as shown in fig. 2 for both cases. The laser oscillation frequency and stress parameter at onset of lasing are given by the formulas

$$\nu_c = (\sigma \Omega_c + ak^2)/(\sigma + 1), \quad r_c = 1 + (\Omega_c - ak^2)^2/(\sigma + 1)^2 \tag{3}$$

with $\Omega_c = -\Delta_s$ (case A) and $\Omega_c = -\Delta_s + r_c\delta_3$ (case B). An important point to note is that for $\Omega < 0$, the homogeneous state $(k = 0)$ has the lowest threshold $(r_c = 1 + \Omega_c^2/(\sigma + 1)^2)$, whereas for $\Omega > 0$ the mode with $k_0 = \pm\sqrt{\Omega_c/a}$ has the lowest threshold. The formula for the laser frequency $\nu_c$ which is a simple generalization of the classic frequency pulling formula, has the obvious physical interpretation that, for $\Omega > 0$ the laser will seek to develop some transverse pattern that minimizes the detuning, thereby maximizing the emission. In a large aspect ratio system one might anticipate the spontaneous formation of local patches of patterns with random relative orientations.

When $r > r_c$, the two-level laser undergoes a supercritical (Hopf) bifurcation to a spatially homogeneous lasing state for $\Omega < 0$ or to a stable travelling wave state $(e, p, n) = (\bar{e}\,e^{i\theta}, \bar{p}\,e^{i\theta}, \bar{n})$, $\theta = k \cdot x + \omega t$ for $\Omega > 0$ with

$$\bar{e} = \sqrt{b(r - r_c)}\,e^{-i\nu_c t}, \quad \bar{p} = [1 - i(\nu_c/\sigma)\sqrt{b(r - r_c)}]e^{i\nu_c t}, \quad \bar{n} = r - r_c,$$

and $\omega = -(\sigma + 1)^{-1}(\sigma\Omega + ak^2)$. In the Raman laser a similar transition to a homogeneous $(\Omega < 0)$ or travelling wave $(\Omega > 0)$ state is observed although in this case the bifurcation may be sub- or supercritical. The lasing emission intensity $\bar{e}^2$ (can be assumed real) is now given as the solution to the following quadratic equation in $\bar{e}^2$,

$$\left(\frac{\delta_1}{b} - \delta_2\right)^2 \bar{e}^4 + \left[2(\Omega - ak^2)\left(\frac{\delta_1}{b} - \delta_2\right) + \frac{(1 + \sigma)^2}{b}\right]\bar{e}^2 + (\Omega - ak^2)^2 - (r - 1)(1 + \sigma)^2$$
$$= 0$$

with
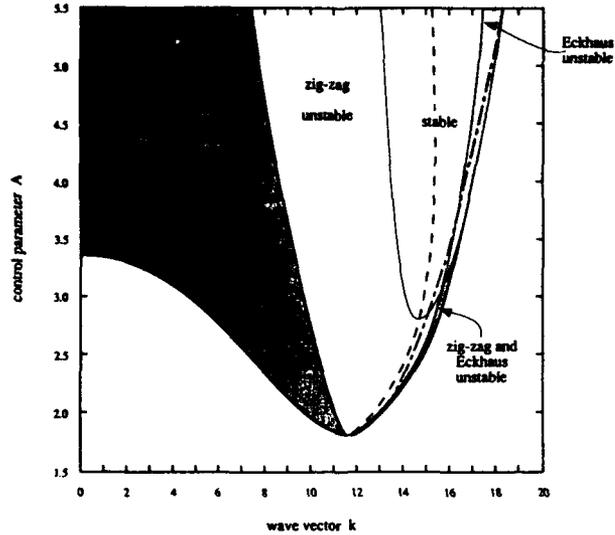
Fig. 3. Existence domain for travelling wave solutions and their associated stability properties for a Raman laser with $\Omega > 0$.

$$\omega = (\sigma + 1)^{-1}[\sigma\delta_2\bar{e}^2 - \sigma\Omega - ak^2 + \delta_1\bar{e}^2/b], \quad \bar{n} = \bar{e}^2/b, \quad \bar{p} = \bar{e}(1 + i\alpha),$$

where

$$\alpha = (\sigma + 1)^{-1}[\bar{e}^2(\delta_2 - \delta_1/b) - (\Omega - ak^2)].$$

It is a straightforward matter to show that the bifurcation to this state will be subcritical if the condition $2(\Omega - ak^2)(\delta_1/b - \delta_2) + (\sigma + 1^2)/b < 0$ is satisfied. Moreover, it can be shown using weakly nonlinear analysis near onset that this *exact* travelling wave solution appears as an instability of a standing wave. This point is discussed briefly in the following section. Numerical integration of both sets of laser equations (1) and (2), for one transverse spatial dimension and a variety of initial conditions suggests that the above travelling wave with $k_0 = \sqrt{\Omega/a}$ is a globally attracting state of the system for $\Omega > 0$ [8].

For illustrative purposes, we show in fig. 3 the neutral stability curve in the $(r, k)$ plane, indicating the region of existence of the travelling wave solution and its stability characteristics for the case of the Raman laser (case B). Observe the coexistence of Eckhaus, zig-zag unstable and stable travelling wave domains. Our parameters were chosen to be identical to those used in ref. [6] with the additional diffraction parameter $a = 0.05$. Along the curve (— · —) the bifurcation is subcritical and the curve (- - -) denotes the maximum emission amplitude of the laser satisfying $\Omega = ak_0^2$. Note also that the lasing emission threshold for the travelling wave is substantially lower than for the homogeneous state. The Eckhaus, zig-zag and stable boundaries can be determined analytically from the phase equation presented in the following section. We remark that the relative disposition of these boundaries depends strongly on the sign of $\delta$. The stable travelling wave region is the optics analog of the Busse balloon relevant to pattern formation in large aspect ratio fluids. A further comment is in order with regard to the distinction between 1D and 2D instabilities in these laser systems. In 1D, the zig-zag instability is absent and only the Eckhaus remains. We therefore expect to observe a fundamental difference between 1D and 2D pattern forming instabilities. We will explore further the natures of the 2D pattern forming instabilities when we discuss some numerical solutions to eqs. (2) later.
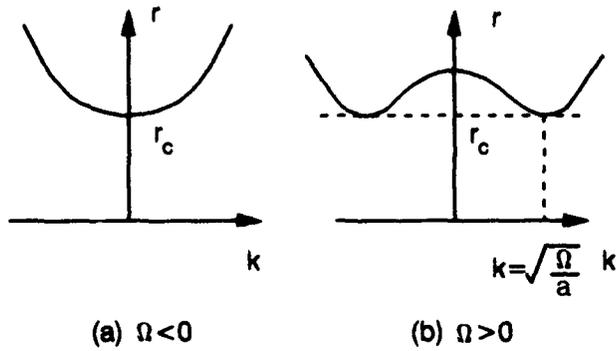
Fig. 4. Neutral stability curves for (a) $\Omega > 0$ and (b) $\Omega < 0$, applicable to both types of laser.

## 3. Near and beyond threshold behavior: amplitude and phase equations

Neutral stability curves, corresponding to negative and positive detuning respectively, are sketched in figs. 4a and 4b and apply to both laser models.

We now specialize to the Raman laser (case B) and assume $\Omega > 0$ as this case offers the richest phenomenology. Carrying out the weakly nonlinear analysis near threshold along the lines described in refs. [8, 9] for the two-level laser, we arrive at the following coupled complex Newell–Whitehead–Segel equations describing pattern formation near the onset of lasing,

$$(\sigma + 1)\frac{\partial B_1}{\partial t} + 2a\sqrt{\frac{\Omega}{a}}\frac{\partial B_1}{\partial x} - ia\nabla^2 B_1 + \frac{\sigma a^2}{(\sigma + 1)^2}\left(2i\sqrt{\frac{\Omega}{a}}\frac{\partial}{\partial x} + \frac{\partial^2}{\partial y^2}\right)^2 B_1$$

$$= \sigma(r - 1)B_1\left[(1 + i\delta_3) - \left(\frac{1}{b} - i\left(\delta_2 + \frac{\delta_1}{\sigma b}\right)\right)(|B_1|^2 + 2|B_2|^2)\right],$$

$$(\sigma + 1)\frac{\partial B_2}{\partial t} - 2a\sqrt{\frac{\Omega}{a}}\frac{\partial B_2}{\partial x} - ia\nabla^2 B_2 + \frac{\sigma a^2}{(\sigma + 1)^2}\left(-2i\sqrt{\frac{\Omega}{a}}\frac{\partial}{\partial x} + \frac{\partial^2}{\partial y^2}\right)^2 B_2$$

$$= \sigma(r - 1)B_2\left[(1 + i\delta_3) - \left(\frac{1}{b} - i\left(\delta_2 + \frac{\delta_1}{\sigma b}\right)\right)(|B_2|^2 + 2|B_1|^2)\right], \tag{4}$$

with $\Omega = -\Delta_s + \delta_3 > 0$. Setting $\delta_i = 0, i = 1, 2, 3$, we recover the coupled complex Newell–Whitehead–Segel equations for the two-level laser near onset given as eq. (6.246) in ref. [9]. These coupled amplitude equations are generally valid near threshold for the two-level laser with $\Omega > 0$ but their validity is restricted to the case where the bifurcation in the Raman laser near threshold is supercritical and $\Omega > 0$. It can be shown for both cases that the standing wave solution $B_1 B_1^* = \frac{1}{3}b$ is unstable to a travelling wave solution [8]. These latter waves are seen to move with the group velocity $v_g = \pm(2\sqrt{a\Omega}/\sigma + 1)$.

By using the exact travelling wave solution above threshold $(e, p, n) = (\bar{e}\,e^{i\theta}, \bar{p}\,e^{i\theta}, \bar{n})$, and assuming that the amplitude is slaved to the phase, we can derive an evolution equation for the phase $\theta$ itself,

$$\theta_t = \omega - \frac{1}{\tau(k)}\left(\frac{2ak^2}{\bar{e}}(x_1 + x_1^*)\frac{d^2\bar{e}^2}{dk^2} + a\bar{e}(x_1 + x_1^*) + 4s(k)\frac{d\omega}{dk^2}k^2\right)\theta_{xx}$$

$$- \frac{a}{\tau(k)}(x_1 + x_1^*)\bar{e}\theta_{yy},$$

where

$$x_1 + x_1^* = (2\bar{e}/\sigma)[\alpha + (b\delta_2 - \delta_1/\sigma)\alpha^2 - (b\delta_2 + \delta_1/\sigma) - \delta_2\bar{e}^2(1 - 2\delta_1\alpha/\sigma)],$$

$$\tau(k) = -(2\bar{e}^2/\sigma)[1 + \sigma + 2\alpha(b\delta_2 - \delta_1)],$$

$$s(k) = \frac{d\bar{e}^2}{dk^2}\left[\frac{\alpha}{\sigma} + \left(b\delta_2 - \frac{\delta_1}{\sigma}\right)\frac{\alpha^2}{\sigma} - \frac{1}{\sigma}\left(b\delta_2 + \frac{\delta_1}{\sigma}\right) - \frac{\delta_2}{\sigma}\left(1 - 2\delta_1\frac{\alpha}{\sigma}\right)\bar{e}^2 + \frac{2}{b}\delta_2\bar{e}^2\left(1 - 2\delta_1\frac{\alpha}{\sigma}\right)\right.$$

$$- \left(b\delta_2 + \frac{\delta_1}{\sigma}\right) - \frac{3\bar{e}^2}{\sigma + 1}\left(\delta_2 - \frac{\delta_1}{\sigma}\right)\left(1 + \alpha\left(b\delta_2 - \frac{\delta_1}{\sigma}\right)\right)$$

$$\left. + \frac{\Omega - ak^2}{1 + \sigma}\left(1 + \alpha\left(b\delta_2 - \frac{\delta_1}{\sigma}\right)\right)\right] - \frac{2a}{\sigma + 1}\bar{e}^2\left(1 + \alpha\left(b\delta_2 - \frac{\delta_1}{\sigma}\right)\right).$$

This rather complicated expression enables us to determine analytically the boundaries for phase instabilities and compare those with results obtained from the exact problem and amplitude equation approximation.

## 4. Numerical simulations beyond threshold

We have numerically integrated eqs. (2) for the Raman laser using parameter values which yielded fig. 3, for a variety of initial conditions both without and with a weak external probe beam. The idea with the external probe is that it may be possible to injection lock local spatial domains of the wide aperture laser in a controllable manner. Choosing the pump parameter $A = 3.0$ (see fig. 3) and initial data corresponding to a plane travelling wave with wavenumber lying in the stable band, confirmed that the travelling wave was indeed stable. If a weak external probe is applied so as to favor the zig-zag instability, which occurs at right angles to the travelling wave direction, we observe the appearance of a rather robust solitary wave train as an alternating dark/bright pattern.

Figures 5A and 5B show a sequence of frames from a movie of the near- and far-field output cross-section of the laser, when the pump $A = 3.0$ and the field is initiated from noise. The sequence of frames (a)–(f) in fig. 5A shows the build-up from noise (random pattern) to a finite amplitude emission consisting of a sea of optical vortices and bright ridges which arise from a combination of Eckhaus and zig-zag instabilities. The bright ridges are aligned at right angles to the direction of the travelling wave which is moving from the top left to bottom right of each frame. This complicated spatio-temporal pattern persists indefinitely in time with no sign of any regular recurrence. The far-field emission shows the initial appearance of a weak far-field ring with radius equal to the value $k = k_c$ at threshold. As the amplitude grows the ring expands slightly consistent with the fact that the wavenumber $k$ increases as one moves above threshold. The ring in the far-field signifies the absence of any preferred direction due to the degeneracy in the local plane wave directions. Once the amplitude becomes significant the travelling wave starts to choose a fixed direction and the emission adjusts to a reasonably well-defined spot off-axis. This off-axis spot remains intense with a randomly fluctuating lower amplitude background indicative of a weakly turbulent state of the laser emission.
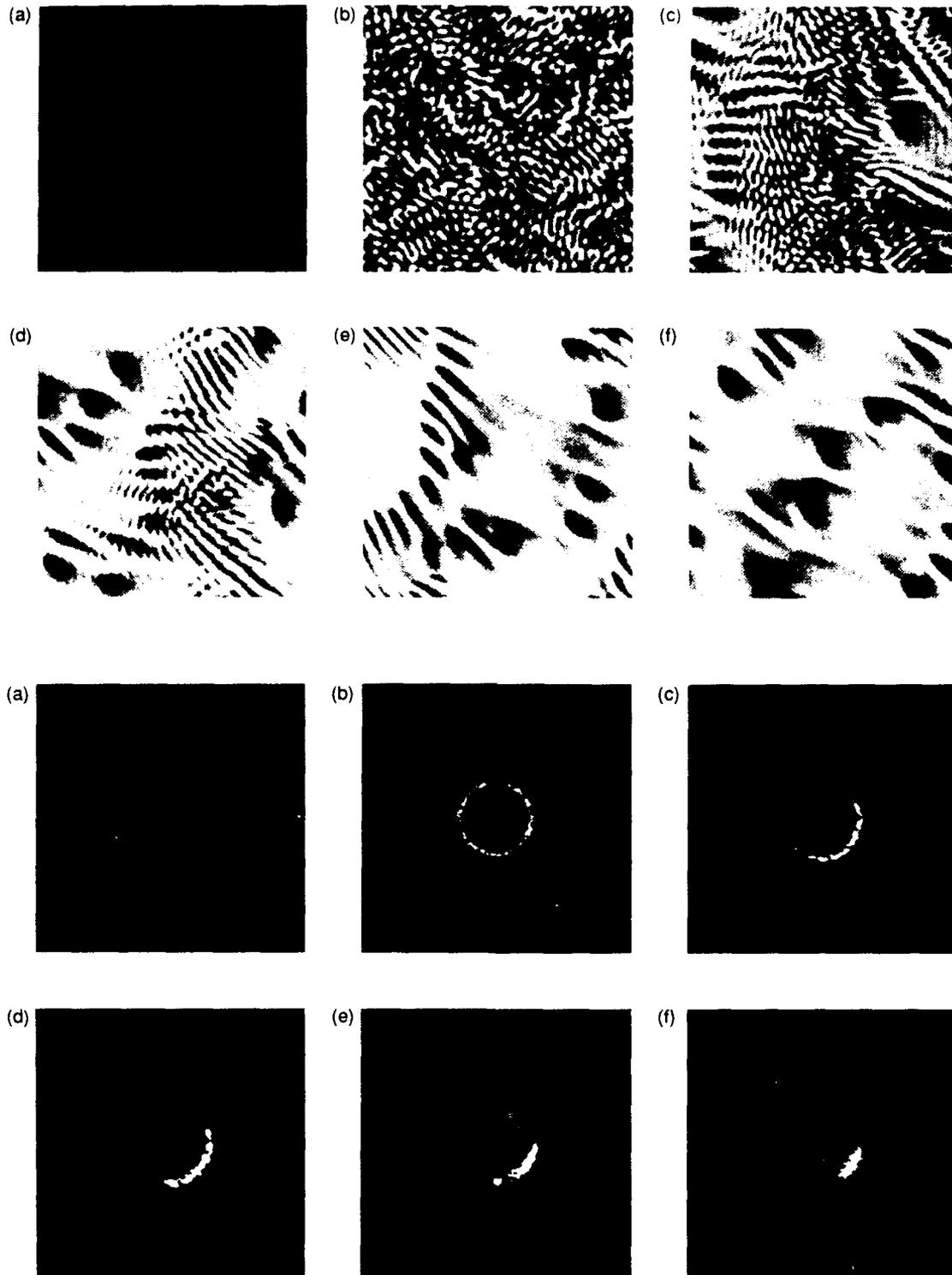
B. FAR-FIELD EMISSION

Fig. 5. (A) Frames from a movie showing a succession of trasnsverse pattern evolutions from noise in the near-field. (B) Far-field emission corresponding to each near-field frame in (A).

## 5. Conclusions

Both two-level and Raman wide aperture lasers are capable of displaying a rich variety of pattern forming instabilities. The nonlinear detuning term in the Raman laser is responsible for promoting a broader range of instability behavior than can occur in the two-level laser. The nature of the instability depends on the transverse dimension of the laser system and it is anticipated that much of the spatio-temporal behavior of these lasers will carry over to technologically important wide aperture semiconductor laser systems. The latter systems require a much more complicated material description, involving many-body interactions between carriers and holes at the microscopic level and a major challenge that remains is to derive order parameter equations near onset that are capable of predicting and suggesting means of stabilization of novel pattern shapes in a broad area and vertical cavity surface emitting semiconductor lasers.

## Acknowledegement

## References

[1] A.C. Newell, T. Passot and J. Lega, Order parameter equations for patterns, Annu. Rev. Fluid Mech. 25 (1993), to appear.
[2] R. Chang, W.J. Firth, R. Indik, J.V. Moloney and E.M. Wright, Three-dimensional simulations of degenerate counterprogating beam instabilities in a nonlinear medium, Opt. Commun. 88 (1992) 167.
[3] P. Coullet, L. Gil and F. Rocca, Optical vortices, Opt. Commun. 73 (1989) 403.
[4] F.T. Arecchi, G. Giacomelli, P.L. Ramazza and S. Residori, Vortices and defect statistics in two-dimensional optical chaos, Phys. Rev. Lett. 67 (1991) 3749.
[5] D. Hennequin, C. Lepers, E. Louvergneaux, D. Dangoisse and P. Glorieux, Spatio-temporal dynamics of a $CO_2$ laser, in: XVIII International Quantum Electronics Conference Technical Digest, Vienna, June 14–19, 1992.
[6] R.G. Harrison, W. Lu and P.K. Gupta, Origin of periodic, chaotic and bistable emission of Raman lasers, Phys. Rev. Lett. 63 (1989) 1372.
[7] R.G. Harrison and D.J. Biswas, Demonstration of self-pulsing instability and transitions to chaos in single-mode and multi-mode homogeneously broadened Raman laser, Phys. Rev. Lett. 55 (1985) 63.
[8] P.K. Jakobsen, J.V. Moloney, A.C. Newell and R. Indik, Space-time dynamics of wide-gain-section lasers, Phys. Rev. A 45 (1992) 8129.
[9] A.C. Newell and J.V. Moloney, Nonlinear Optics (Addison-Wesley, Reading, 1992).

PHYSICA D

# Applications of self-trapping in optically coupled devices

G.P. Tsironis, W.D. Deering and M.I. Molina

*Physics Department, University of North Texas, Denton, TX 76203, USA*

We present two generalizations to standard optical coupler configurations. Both cases show improved switching characteristics.

The switching characteristics of nonlinear directional couplers have been the object of attention since the work of Jensen [1] and Maïer [2]. These authors demonstrated the possibility of an all-optical switching in a two waveguide device based on the occurrence of power-triggered switching (power self-trapping). The present study will focus on means of improving the switching profile and decreasing power levels required for the device to operate by considering direct extensions of Jensen's original idea. Consider a one-dimensional array of $N$ (different) nonlinear directional couplers. The nonlinear waveguide medium is assumed to be of the Kerr type, i.e., with refractive index given by $n_i = n_{0i} + n_{2i}|E_i|^2$ where $n_{0i}$ is the linear index of refraction of the $i$th guide, $n_{2i}$ is its Kerr coefficient and $E_i$ is the electric field inside the guide. Treating the system in the coupled-mode approximation and restricting each waveguide to separately support a single mode only and neglecting cross-phase modulation, the equations for the mode field amplitudes $a_n$ are

$$i\frac{da_n}{dz} = \epsilon_n a_n + V_{n,n-1}\exp[i(\phi_{n-1}-\phi_n)z]a_{n-1}$$
$$+ V_{n,n+1}\exp[i(\phi_{n+1}-\phi_n)z]a_{n+1}$$
$$- Q_n|a_n|^2 a_n, \qquad (1)$$

where $n = 1, 2, \ldots, N$, $\epsilon_n$ is a self-energy term, $V_{n,n-1}$ and $V_{n,n+1}$ are nearest-neighbor coupling coefficients. The nonlinear coefficient $Q_n$ is the

self-phase modulation coefficient, the "phase" $\phi_n$ is the wavenumber of the field along the direction of propagation in the $n$th waveguide and $P$ is the total power injected into the waveguide system at $z = 0$. Power conservation $(\sum_n |a_n|^2 = P)$ requires $V_{n,n+1} = V^*_{n+1,n}$. The transformation $c_n = a_n \exp(i\phi_n z)/\sqrt{P}$ reduces system (1) to

$$i\frac{dc_n}{dz} = \delta_n c_n + V_{n,n-1}c_{n-1} + V_{n,n+1}c_{n+1}$$
$$- \chi_n|c_n|^2 c_n, \qquad (2)$$

where $\delta_n \equiv \epsilon_n - \phi_n$, $\chi_n = PQ_n$ and $\sum_n |c_n|^2 = 1$. Note that in eq. (2) the "phases" $\phi_n$ play the role of local "self-energy" terms or local "site energies".

System (2) is a fairly general form of the discrete nonlinear Schrödinger equation (DNSE) or discrete selftrapping equation introduced by Eilbeck, Lomdahl and Scott in the context of coupled nonlinear oscillators [3]. In most cases of physical interest, systems of *identical* waveguides are considered, in which case $\delta_n = \delta$, $V_{n,n-1} = V_{n,n+1} = V$ and $\chi_n = \chi$. In Jensen's two identical coupler device ($N = 2$), in particular, and for the initial condition which places all the initial power in one of the guides (say guide 1), the *exact* solution shows the presence of a "dynamical" self-trapping transition [1,2,4]. For values $\chi/4V < 1$ there is a complete transfer of power between the two guides in a trigonometric-like fashion, while for $\chi/4V > 1$,
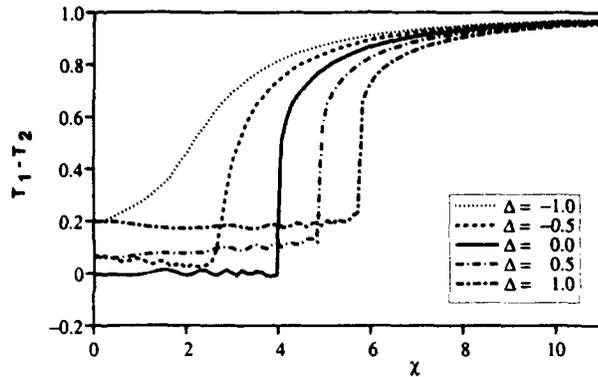
Fig. 1. Nonlinear two-waveguide with phase mismatch: space-averaged transmittance difference $T_1 - T_2$ = $\langle |c_1|^2 \rangle - \langle |c_2|^2 \rangle$ versus power ($\chi$) for several phase mismatch values ($\Delta$) ($V \equiv 1$).



Fig. 2. Dynamical phases for the DNT system as a function of the linear coupling W and power $\chi$. The regimes labeled "free", "chaotic" and "self-trapped", refer to complete, chaotic and incomplete power transfer, respectively. The top half of the figure displays the case for $V$ and $W$ negative while the lower half is for $V$ and $W$ positive [7].

the transfer is incomplete and the power remains trapped in the original waveguide. Let us now consider a simple extension of the above system: two guides with different linear indices of refraction but rather similar Kerr coefficients. Since the waveguides are dissimilar in this case the propagation properties of the fields in the two guides are different, resulting in a *non-zero phase mismatch* $\phi_2 - \phi_1 \neq 0$ in the system of equations (1). The resulting set of equations is deceivingly complicated since these equations can be written equivalently in the form presented in eq. (2), leading to

$$i\frac{dc_1}{dz} = Vc_2 - \chi|c_1|^2 c_1, \qquad (3)$$

$$i\frac{dc_2}{dz} = \Delta c_2 + Vc_1 - \chi|c_2|^2 c_2. \qquad (4)$$

Assuming for simplicity $\epsilon_1 = \epsilon_2 = 0$ we obtain $\Delta = \delta_2 - \delta_1 = \phi_1 - \phi_2$ as the effective mismatch between the two guides. Equations (3), (4) represent the exact equations of motion for a nonlinear *nondegenerate* dimer [5,6]. Figure 1 shows numerical simulation results for the space-averaged difference in transmittance $\langle |c_1|^2 \rangle - \langle |c_2|^2 \rangle$ as a function of $\chi$ (proportional to the input power $P$) for several values of $\Delta$ for $c_1(0) = 1$ and $c_2(0) = 0$. For $\Delta = 0$ we have the degenerate dimer self-trapping transi-

tion at $\chi/4V = 1$. For $\Delta < 0$, the position of the transition shifts to *lower* power levels, while for $\Delta > 0$, the tendency is reversed and *more* power is required. A phase diagram illustrating the different regimes of behavior can be found in ref. [6].

Another simple variant of the Jensen model is the doubly nonlinear trimer (DNT) system: two identical nonlinear waveguides interacting with a third *linear* guide in a closed configuration [7] ($\delta_1 = \delta_2 = \delta_3, \chi_1 = \chi_2, \chi_3 = 0$). In practice this could be achieved by placing three waveguides in a triangular array with (nonlinear) guides 1 and 2 in close proximity and (linear) guide 3 symmetrically below the mid point of the line joining the other two. The equations for the field amplitudes are

$$i\frac{dc_1}{dz} = Vc_2 + Wc_3 - \chi|c_1|^2 c_1, \qquad (5)$$

$$i\frac{dc_2}{dz} = Vc_1 + Wc_3 - \chi|c_2|^2 c_2, \tag{6}$$

$$i\frac{dc_3}{dz} = W(c_1 + c_2), \tag{7}$$

where $W$ is the coupling to the linear guide. Numerical simulations show that the self-trapping transition is sustained in the presence of the third linear guide. Furthermore, by a judicious choice of the value of $W$, the transition can be *tuned* to occur at much lower input power levels (fig. 2).

The study of simple extensions and modifications to Jensen's original device demonstrates that it is indeed possible to improve on the characteristics of that device. In particular, lower power *and* reasonably abrupt switching are possible in the context of the models we described above. The attainment of these desirable features is a positive step in the direction of the fabrication of ultrafast all-optical switching devices at reasonable power levels.

## References

[1] S.M. Jensen, IEEE J. Quant. Electron. QE-18 (1982) 1580.

[2] A.A. Maïer, Sov. J. Quantum Electron. 14 (1984) 101.

[3] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, Physica 16 D (1985) 318.

[4] V.M. Kenkre and D.K. Campbell, Phys. Rev. B 34 (1986) 4959;
G.P. Tsironis and V.M. Kenkre, Phys. Lett. A 127 (1989) 209.

[5] A.C. Scott, Physica Scripta 42 (1990) 14.

[6] G.P. Tsironis, Phys. Lett. A 173 (1993) 381.

[7] M.I. Molina and G.P. Tsironis, Phys. Rev. A 46 (1992) 1124;
M.I. Molina, G.P. Tsironis and W.D. Deering, Physica D 66 (1993) 135.

**PHYSICA D**

# On the quantum inverse scattering method for the DST dimer

V.Z. Enol'skii[a], V.B. Kuznetsov[b] and M. Salerno[c]

[a] *Department of Theoretical Physics, Institute of Metal Physics, Vernadsky str. 36, Kiev-680, 252142, Ukraine*
[b] *Department of Mathematical and Computational Physics, Institute of Physics, Saint-Petersburg University, Saint-Petersburg 198904, Russia*
[c] *Department of Theoretical Physics, University of Salerno 84100, Salerno, Italy*

The quantum inverse scattering method is used to solve the spectral problem of the discrete self-trapping dimer, in the case of both a quadratic and a linear $r$-matrix algebra representation. The first case is solved by the algebraic Bethe ansatz, while in the case of the linear $r$-matrix algebra we use the method of separation of variables. In this last case it is shown that the wave functions of the quantum discrete self-trapping dimer are related to the solutions of a Heun's type equation and that the system is equivalent to the two-site hyperbolic Gaudin magnet separable in elliptic coordinates.

## 1. Introduction

During the past years the quantum inverse scattering method (QISM) has been shown to be a very powerful tool in the field of quantum integrability. This method was used to perform the exact quantization of several nonlinear field theories of physical interest such as nonlinear Schrödinger, sine-Gordon etc., and provides a unifying bridge with the works of Bethe, Yang, Baxter, etc. on solvable lattice models of statistical physics [7,13]. Besides this, the QISM has shown to have interesting links with the theory of factorizable $S$-matrices, braid groups and the newly born discipline of quantum groups. In the context of quantum integrable systems the QISM is used to find the common eigenvectors and the eigenvalues of the quantum integrals of motion. The method consists in including a commutative subalgebra of operators, corresponding to the quantum integrals of motion, in some associative algebra of generators $T_{\alpha\beta}(u)$ ($\alpha, \beta = 1, \ldots, d$) satisfying the following quadratic $R$-matrix relation:

$$R(u - v)T_1(u)T_2(v) = T_2(v)T_1(u)R(u - v), \tag{1.1}$$

or the linear $r$-matrix relation

$$[T_1(u), T_2(v)] = [r(u - v), T_1(u) + T_2(v)], \tag{1.2}$$

with

$$T_1(u) = T(u) \otimes \text{id}, \quad T_2(v) = \text{id} \otimes T(v). \tag{1.3}$$

Here $T(u)$ denotes a $(d \times d)$ matrix with $q$-numbers (operator) elements, id is the unit $(d \times d)$-matrix, $R(u)$ and $r(u)$ are $(d^2 \times d^2)$ $c$-number matrices which satisfy, respectively, the quantum and classical Yang–Baxter equations [7]. The matrix $T(u)$ is the monodromy matrix (or Lax operator)

and the commutative subalgebra is simply the one generated by the trace of $T(u)$ in the quadratic case and by the trace of $T^2(u)$ in the linear case. Sometimes it is possible to include the family of integrals of motion into both linear and quadratic algebras. In this case two different Lax operators for the system must exist. Below we shall restrict ourselves to the case $d = 2$. By fixing $R$ and chosing a representation of the associative algebra (1.1) (or (1.2)) one has an integrable quantum system whose spectral problem can be solved by purely algebraic means.

The method of solution, however, may depend on the particular representation ($L$-operator) chosen for the algebra of generators. Two principal methods exist to solve this problems. The first one is the so-called algebraic Bethe ansatz (BA). This technique is only applicable when the representation of the algebra of generators

$$T(u) = \begin{pmatrix} A(u) & B(u) \\ C(u) & D(u) \end{pmatrix} \tag{1.4}$$

does possess the highest vector $|0\rangle$ (pseudovacuum) such that

$$B|0\rangle = 0. \tag{1.5}$$

The application of this method however does not automatically guarantee the completeness of the obtained solutions (it is however believed that the cases in which completeness problems arise constitute a set of zero measure).

The second method is the functional BA, also called the method of separation of variables, which was originally introduced by Sklyanin [20,21]. This method does not depend on the existence of the pseudovacuum and has the advantage of avoiding the completeness problems.

The aim of the present paper is twofold. First, we show on a concrete example how one uses in the QISM the above methods to calculate the eigenvalues and the common eigenvectors of the quantum integrals of motion. To this end we will use as a working example the quantum discrete self-trapping (QDST) dimer [3,2,19,18]. This system was introduced by Scott and has been shown to be completely integrable in terms of a Lax pair related to a quadratic $R$-matrix algebra [5]. In this representation there exists the pseudovacuum so that the algebraic BA can be applied. Second, we present new results on the QISM analysis of the QDST dimer. More precisely, we show that it is possible to construct a different Lax pair for the QDST dimer which is related to a linear $r$-matrix algebra. In this case the representation space does not contain the pseudovacuum state so that the method of separation of variables must be used. These features make the QDST dimer to be an ideal system on which the different ansatze used in the theory of quantum integrability can be tested and compared. The proposed paper is devoted exactly to the discussion of these questions. As a result we show that in the case of linear $r$-matrix algebra the resonant QDST dimer is equivalent to the hyperbolic Gaudin magnet [8] and the eigenfunctions of the integrals of motion are expressed in terms of generalized Hermite polynomials or confluent Heun's polynomials [1,10,12,16,17,23,24]. To this end we use the technique developed in refs. [11,14,15] to show that the QDST dimer separates in terms of elliptic coordinates. We note, that confluent Heun's polynomials also appeared in the quantum two-dimensional Coulomb problem which was presented in details in the book [12] and we think it is remarkable that the eigenfunctions of the QDST dimer are expressed in terms of them. This result is also compared with a direct solution of the eigenvalues problem of the QDST dimer in terms of su(2)-variables.

The paper is organized as follows. In section 2 we recall the quadratic $R$-matrix algebra for the QDST dimer and the algebraic BA is briefly reviewed. In section 3 a new Lax operator related to a

linear $r$-matrix algebra is introduced. In section 4 we show how the method of separation of variables allows to solve the spectral problem of the QDST dimer. Finally, in the last section, we check the results of section 4 by expressing the QDST Hamiltonian in terms of $su(2)$-variables and performing an alternative analysis of our problem in functional space.

## 2. Quadratic $R$-matrix algebra for the QDST dimer

The QDST dimer is defined by the following integrals of motion: the energy $H$ and the particle number $N$ [18],

$$H = -\tfrac{1}{4}\gamma[(b_1^\dagger b_1 - b_2^\dagger b_2)^2 + N^2] + (\omega_1 - \tfrac{1}{2}\gamma)b_1^\dagger b_1 + (\omega_2 - \tfrac{1}{2}\gamma)b_2^\dagger b_2 - \varepsilon(b_1^\dagger b_2 + b_2^\dagger b_1), \quad (2.1)$$

$$N = b_1^\dagger b_1 + b_2^\dagger b_2, \tag{2.2}$$

where $b_i^\dagger$ and $b_i$ are creation and annihilation operators satisfying the boson commutation relations and $\gamma$, $\varepsilon$, $\omega_i$, $i = 1, 2$, being real parameters. The corresponding system represents a two degrees of freedom integrable Hamiltonian system with equations of motion given by

$$\dot{b}_i = [H, b_i], \quad \dot{b}_i^\dagger = [H, b_i^\dagger], \quad i = 1, 2. \tag{2.3}$$

In the classical limit, these equations were shown to be exactly solvable in terms of elliptic functions [18]. In ref. [5] the QDST dimer was analized within the QISM on the background of quadratic $R$-matrix algebra defined by the principal relation (1.1) with $d = 2$. It was found that the $R$-matrix for QDST has the simplest rational form

$$R(u) = \begin{pmatrix} f(u) & 0 & 0 & 0 \\ 0 & 1 & g(u) & 0 \\ 0 & g(u) & 1 & 0 \\ 0 & 0 & 0 & f(u) \end{pmatrix}, \tag{2.4}$$

with

$$f(u) = (u - i\gamma)/u, \quad g(u) = -i\gamma/u. \tag{2.5}$$

The $L$-operator of the system is given by

$$L(u) = L^{(1)}L^{(2)} = \begin{pmatrix} A(u) & B(u) \\ C(u) & D(u) \end{pmatrix}, \tag{2.6}$$

where $L^{(m)}$ are the matrices

$$L^{(m)}(u) = \begin{pmatrix} u - i(\gamma b_m^\dagger b_n + \omega_m) & \sqrt{\varepsilon\gamma}\, b_m^\dagger \\ \sqrt{\varepsilon\gamma}\, b_m & i\varepsilon \end{pmatrix}, \quad m = 1, 2. \tag{2.7}$$

In this representation the pseudovacuum vector $|0\rangle$, satisfying the relations

$$\langle 0|B = 0, \quad B|0\rangle = 0 \tag{2.8}$$

exists so that the spectrum of the generating function of the quantum integrals of motion (i.e. the trace of $L(u)$) can be computed by the algebraic BA. Here we recall the main ideas of this analysis, referring to refs. [5,6] for more details. In the algebraic BA a crucial role is played by the so-called Bethe equations

$$\frac{a(u_k)}{d(u_k)} = \prod_{j=1, j \neq k}^{n} \frac{f(u_k - u_j)}{f(u_j - u_k)}, \quad k = 1, \ldots n, \tag{2.9}$$

with $a(u)$ and $d(u)$ given by

$$A(u)|0\rangle = a(u)|0\rangle, \quad D(u)|0\rangle = d(u)|0\rangle. \tag{2.10}$$

If this set of equations is satisfied, then the eigenfunction of the trace operator $A(u) + D(u)$ can be constructed as

$$|\Psi(u_1, \ldots, u_n)\rangle = \prod_{j=1}^{n} B(u_j)|0\rangle, \tag{2.11}$$

where $n$ is the number of quanta and $u_j, j = 1, \ldots, n$, are the Bethe parameters. The eigenvalue $t(u)$ of the trace operator

$$(A(u) + D(u))|\Psi(u_1, \ldots, u_n)\rangle = t(u)|\Psi(u_1, \ldots, u_n)\rangle \tag{2.12}$$

is given by the formula

$$t(u) = a(u)\frac{\prod_{j=1}^{n}(u - u_j + i\gamma)}{\prod_{j=1}^{n}(u - u_j)} + d(u)\frac{\prod_{j=1}^{n}(u - u_j - i\gamma)}{\prod_{j=1}^{n}(u - u_j)}, \tag{2.13}$$

with $u_1, \ldots u_n$ satisfying the the Bethe equations

$$a(u_k) = -d(u_k)\prod_{j \neq k}\frac{u_j - u_k + i\gamma}{u_j - u_k - i\gamma}, \quad k = 1, \ldots, n, \tag{2.14}$$

$$a(u) = [u - i(\omega_1 - \tfrac{1}{2}\gamma)][u - i(\omega_2 - \tfrac{1}{2}\gamma)], \quad d(u) = -\varepsilon^2. \tag{2.15}$$

The Hamiltonian operator can be written in terms of the trace operator as

$$\gamma \hat{H} = \left[\frac{1}{2}\left(\frac{d}{du}\operatorname{tr} L(u)\right)^2 - \operatorname{tr} L(u)\right]_{u=0} - \varepsilon^2 + \tfrac{1}{2}(\omega_1 - \tfrac{1}{2}\gamma)^2 + \tfrac{1}{2}(\omega_2 - \tfrac{1}{2}\gamma)^2. \tag{2.16}$$

The energy levels $E_k$ are given by the formula

$$\gamma E_k = \left[\frac{1}{2}\left(\frac{d}{du}t(u)\right)^2 - t(u)\right]_{u=0} - \varepsilon^2 + \tfrac{1}{2}(\omega_1 - \tfrac{1}{2}\gamma)^2 + \tfrac{1}{2}(\omega_2 - \tfrac{1}{2}\gamma)^2. \tag{2.17}$$

To find the energy levels $E_k$ it is convenient to rewrite (2.13) in the form [4]

$$t(u) = a(u) - \varepsilon^2 + \frac{\Phi_{n+1}(u)}{\prod_{j=1}^{n}(u - u_j)}, \tag{2.18}$$

where the polynomial $\Phi_{n+1}(u)$ of degree $n + 1$ equals

$$\Phi_{n+1}(u) = [u - i(\omega_1 - \tfrac{1}{2}\gamma)][u - i(\omega_2 - \tfrac{1}{2}\gamma)] \left( \prod_{j=1}^{n} (u - u_j - i\gamma) - \prod_{j=1}^{n} (u - u_j) \right)$$

$$-\varepsilon^2 \left( \prod_{j=1}^{n} (u - u_j + i\gamma) - \prod_{j=1}^{n} (u - u_j) \right). \tag{2.19}$$

An alternate expression for $\Phi_{n+1}(u)$ is

$$\Phi_{n+1}(u) = (-in\gamma u + \beta_n) \prod_{j=1}^{n} (u - u_j), \tag{2.20}$$

where $\beta_n$ is a new parameter. Equating coefficients of powers of $u$ between (2.19) and (2.20) leads to the condition $\det(\beta I + P) = 0$ [4], where $I$ is the $(n + 1) \times (n + 1)$ unit matrix and $P$ is the $(n + 1) \times (n + 1)$ lower Hessenberg matrix,

$$P = \begin{pmatrix} q_{0,n} & -i\gamma & 0 & 0 & \dots & 0 \\ q_{1,n} & q_{1,n-1} & -2i\gamma & 0 & \dots & 0 \\ q_{2,n} & q_{2,n-1} & q_{2,n-2} & -3i\gamma & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \dots & \vdots \\ & & & & & -ni\gamma \\ q_{n,n} & q_{n,n-1} & q_{n,n-2} & \dots & & q_{n,0} \end{pmatrix}, \tag{2.21}$$

with the matrix elements $q_{m,n-l}$, $m, l = 0, \dots, n$ given by the formula

$$q_{m,n-l} = -\theta(m - l)(-i\gamma)^{m-l-2} \binom{n-l}{n-m-2} [1 - \delta_{m,|l-2|}\theta(l - 2)]$$

$$-i\theta(m - l)(\omega_1 + \omega_2 - \gamma)(-i\gamma)^{m-l-1} \binom{n-l}{n-m-1} [1 - \delta_{m,|l-1|}\theta(l - 1)]$$

$$+ (-i\gamma)^{m-l}\theta(m - l)[\omega_1\omega_2 - \tfrac{1}{2}\gamma(\omega_1 + \omega_2) + \tfrac{1}{4}\gamma^2 + (-1)^{m-l}\varepsilon^2]$$

$$\times \binom{n-l}{n-m} (1 - \delta_{m,l}). \tag{2.22}$$

From the $n + 1$ characteristic values $\beta_n^{(k)}$, $k = 1, \dots, n + 1$, one finally obtains the energy levels of the QDST dimer as

$$E_k = n(\omega_1 + \omega_2) - \tfrac{1}{2}n(n + 2)2\gamma - \beta_n^{(k)}/\gamma. \tag{2.23}$$

The proposed method works effectively for the calculation of the energy levels for small numbers of quanta; the explicit formulae are in refs. [5,6].

## 3. Linear r-matrix algebra for the QDST dimer

In this section we introduce a new Lax operator for the QDST dimer related to a linear $r$-matrix algebra. The existence of another Lax representantion is an interesting phenomenon which is related to the possibility of constructing the different bases in the space of quantum states. In this representation the QDST dimer becomes equivalent to the two site hyperbolic quantum Gaudin magnet [8] and the spectral problem for the quantum integrals of motion can be solved by the method of separation of variables. In order to construct a linear $r$-matrix algebra representation for the QDST we take as space of the quantum states the representation space of the direct sum of two $su(1, 1)$ Lie algebras with generators $s_\alpha, \alpha = 1, 2$ satisfying the commutation relations

$$[s_\alpha^j, s_\beta^k] = -i\delta_{\alpha\beta}\varepsilon_{jkl}g_{lm}s_\alpha^m, \quad g = \text{diag}(1, -1, -1). \tag{3.1}$$

In the following, the norm and the scalar product of operator vectors $s_\alpha$ will always be computed with respect to the metric $g$, i.e.

$$s_\alpha^2 = (s_\alpha, s_\alpha) = (s_\alpha^1)^2 - (s_\alpha^2)^2 - (s_\alpha^3)^2, \quad (s_\alpha, s_\beta) = s_\alpha^1 s_\beta^1 - s_\alpha^2 s_\beta^2 - s_\alpha^3 s_\beta^3. \tag{3.2}$$

The Casimir operators of the algebra (3.1) have the form $s_\alpha^2 = k_\alpha(k_\alpha - 1)$, with $k_\alpha = 1, \frac{3}{2}, 2, \frac{5}{2}, \ldots$, for discrete series representation, while $k_\alpha$ vary continuously from zero to infinity, $0 < k_\alpha < \infty$, for representation of the universal enveloping algebra $\widetilde{su}(1, 1)$.

In terms of these variables the $L$-operator of our system is introduced as

$$L(u) = (\sigma, L(u)) = \begin{pmatrix} A(u) & B(u) \\ C(u) & -A(u) \end{pmatrix}, \tag{3.3}$$

where $L(u)$ is the operator vector

$$L(u) = \frac{s_1}{u - a} + \frac{s_2}{u - b} - \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \tag{3.4}$$

and $\sigma$ denotes the $su(1, 1)$ generators in $2 \times 2$ representation

$$\sigma^1 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$

It is easy to check that the $L$-operator (3.4) obeys the following linear $r$-matrix algebra:

$$[L^j(u), L^k(v)] = \frac{i\varepsilon_{jkl}}{u - v}g_{lm}(L^m(u) - L^m(v)), \quad j, k, l, m = 1, 2, 3. \tag{3.5}$$

This expression can be rewritten in matrix form as

$$[L_1(u), L_2(v)] = [r(u - v), L_1(u) + L_2(v)], \tag{3.6}$$

with the matrix $r(u)$ being proportional to the permutation matrix as

$$r(u) = \frac{1}{u} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In (3.6) the notations for $L_i$ are the same as in (1.3).

By defining the quantum determinant of the $L$-operator as

$$\det L(u) = -L^2(u) = -A^2(u) - \tfrac{1}{2}\{B(u), C(u)\}, \tag{3.7}$$

where the brackets $\{\cdot, \cdot\}$ denote the anticommutator, one can construct the following commutative algebra:

$$[L^2(u), L^2(v)] = 0. \tag{3.8}$$

We have then

$$L^2(u) = \frac{s_1^2}{(u-a)^2} + \frac{s_2^2}{(u-b)^2} + \frac{H_1}{u-a} + \frac{H_2}{u-b} + \frac{1}{4} \tag{3.9}$$

is the generating function of the quantum integrals of motion

$$H_1 = \frac{2}{a-b}(s_1, s_2) - s_1^1, \tag{3.10}$$

$$H_2 = \frac{2}{b-a}(s_1, s_2) - s_2^1. \tag{3.11}$$

The system defined by these two integrals is called the two site hyperbolic Gaudin magnet [8,14,20]. In order to make contact with the QDST dimer it is convenient to construct the generators of the algebra (3.1) in terms of new operators $p_\alpha, x_\alpha$ as

$$s_\alpha^1 = \tfrac{1}{4}(p_\alpha^2 + x_\alpha^2), \quad s_\alpha^2 = \tfrac{1}{4}(p_\alpha^2 - x_\alpha^2), \quad s_\alpha^3 = \tfrac{1}{4}\{p_\alpha, x_\alpha\} \tag{3.12}$$

with $p_\alpha, x_\alpha$ satisfying Heisenberg's commutation relations

$$[p_\alpha, x_\beta] = -i\delta_{\alpha\beta}, \quad [p_\alpha, p_\beta] = [x_\alpha, x_\beta] = 0, \quad \alpha, \beta = 1, 2. \tag{3.13}$$

One easily checks that for such a realization of the algebra (3.1) the Casimir operators have the value $s_\alpha^2 = -\frac{3}{16}, \alpha = 1, 2$ (we deal with the metaplectic representation of the su(1, 1) algebra, $k_\alpha = \frac{1}{4}, \frac{3}{4}$). Performing the canonical transformation

$$p_1 = \tfrac{1}{2}(b_1^\dagger + b_1 + b_2^\dagger + b_2), \quad x_1 = (b_1^\dagger + b_2^\dagger - b_1 - b_2)/2i,$$

$$p_2 = (-b_1^\dagger + b_2^\dagger + b_1 - b_2)/i, \quad x_2 = \tfrac{1}{2}(b_1^\dagger - b_2^\dagger + b_1 - b_2), \tag{3.14}$$

one obtains

$$H_1 + H_2 = -\tfrac{1}{2}(b_1^\dagger b_1 + b_2^\dagger b_2 + 1),$$

$$2(H_1 - H_2) = ((b_1^\dagger b_1 - b_2^\dagger b_2)^2 + \tfrac{1}{2})/(a - b) - (b_1^\dagger b_2 + b_2^\dagger b_1). \tag{3.15}$$

By comparing these expressions with (2.2), (2.1) we see that the QDST system is equivalent to the two site Gaudin magnet if $\omega_1 = \omega_2$ and $a - b = -4\varepsilon/\gamma$. From (3.3), using relations (3.12), (3.14) one gets the $L$-operator of the QDST dimer satisfying the linear $r$-matrix algebra (3.6).

## 4. The method of separation of variables for the QDST dimer

The Lax operator (3.3) introduced in the previous section does not admit the highest vector $|0\rangle$ for which

$$B(u)|0\rangle = 0, \quad \langle 0|B(u) = 0. \tag{4.1}$$

This is due to the presence in (3.4) of the $c$-number term $\tfrac{1}{2}\mathrm{col}(1,0,0)$. In the method of separation of variables eqs. (4.1) are replaced by the operator equation

$$B(u) = 0 \tag{4.2}$$

for the operator zeros $u_j$ of $B$. If we substitute (by the left) $u = u_j$ in (3.3), the $L$-operator acquires the form

$$L_j = L(u_j) = \begin{pmatrix} A(u_j) & 0 \\ \star & -A(u_j) \end{pmatrix} \tag{4.3}$$

from which we see that $\mathrm{i}v_j = A(u_j), j = 1,\ldots,n$, are the "operator eigenvalues" of $L(u)$. We aim at solving the spectral problem for the generating function of the integrals of motion, which in this case is given by the operator determinant

$$-L^2(u) = \det L(u). \tag{4.4}$$

Since the operation of taking the determinant of $L(u)$ commutes with the substitution $u \to u_j$ we can express (4.4) in $u_j, v_j$ variables as

$$L^2(u_j) = -v_j^2, \quad j = 1,\ldots,n. \tag{4.5}$$

By using the linear $r$-matrix algebra for the $L$-operator, one can prove the following commutators:

$$[v_j, u_k] = -\mathrm{i}\delta_{jk}, \quad [u_i, u_j] = [v_i, v_j] = 0, \tag{4.6}$$

i.e. the $u_j$ and $v_j$ are canonical operators. This fact allows to separate the eigenvalue problem for $L^2(u)$

$$L^2(u)\Psi = t(u)\Psi \tag{4.7}$$

into a set of one dimensional problems

$$-v_i^2 \psi_i(u_i) = t(u_i)\psi(u_i), \tag{4.8}$$

with

$$\Psi = \prod_i \psi_i(u_i). \tag{4.9}$$

In the following we show how this works for the QDST dimer. In this case the spectral problem has the form (see (3.9))

$$L^2(u)\Psi = t(u)\Psi \quad \Leftrightarrow \quad H_i\Psi = h_i\Psi,$$

$$t(u) = -\frac{\frac{3}{16}}{(u-a)^2} - \frac{\frac{3}{16}}{(u-b)^2} + \frac{h_1}{u-a} + \frac{h_2}{u-b} + \frac{1}{4}. \tag{4.10}$$

The separation variables are determined as the roots of the equation

$$B(u) = -L^1(u) + L^2(u) = 0. \tag{4.11}$$

In explicit form we have

$$\frac{x_1^2}{u-a} + \frac{x_2^2}{u-b} = 1, \tag{4.12}$$

with the roots $u_1, u_2$ given by the equations

$$u_1 + u_2 = a + b + x_1^2 + x_2^2, \quad u_1 u_2 = ab + bx_1^2 + ax_2^2.$$

Let us introduce the following notation:

$$\Theta(u) = \frac{x_1^2}{u-a} + \frac{x_2^2}{u-b} - 1 = -\frac{(u-u_1)(u-u_2)}{(u-a)(u-b)}. \tag{4.13}$$

Taking the residues in $u = a$ and $u = b$ at both sides of (4.13), we have

$$x_1^2 = \frac{(u_1-a)(u_2-a)}{b-a}, \quad x_2^2 = \frac{(u_1-b)(u_2-b)}{a-b}. \tag{4.14}$$

These elliptic coordinates $u_1, u_2$ in $\mathbb{R}^2$ satisfy the inequalities

$$a < u_1 < b < u_2 < \infty. \tag{4.15}$$

For each $u_j$ we define the additional variable $v_j$ as (the left substitution!)

$$v_j = -iA(u_j) = L^3(u_j) = \frac{1}{4(u_j-a)}\{x_1, p_1\} + \frac{1}{4(u_j-b)}\{x_2, p_2\}. \tag{4.16}$$

The matrix elements of the $L$-operator (3.3) in $u_i, v_i$ variables then have the form

$$B(u) = \frac{1}{2} - \frac{s_1^1 - s_1^2}{u - a} - \frac{s_2^1 - s_2^2}{u - b} = \frac{1}{2} \frac{(u - u_1)(u - u_2)}{(u - a)(u - b)}, \tag{4.17}$$

$$A(u) = \frac{i}{4} \frac{\{p_1, x_1\}}{u - a} + \frac{i}{4} \frac{\{p_2, x_2\}}{u - b} = 2iB(u) \left( \frac{1}{u - u_2} D_2 v_2 + \frac{1}{u - u_1} D_1 v_1 \right), \tag{4.18}$$

where

$$D_1 = \frac{(u_1 - a)(u_1 - b)}{u_1 - u_2}, \quad D_2 = \frac{(u_2 - a)(u_2 - b)}{u_2 - u_1}. \tag{4.19}$$

By using (4.19), (4.13) it is easy to show that

$$D_i = \left( \frac{x_1^2}{(u_i - a)^2} + \frac{x_2^2}{(u_i - b)^2} \right)^{-1} > 0, \quad i = 1, 2. \tag{4.20}$$

Note that the meromorphic operator-valued function $A(u)$ is obtained by interpolation with the data $A(u_j) = iv_j$ and the expression for $B(u)$ follows from the definition of the variables $u_j$. By equating the residues at $u = a, b$ in the r.h.s and l.h.s of (4.18) we obtain

$$\tfrac{1}{4}\{p_\alpha, x_\alpha\} = -x_\alpha^2 \left( \frac{1}{e_\alpha - u_2} D_2 v_2 + \frac{1}{e_\alpha - u_1} D_1 v_1 \right), \tag{4.21}$$

where $e_1 = a$ and $e_2 = b$.

The conjugation properties for the introduced operators have the form:

$$u_j^* = u_j, \quad D_j v_j = v_j^* D_j \tag{4.22}$$

from which we see that the operators $v_j$ are not self-adjoint. From the $v_j$ one can construct self-adjoint operators $w_j$ as follows:

$$w_j = \sqrt{D_j} v_j \frac{1}{\sqrt{D_j}}, \quad v_j = \frac{1}{\sqrt{D_j}} w_j \sqrt{D_j}, \tag{4.23}$$

and from (4.23), (4.6), (4.20) one can show that the operators $w_j, u_j$ are also canonical ones.

Substituting $u = u_j$ into (3.9) and taking into the account (3.7) we obtain two operator equalities

$$-v_j^2 = -\frac{\tfrac{3}{16}}{(u_j - a)^2} - \frac{\tfrac{3}{16}}{(u_j - b)^2} + \frac{1}{u_j - a} H_1 + \frac{1}{u_j - b} H_2 + \frac{1}{4}, \quad j = 1, 2. \tag{4.24}$$

Acting by the right and left hand sides of this operator equality on $\Psi$, which is a common eigenfunction of $H_1$ and $H_2$, we obtain

$$v_j^2 \Psi + t(u_j)\Psi = 0, \quad j = 1, 2, \tag{4.25}$$

where the function $t(u)$ is given in (4.10). In terms of $w_j$ the above equations take the form

$$w_j^2 \sqrt{D_j}\, \Psi + t(u_j)\sqrt{D_j}\, \Psi = 0. \tag{4.26}$$

Let us demand that the wave function $\Psi$ factorizes as

$$\Psi = \frac{1}{\sqrt[4]{D_1 D_2}} \exp[-\tfrac{1}{2}(u_1 + u_2)]\phi, \quad \phi = \phi_1(u_1)\phi_2(u_2), \tag{4.27}$$

where the partial functions $\phi_j(u_j)$ are defined on the corresponding intervals (4.15). Note that the factor $1/\sqrt[4]{D_1 D_2}$ in (4.27) is connected to the Jacobian of the transformation from the $x_j$ to the $u_j$ variables,

$$|\det \partial(x_1, x_2)/\partial(u_1, u_2)| = 1/\sqrt[4]{D_1 D_2}. \tag{4.28}$$

By substituting (4.27) in (4.26) we obtain the following separation equations:

$$-\phi''(u) + \left(1 - \frac{\tfrac{1}{2}}{u-a} - \frac{\tfrac{1}{2}}{u-b}\right)\phi'(u) + \left(\frac{\tilde{h}_1}{u-a} + \frac{\tilde{h}_2}{u-b}\right)\phi(u) = 0. \tag{4.29}$$

Here (and below) we have omitted the indices since the $\phi$'s satisfy equations of the same form, but on different intervals. Equation (4.29) is of confluent Heun's type (see, e.g. refs. [1,12]), having $a$ and $b$ as regular singularities and $\infty$ as irregular singularity and with the spectral parameters $\tilde{h}_j, j = 1, 2$, given by

$$\tilde{h}_1 = h_1 + \tfrac{1}{4} - 1/8(a-b), \quad \tilde{h}_2 = h_2 + \tfrac{1}{4} - 1/8(b-a). \tag{4.30}$$

It is worth to remark that although the $\phi's$ satisfy the same equation they are defined on different intervals. In order to get the eigenvalues we have to solve two boundary problems, one for the angular interval $a < u_1 < b$ and the other for the radial interval $b < u_2 < \infty$. Furthermore, both $\Psi$ and $\exp[\tfrac{1}{2}(-u_1 - u_2)]\phi$ are demanded to be square integrable in their variables.

Let us take the following form for the solution of eq. (4.29):

$$\phi(u) = (u-a)^{k_1}(u-b)^{k_2}\Phi(u), \quad \Phi(u) = \prod_{j=1}^{n}(u - \lambda_j), \tag{4.31}$$

where $k_i = 0, \tfrac{1}{2}$. These partial functions are polynomials multiplied in some cases by the square root of $u - a$ or/and $u - b$. By substituting (4.31) in (4.29) we get the following equation for $\Phi$:

$$-\Phi'' + \left(1 - \frac{2k_1 + \tfrac{1}{2}}{u-a} - \frac{2k_2 + \tfrac{1}{2}}{u-b}\right)\Phi' + \left(\frac{\tilde{\tilde{h}}_1}{u-a} + \frac{\tilde{\tilde{h}}_2}{u-b}\right)\Phi = 0, \tag{4.32}$$

where

$$\tilde{\tilde{h}}_1 = \tilde{h}_1 + k_1 - \frac{1}{2}\frac{k_1 + k_2 + 4k_1 k_2}{a-b}, \quad \tilde{\tilde{h}}_2 = \tilde{h}_2 + k_2 + \frac{1}{2}\frac{k_1 + k_2 + 4k_1 k_2}{a-b}. \tag{4.33}$$

Equation (4.32) has just polynomial solutions with the zeros $\lambda_j$ satisfying the nonlinear equations

$$\sum_{j \neq i}\frac{1}{\lambda_i - \lambda_j} - \frac{1}{2} + \frac{k_1 + \tfrac{1}{4}}{\lambda_i - a} + \frac{k_2 + \tfrac{1}{4}}{\lambda_i - b} = 0, \quad i = 1, \dots, n. \tag{4.34}$$

It is possible to prove that the zeros $\lambda_i$ have the following properties (generalized Stieltjes theorem [22]): (i) they are simple, (ii) they are placed along the real axis inside the intervals (4.15), (iii) they are the critical points of the function $|G|$,

$$G(\lambda_1,\ldots,\lambda_n) = \exp[-\tfrac{1}{2}(\lambda_1 + \cdots + \lambda_n)] \prod_{p=1}^{n} (\lambda_p - a)^{k_1 + 1/4}(\lambda_p - b)^{k_2 + 1/4} \prod_{r>p}(\lambda_r - \lambda_p). \quad (4.35)$$

To calculate the eigenvalues $\tilde{h}_i$ we must first solve eqs. (4.34) and then use the formulae

$$\tilde{h}_1 = (2k_1 + \tfrac{1}{2}) \sum_{j=1}^{n} \frac{1}{a - \lambda_j}, \quad \tilde{h}_2 = (2k_2 + \tfrac{1}{2}) \sum_{j=1}^{n} \frac{1}{b - \lambda_j},$$

$$\tilde{h}_1 + \tilde{h}_2 = -n = h_1 + h_2 + k_1 + k_2 + \tfrac{1}{2}. \quad (4.36)$$

From (4.36) we see that the spectrum of $H_1 + H_2$ has the form

$$\langle H_1 + H_2 \rangle = h_1 + h_2 = -\tfrac{1}{2}\langle b_1^{\dagger}b_1 + b_2^{\dagger}b_2 \rangle - \tfrac{1}{2} = -(n + k_1 + k_2 + \tfrac{1}{2}) \quad (4.37)$$

and therefore the spectrum of $N$ is

$$\langle N \rangle = 2(n + k_1 + k_2). \quad (4.38)$$

For each fixed $n$, $k_1$ and $k_2$ (i.e. for each fixed $\langle N \rangle$) we have $n + 1$ different eigenfunctions $\Phi(u, \{\lambda_k\}_{k=1}^{n})$ which differ by the distribution of their zeros among two intervals: $(a, b)$ and $(b, \infty)$. The $k$th polynomial has $k$ zeros on the first interval and $n - k$ zeros on the second ($k = 0, \ldots, n$) and for fixed $\langle N \rangle$ we have $\langle N \rangle + 1$ eigenfunctions.

Let us take the alternate representation for the polynomial $\Phi(u)$

$$\Phi(u) = \sum_{j=0}^{n} c_j(u - a)^j, \quad c_n = 1. \quad (4.39)$$

Equivalently one can take the expansion near another singular point $u = b$. If we denote $u - \cdot = z$, $a - b = \tilde{\varepsilon}$, $h = \tilde{h}_1 + \tilde{h}_2 = -n$, $g = \tilde{\varepsilon}\tilde{h}_1$, we have

$$-z(z + \tilde{\varepsilon})\Phi''(z) + [z^2 + z(\tilde{\varepsilon} - 2k_1 - 2k_2 - 1) - (2k_1 + \tfrac{1}{2})\tilde{\varepsilon}]\Phi'(z) + (zh + g)\Phi(z) = 0,$$

which gives the following three-term recurrent relations for the $c_j$

$$c_j[g - j(j + 2k_1 + 2k_2 - \tilde{\varepsilon})] + c_{j-1}(j - 1 - n) - c_{j+1}\tilde{\varepsilon}(j + 1)(j + 2k_1 + \tfrac{1}{2}) = 0,$$

$$j = 0, \ldots, n \quad (4.40)$$

with $c_{-1} = c_{n+1} = 0$. This equation can be rewritten in the form

$$Ac = 0, \quad (4.41)$$

with $A$ being a three-diagonal matrix which, after symmetrization, reduces to a Jacobi matrix. The procedure of finding the $n + 1$ eigenvalues turns to the problem of calculating of zeros of the characteristic polynomial $\det A = 0$. We remark here that the above prosedure of passing form the solutions $\lambda_j$ of the (4.34) to the polynomial (4.39) coefficients is analogous to the described in the section 2 derivation of the energy levels on the background of Bethe ansatz for the quadratic $R$-matrix algebra.

It is easy to get the expressions of $\Psi$ in terms of the initial variables $x_i$. Let us dispatch the factor $1/\sqrt[4]{D_1 D_2}$ to the weight when integrating upon the $u_j$ variables, so in $x_j$ variables the wave function $\Psi$ has the form

$$\Psi(x_1, x_2) = \exp[-\tfrac{1}{2}(x_1^2 + x_2^2)]x_1^{2k_1}x_2^{2k_2} \prod_{i=1}^{n} \Theta(\lambda_i), \tag{4.42}$$

where $\Theta(\lambda_i)$, $i = 1, \ldots, n$, is defined in (4.13) and $\{\lambda_i\}_{i=1}^{n}$ satisfy eqs. (4.34). One can verify that $\Psi$ is an eigenfunction of the two commuting operators $H_1$ and $H_2$ labelled by the quantum numbers $n, k_1$ and $k_2$ and by an additional index enumerating the $n + 1$ different sets of solutions for $\{\lambda_j\}_{j=1}^{n}$. If $x_2 = 0$, we have that (4.42) reduces to the usual Hermite polynomials, so our functions are generalized Hermite plynomials or confluent Heun's polynomials [1,12,17,24]. Heun's polynomials were studied in ref. [17] and the confluent Heun's polynomials first appeared in connection with the two-dimensional Coulomb problem [12] and is remarkable that they have also appear in connection with the eigenfunctions of the QDST dimer.

## 5. su(2) algebra and QDST dimer

In this section we show that the QDST dimer is naturally related to the su(2) algebra. We use this fact to check the results of the previous section. To this end, let us introduce the operators

$$t_3 = \tfrac{1}{2}(b_1^\dagger b_1 - b_2^\dagger b_2), \quad t_+ = t_1 + it_2 = b_1^\dagger b_2, \quad t_- = t_1 - it_2 = b_2^\dagger b_1, \tag{5.1}$$

with the operators $b_i, b_i^\dagger$ satisfying the commutation relations $[b_i, b_j^\dagger] = \delta_{ij}$ $[b_i, b_j] = [b_i^\dagger, b_j^\dagger] = 0$. The operators (5.1) are the generators of the su(2) algebra

$$[t_3, t_\pm] = \pm t_\pm, \quad [t_+, t_-] = 2t_3. \tag{5.2}$$

The Casimir operator is given by

$$c_2 = t_3^2 + \tfrac{1}{2}t_+ t_- + \tfrac{1}{2}t_- t_+ = s(s + 1), \tag{5.3}$$

where $s$ is the eigenvalue of the operator $\tfrac{1}{2}(b_1^\dagger b_1 + b_2^\dagger b_2)$. In terms of the generators (5.2) the Hamiltonian of the QDST dimer becomes

$$H = -\gamma t_3^2 + \Delta t_3 - 2\varepsilon t_1 - s(\gamma s + \gamma - (\omega_1 + \omega_2)), \tag{5.4}$$

where $\Delta = \omega_1 - \omega_2$ and the average of the particle number operator equals

$$\langle N \rangle = 2s. \tag{5.5}$$

If we take the unitary irreducible representation of the su(2) algebra in the vector space $\mathbb{C}^{2s+1}$

$$t_3|s,m\rangle = m|s,m\rangle, \tag{5.6}$$

$$t_\pm|s,m\rangle = \sqrt{(s \mp m)(s \pm m + 1)}|s,m \pm 1\rangle, \tag{5.7}$$

where

$$s = 0, \tfrac{1}{2}, 1, \ldots = \tfrac{1}{2}n, \tag{5.8}$$

then the spectrum of $H$ is simply obtained by diagonalizing a three diagonal matrix. These matrices were written down explicitly in a number of cases by Scott [18].

Here we show that the eigenfunctions $\Psi$ of $H$ and $N$ are the confluent Heun's polynomials, in agreement with the result of the previous section. To this end we consider the following realization of the su(2) algebra in the space of holomorphic functions:

$$t_3 = x\partial_x - s, \quad t_- = \partial_x, \quad t_+ = -x^2\partial_x + 2sx. \tag{5.9}$$

The operators $t_+$ and $t_-$ are adjoint to each other in the scalar product $(x \in \mathbb{C})$,

$$\langle\varphi|\psi\rangle = \frac{2s + 1}{\pi} \int \frac{\overline{\varphi(x)}\psi(x)}{(1 + |x|^2)^{2+2s}} \, d^2x, \tag{5.10}$$

while $t_3$ is a self-adjoint operator. The basis vectors $|s, m\rangle$ in this representation have the form $f_m = c_m x^{s+m}$ with $c_m = [(2s)!/(s - m)!(s + m)!]^{1/2}$.

It is convenient to realize the algebra (5.9) in the space of functions $x^{-s}\Psi(x)$ with $x = z + \sqrt{z^2 - 1}$. Then the generators are

$$\tilde{s}_3 = \sqrt{z^2 - 1}\,\partial_z, \quad \tilde{s}_\pm = (z \pm \sqrt{z^2 - 1})(s \mp \sqrt{z^2 - 1}\,\partial_z). \tag{5.11}$$

In terms of $z$ the spectral problem $H\Phi(z) = h\Phi(z)$ becomes

$$\Phi'' + \left(\frac{z}{z^2 - 1} - 2\frac{\varepsilon}{\gamma} - \frac{\Delta}{\gamma\sqrt{z^2 - 1}}\right)\Phi'$$
$$+ \frac{1}{z^2 - 1}\left(2\frac{\varepsilon s}{\gamma}z + s + 1 - \frac{\omega_1 + \omega_2}{\gamma} - \frac{h}{\gamma}\right)\Phi = 0. \tag{5.12}$$

This equation, in the case $\Delta = 0$, coincides with eq. (4.29) with $a = 2\varepsilon/\gamma, b = -2\varepsilon/\gamma$.

We finally remark that eq. (5.12), rewritten in the $y = \ln x$ variables, gives the quasi-exactly solvable system considered in ref. [9].

## 6. Conclusions

We have shown that two different representations for the wave functions of the QFₜ₁ dimer can be constructed. The first one is connected with the quadratic $R$-matrix algebra and tₙ second one with the linear $r$-matrix algebra. These two representations must be connected by some unitary transformation. It is of interest to find this transformation explicitly, since it would elucidate the link between the standard algebraic BA and Baxter equation (see [21]) and the polynomial solutions of the

separation equations. This question is presently under investigation. A further interesting problem under investigation is the application of the method of separation of variables to $q$-deformations of the QDST dimer. This would give the possibility of introducing a new class of special functions i.e. the $q$-deformed Heun's polynomials.

## Aknowledgements

## References

[1] H. Bateman and A. Erdelyi, Higher Transcendental Functions, Vol. 3 (McGraw-Hill, New York, 1955).

[2] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, The discrete self-trapping equation, Physica D 16 (1975) 318–338.

[3] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, Soliton structure in crystalline acetanilide, Phys. Rev. B 30 (1984) 4703–4712.

[4] V.Z. Enol'skii and M. Salerno, On the calculation of the energy spectrum of quantum integrable systems, Phys.Lett. A 155 (1991) 121–125.

[5] V.Z. Enol'skii, M. Salerno, N.A. Kostov and A.C. Scott, Alternate quantization of the discrete self-trapping dimer, Phys. Scripta 43 (1991) 229–235.

[6] V.Z. Enol'skii, M. Salerno, A.C. Scott and J.C. Eilbeck, There is more than one way to skin a Schrödinger cat, Physica D 59 (1992) 1–24.

[7] L.D. Faddeev, in: Recent Advances in Field Theory and Statistical Mechanics, Proc. Les Houches Summer School XXXIX, eds. J.B. Zuber and R. Stora (North-Holland, Amsterdam, 1984) pp. 719–756.

[8] M. Gaudin, La Fonction d'Onde de Bethe (Masson, Paris, 1983).

[9] A. Gonzáles-López, N. Kamran and P.J. Olver, Normalizability of One-dimensional quasi-exactly solvable Schrödinger operator, preprint (1992).

[10] E.W. Hobson, The theory of spherical and ellipsoidal harmonics (Chelsea, 1985).

[11] I.V. Komarov and V.B. Kuznetsov, Quantum Euler–Manakov top on the 3-sphere $S_3$, J. Phys. A 24 (1991) L737–L742.

[12] I.V. Komarov, L.I. Ponomarev and S.Yu. Slavyanov, Spheroidal and Coulomb Spheroidal Functions (Nauka, Moscow, 1976) [in Russian].

[13] P.P. Kulish and E.K. Sklyanin, Integrable Quantum Field Theories, Lecture Notes in Physics, Vol. 151, eds. J. Hietarinta and C. Montonen (Springer, Berlin, 1982) pp. 61–119.

[14] V.B. Kuznetsov, Quadrics on real Riemannian spaces of constant curvature, separation of variables and connection with Gaudin magnet, J. Math. Phys. 33 (1992) 3240–3254.

[15] V.B. Kuznetsov, Equivalence of two graphical calculi, J. Phys. A 25 (1992).

[16] W.D. Niven, On ellipsoidal harmonics, Phil. Trans. 182A (1892) 231–278.

[17] J. Patera and P. Winternitz, A new basis for the representation of rotation group: Lamé and Heun polynomials J. Math. Phys. 14 (1973) 1130–1139.

[18] A.C. Scott, A non resonant discrete self-trapping system, Phys. Scripta 42 (1990) 14–19.

[19] A.C. Scott, P.S. Lomdahl and J.S. Eilbeck, Between the local mode and normal mode limits, Chem. Phys. Lett. 113 (1985) 29–36.

[20] E.K. Sklyanin, Separation of variables in the Gaudin model, J. Soviet Math. 47 (1989) 2473–2488.

[21] E.K. Sklyanin, Separation of variables in the classical integrable $SL(3)$ magnetic chain, preprint RIMS-871, Kyoto (1992).

[22] T.J. Stieltjes, Sur certain polynômes qui verifient une équation différentielle, Acta Math. 6/7 (1885) 321–326.

[23] G. Szegö, Ortogonal polynomials, Am. Math. Soc. (1967).

[24] E.T. Whittaker and G.N. Watson, A Course of Modern Analysis (Cambridge Univ. Press, Cambridge, 1927).

PHYSICA D

# An ecumenical nonlinear von Neumann equation: fluctuations, dissipation, and bifurcations

V.M. Kenkre

*Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131, USA*

An evolution equation is presented to describe the transport of a quantum mechanical quasiparticle such as an electron, or electronic or vibrational excitation, interacting strongly with lattice vibrations. A generalization of the discrete nonlinear Schrödinger equation to incorporate dissipation and fluctuation effects arising from interactions with a thermal reservoir, the equation predicts a multitude of interesting phenomena including bifurcations.

## 1. Introduction

Alwyn Scott has initiated, performed, stimulated, and guided, an amazing amount of current research in nonlinear physics. The community of nonlinear scientists owes much to him. It is a pleasure and an honour to be asked to make this contribution to the Festschrift on the occasion of his sixtieth birthday. The range of Alwyn's work in nonlinear science is vast. It spans a variety of systems and approaches, and deals with practical matters such as the problem of launching solitons and the spectra of acetanilide, as well as formal matters such as the dynamics of nonlinear systems and the generalization of nonlinearities [1–4].

The subject of the present article is the *ecumenical* nonlinear von Neumann equation

$$
i\hbar \frac{d\rho_{mn}}{dt} = [V, \rho]_{mn} - \chi(\rho_{mm} - \rho_{nn})\rho_{mn}
$$
$$
- i\frac{\chi}{\Gamma} \rho_{mn}([V, \rho]_{mm} - [V, \rho]_{nn})
$$
$$
- i\alpha(1 - \delta_{mn})(\rho_{mn} - \rho_{mn}^{eq}). \tag{1}
$$

Before describing the *what*, *whence*, and *whither* of this equation, it is relevant to comment on the

relation of the equation to Alwyn Scott. There exists a nonlinear transport equation which goes under the name of the discrete nonlinear Schrödinger equation (DNLSE), whose study has occupied many investigators, including the present author [5–8], for several years. Equation (1) is the augmented form taken on by the DNSLE, when the latter is opened to heat reservoirs and prepared for stochastic activities through the introduction of dissipation and fluctuation. The DNLSE, on the other hand, is identical in form to the so-called discrete self-trapping equation (DSTE), introduced earlier, and studied extensively, by Alwyn Scott and his collaborators.

While identical in form, and therefore in the relevance of a number of results such as concerning stationary states, the DSTE and the DNLSE differ significantly in physical meaning. This is evident from the role played by quantum mechanics in the two equations. The complex nature of the amplitudes in the DNLSE are a requirement of quantum mechanics and the study of the interplay of phases and nonlinearity is an undertaking of direct physical import. The complex nature of the amplitudes in the DSTE, on the other hand, is a convenience of the type familiar in electromagnetism where fields, which are actually real, are described by complex

quantities for computational ease. There is, thus, work on the "quantization" of the DSTE, which appears to have no counterpart at all in the context of the DNLSE. Also, the two equations differ in physical origin. We begin with a brief motivational discussion of the microscopic origin of the DNLSE in the next section. After a quick review of earlier work on the DNLSE in the following section, we introduce noninfinite relaxation and thermal fluctuations, and arrive at the ecumenical equation (1), which has been so termed because it is capable of resolving the competition and conflicts of nonlinearity and damping, or of fluctuation and dissipation, since it unifies the description of *all* these effects in one fell swoop. The fascinating consequences of the augmented equation will occupy us in the rest of the article. The subject under investigation is the transport of a low or intermediate mobility quasiparticle under strong interactions with vibrations. Examples are an electron in a narrow-band material, a vibrational or electronic excitation in a polymer, and a light interstitial such as a proton or muon in a metal. The notation in (1) is standard: $\rho$ represents the density matrix of the quasiparticle, the matrix elements are taken in the representation of some localized states such as a Wannier set, and the four terms on the right side arise from intersite transfer, nonlinearity stemming from interactions with vibrations, finiteness of vibrational relaxation, and thermal fluctuations, respectively.

## 2. Physical motivation of the DNLSE

Consider a moving quasiparticle described by the ket $|\Psi(t)\rangle$ whose time evolution is governed by the Hamiltonian $H$ through $i\hbar(d|\Psi(t)\rangle)/dt = H|\Psi(t)\rangle$, i.e., the standard Schrödinger equation. On multiplying the equation by a Wannier-like localized bra $\langle m|$, one obtains

$$i\hbar \frac{dc_m}{dt} = \sum_n V_{mn}c_n + E_m c_m , \qquad (2)$$

for the amplitudes $c_m$ where the $E$'s and the $V$'s are, respectively, the diagonal and off-diagonal matrix elements of the Hamiltonian in the localized basis. If interactions with vibrations are present, the $V$'s and $E$'s are dependent on the vibrational coordinates $x$. For simplicity, let us take $V_m$ to be independent of $x$, and $E_m$ to depend linearly only on the internal coordinate $x_m$ at site $m$. In the absence of interactions, $x_m$ might have evolved sinusoidally with frequency $\omega$ and equilibrium position 0. It might naturally obey, in the presence of interactions,

$$\frac{d^2 x_m}{dt^2} + \omega^2 x_m = -\text{const.} \times |c_m|^2 , \qquad (3)$$

where the equilibrium position of the oscillator is changed by an amount proportional to the probability that site $m$ is occupied by the quasiparticle. In the presence of a time scale disparity, if the vibrations are slaved by the quasiparticle probabilities, the time derivatives in (3) may be put equal to zero. The vibrational coordinate is then proportional to the quasiparticle occupation probability, and we obtain the DNLSE

$$i\hbar \frac{dc_m}{dt} = \sum_n V_{mn}c_n - \chi |c_m|^2 c_m . \qquad (4)$$

The above description of the origin of the DNLSE is meant to convey only its essential physical content. While we do not wish to discuss here the many subtleties which the microscopic derivation of the DNLSE entails [9,10], we now give a brief description of how the DNLSE can be obtained through a *semiclassical approximation* from the Hamiltonian evolution of the standard model of a quasiparticle interacting strongly with vibrations, viz.,

$$H = \sum_m \varepsilon_m a_m^\dagger a_m + \sum_{m,n} V_{m,n} a_m^\dagger a_n$$

$$+ \sum_q \hbar\omega_q (b_q^\dagger b_q + \tfrac{1}{2})$$

$$+ N^{-1/2} \sum_q \hbar\omega_q g_q$$

$$\times \exp(i q \cdot R_m) (b_q + b_{-q}^\dagger) a_m^\dagger a_m . \qquad (5)$$

Equation (5) uses standard notation which we will not detail here, except for stating that $a$ and $b$ refer, respectively, to the quasiparticle and the vibrations with which it interacts strongly. With

$$\hat{p} = a_1^\dagger a_1 - a_2^\dagger a_2 , \tag{6a}$$

$$\hat{q} = -i(a_1^\dagger a_2 - a_2^\dagger a_1) , \tag{6b}$$

$$\hat{r} = a_1^\dagger a_2 + a_2^\dagger a_1 , \tag{6c}$$

$$\hat{y} = -(b^\dagger + b)/2g , \tag{6d}$$

the Hamiltonian (5) can be written, for a simple two-site symmetric system (a dimer) interacting with a single vibrational coordinate, as

$$H = V\hat{r} + g\omega\hat{p}(b^\dagger + b) + \omega(b^\dagger b + \tfrac{1}{2}) . \tag{7}$$

Here $V$ denotes the interaction matrix element for the quasiparticle transfer between the two sites of the dimer, $g$ measures the coupling strength with the vibration of frequency $\omega$, and we have put the (identical) site energy of the two sites equal to zero. Henceforth we put $\hbar = 1$. Equation (7) then results in the following evolution for the quantities defined in (6):

$$\frac{d\hat{p}}{dt} = 2V\hat{q} . \tag{8a}$$

$$\frac{d\hat{q}}{dt} = -2V\hat{p} - \chi\hat{r}\hat{y} , \tag{8b}$$

$$\frac{d\hat{r}}{dt} = \chi\hat{q}\hat{y} , \tag{8c}$$

$$\frac{d^2\hat{y}}{dt^2} = -\omega^2(\hat{y} - \hat{p}) . \tag{8d}$$

Let us now take the expectation values of the operators in an initial state $\psi$, and denote these values by removing the circumflexes, i.e., $\zeta = \langle \psi|\hat{\zeta}|\psi\rangle$. The single assumption that $\hat{y}$ is classical converts the exact dynamics into the DNLSE, as it is trivial to show [6–8] that eqs. (8) rewritten without the circumflexes represent the DNLSE for the two-site system, the $p$, $q$ and $r$ being nothing other than the standard Feynman den-

sity matrix element combinations of the two-site system (dimer):

$$p = \rho_{11} - \rho_{22} , \quad q = i(\rho_{12} - \rho_{21}) ,$$

$$r = \rho_{12} + \rho_{21} . \tag{9}$$

While simple, the foregoing demonstration makes transparent what crucial assumption is responsible for the passage from microscopics to evolution equations such as the DNLSE. It is clearly the assumption that the vibrations may be considered classical. This fact, while well known to many, has often gone unappreciated. This has led to numerous incorrect, although wishful, assertions on the one hand and meaningless queries on the other.

## 3. Nonlinear capture and trimer, $N$-mer evolution

Before proceeding with the process of the augmentation of the DNLSE into the ecumenical form (1), i* is of interest to recall the large number of interesting and useful results which have emerged from the DNLSE in a variety of situations. Space limitations force us to do no more than list the primary contexts of some of that work, along with some slightly greater detail on two specific areas, viz. nonlinear trapping [11,12] and exact analysis of some spatially extended systems [13,14]. For further description, the reader may refer to several more detailed reviews [6–8].

The following is a list of some early work done on the DNLSE, i.e., from the infinite relaxation, zero temperature limit of (1):

(1) Exact solutions for the dynamics of the two-state system for arbitrary initial conditions and the elucidation of polaronic motion and self-trapping on the basis of those solutions [5].

(2) Application to fluorescence depolarization, wherein the moving quasiparticle is an electronic excitation and the observable is the

intensity of light emitted with different polarizations [6,15].

(3) Application to neutron scattering of hydrogen trapped around impurities such as oxygen in metals such as niobium, where the moving quasiparticle is the proton [16].

(4) Application to muon spin relaxation in antiferromagnetic solids such as bcc iron wherein the quasiparticle is a muon moving within a solid, and the observable is spin polarization [7,17].

(5) Generalization of the DNLSE to incorporate anharmonic potentials and nonlinear restoring forces, resulting in nonlinearities other than bilinear, and the appearance, and counterintuitive disappearance, of multiple stationary states [6,18].

(6) Calculation of memories in nonlinear generalized master equations, and the development of a perturbative scheme which is exact in the nonlinearity but perturbative in the intersite transfer [19].

(7) Studies of the interplay of nonlinearity and disorder on the basis of ensembles with various distributions of the nonlinearity parameter [6,7].

(8) Theory of nonlinear trapping of excitation directed at fluorescence quenching in molecular aggregates [11,12].

(9) Analytic solutions for a restricted class of initial conditions in some spatially extended systems such as trimers and symmetric $N$-mers [13,14].

We refer briefly to the last two of these below.

### 3.1. Nonlinear trapping

Quasiparticle trapping is an important phenomenon, and is of particular interest in areas of investigation such as photosynthesis, in which the harvesting of energy necessary for the operation of the reaction centers is followed by the process of the transfer of the harvested energy to the reaction center. The excitation carrying

energy, but no matter, is the quasiparticle, and the reaction center is the trap [20,21]. Let us assume that an excitation moves on a chain via nearest neighbour interactions $V$, and trapped by a site which has the nonlinear behavior arising from strong interactions with vibrations leading to the nonlinearity described by the cubic term in the DNLSE. Of the several possible models of capture, consider two: one in which one of the sites in the chain is itself the trap site and possesses the cubic nonlinearity, and another in which the excitation moves in a region of space called the antenna and communicates with a trap which is external to the antenna. In a simple example of the latter situation, the $N$ antenna sites all communicate equally with the external trap through a matrix element $W$, while transferring excitation among themselves through nearest neighbour elements $V$. The first model is represented by

$$i\frac{dc_m}{dt} = V(c_{m+1} + c_{m-1}) - \delta_{m,0}\chi|c_0|^2 c_0 , \qquad (10)$$

in obvious notation, while the second is represented by

$$i\frac{dc_m}{dt} = V(c_{m+1} + c_{m-1}) + Wc_\theta , \qquad (11a)$$

$$i\frac{dc_\theta}{dt} = W\sum_m c_m - \chi|c_\theta|^2 c_\theta . \qquad (11b)$$

Numerical calculations by Dunlap, Kenkre and Reineker [11] on the first model have shown that a transition appears to occur as the nonlinearity parameter $\chi$ crosses the value $(3.2)V$. The second model has been solved analytically by Kenkre and Kuš [12] who have shown that the probability of the trap site, which we will call $P$, obeys

$$\frac{d^2 P}{d\tau^2} = \frac{1}{2}(I_1 - \eta I_2) - (\xi I_2 + \eta^2 + 1)P$$
$$- 3\xi\eta P^2 - 2\xi^2 P^3 , \qquad (12)$$

where, with the notation that $w$ is defined as $\sqrt{N}\,W$, the scaled time $\tau$ and the parameters $\xi$

and $\eta$ are $\tau = 2wt$, $\xi = \chi/4w$, and $\eta = V/w$, the $I$'s being constants of the motion. Clearly, $\xi$ measures the amount of nonlinearity relative to the (scaled) interaction matrix element for transfer of excitation between the antenna and the trap while $\eta$ measures the motion of the excitation within the antenna relative to the extent of its transfer from the antenna to the trap. Equation (12) can be solved in terms of Weierstrassian elliptic functions. A completely general solution to the problem has been obtained for arbitrary initial conditions, and a number of physical features have been extracted [12]. The analysis has also been generalized to finite relaxation and finite temperature situations.

### 3.2. Spatially extended systems: analytic results

The methods of analysis used in the early work [5] to obtain exact solutions in two-site systems were generalized by Andersen and Kenkre to some larger systems including trimers and a class of extended systems termed $N$-mers [13,14]. They obtained the explicit time dependence for the trimer for a class of special initial conditions, discussed a transition at $\chi = -6V$ as well as the effect of the sign of the nonlinearity, and showed the connection of their results to the trimer stationary states obtained earlier by Eilbeck et al. [3]. They generalized the work to $N$-mers, also obtaining explicit analytical solutions. The systems considered were site-degenerate with $V_{mn} = V$ between any two sites, the initial excitation being localized on a single site, or more generally, distributed equally among all sites of one of two groups, one of $m_A$ sites and the other of $m_B = N - m_A$ sites. The calculational trick of Andersen and Kenkre [13,14] consists of the generalization of (9) to define new quantities $p$, $q$, and $r$, which allow one to convert the trimer or $N$-mer problem as described by

$$\frac{dc_m}{dt} = -iV\left(-c_m + \sum_n c_n\right) + i\chi|c_m|^2 c_m, \qquad (13)$$

into a completely tractable non-degenerate dimer problem [22] involving the solution of

$$\frac{d^2p}{dt^2} = -2\gamma_0 - 6\gamma_1 p - 6\gamma_2 p^2 - 2\gamma_3 p^3. \qquad (14)$$

Some of this work is reappearing in the literature and is being applied to experimental situations.

## 4. Finite relaxation and thermal fluctuations

How do we augment the DNLSE (4) to treat situations *not* involving time scale disparities between the quasiparticle and the vibrations, to eliminate unrealistic assumptions such as that of infinite vibrational relaxation, and to include unavoidable, and important, interactions with reservoirs which give rise to temperature effects and fluctuations? This is the question that we now address. Equation (4) is the result of a time disparity assumption made on (2) and (3). We now retain (2), and replace (3) by

$$\frac{d^2x_m}{dt^2} + \gamma\frac{dx_m}{dt} + \omega^2 x_m = -\text{const.} \times |c_m|^2 + R_m(t), \qquad (15)$$

and explore the consequences of dissipation introduced by the rate $\gamma$, and of fluctuation caused by the random force term $R_m(t)$. For simplicity, we have used the restriction that the reservoir interaction occurs only with the vibrational system. Two limits of (15) are particularly interesting: zero damping, and extremely large damping. Some exact solutions can be found for the former case [23] for the dimer. The probability difference $p(t)$ shows cn or dn behavior as in the adiabatic solutions [5] and undergoes a characteristic new transition into a region where it equals the sum of a part which is proportional to the appropriate elliptic function (cn or dn), and a part which is proportional to the cube of the elliptic function. Here, we will focus on the opposite limit of extremely large damping.

If the damping is large enough to justify the

neglect of the second derivative of the oscillator displacements, more formally, if $\omega \to \infty$, $\gamma \to \infty$, $\omega^2/\gamma = \Gamma$, the evolution of the oscillator displacements towards their equilibrium positions now possesses a single characteristic "vibrational relaxation" rate $\Gamma$. Restricting the analysis to the two-site system in the interest of full tractability, we obtain from (2) and the large damping limit of (15),

$$\frac{dp}{dt} = 2Vq \,, \quad \frac{dq}{dt} = -2Vp - \chi yr \,, \quad \frac{dr}{dt} = \chi yq \,,$$
(16a)

$$\frac{dy}{dt} = -\Gamma(y - p) + F(t) \,, \quad F(t) = 2\Gamma(kT/\omega^2)\,\delta(t) \,,$$
(16b)

where $k$ is the Boltzmann constant.

Equations (16) constitute a Langevin set. Standard techniques allow the derivation of an exact Fokker–Planck equation for the distribution function in $p$, $q$, $r$, $y$ space and of an approximate but useful Fokker–Planck equation for the distribution function $\sigma(p, q, r; t)$ in $p$, $q$, $r$ space alone. The latter is obtained [24] through the application of projection techniques which eliminate the vibrational variable $y$, and is valid in the high damping limit. The Fokker–Planck equation can be solved exactly for its stationary state distribution function, and the formalism can be used in two separate ways: (i) to perform a Kramer's first passage time analysis aimed at an investigation of the stability (against thermal fluctuations) of the nonlinear structures inherent in the DNLSE, and (ii) to carry out a contraction analysis [25] from the Fokker–Planck equation in order to arrive at a closed equation such as (1) for the quasiparticle variables. We refer the reader elsewhere [24] for (i), and concentrate on the results of (ii) here. The contraction analysis has, as its consequence,

$$\frac{dp}{dt} = 2Vq \,,$$
(17a)

$$\frac{dq}{dt} = -2Vp - \chi pr + \frac{2V\chi}{\Gamma} qr - \alpha q \,,$$
(17b)

$$\frac{dr}{dt} = \chi pq - \frac{2V\chi}{\Gamma} q^2 - \alpha(r - r_{eq}) \,.$$
(17c)

Here, $r_{eq}$ is the thermal equilibrium value of $r$, and $\alpha$ is a rate which attempts to drive the system to the thermal state. A high temperature expression [25] for this rate is $\alpha = (2\chi/\Gamma)kT$.

The set of equations (17) have been arrived at by extending the DNLSE to nonadiabatic, finite temperature situations through a blend of analytic arguments and physical assumptions. The system they describe reduces to the trivial linear dimer if $\chi = 0 = \alpha$, to the high-temperature damed linear dimer if $\chi$ vanishes but $\alpha$ does not, to the nonlinear adiabatic dimer if $\chi$ is finite but $\Gamma$ is infinite and $\alpha$ vanishes, to a relatively crude extension [26] of the nonlinear dimer to dissipative situations if $\chi$ is finite but $\Gamma$ is infinite and $\alpha$ vanishes, and to the nonlinear nonadiabatic dimer if $\chi$ and $\Gamma$ are finite and $\alpha$ vanishes. The latter case displays a rich multitude of phenomena including a fascinating interplay of quantum phases and nonlinearity, for which we refer the reader to refs. [27]. Evolution showing a complete combination of the elliptic function evolution of the adiabatic dimer for short times, followed by a self-trapping swing into the localized stationary states of the dimer for longer times, which is itself followed by a delocalization and symmetrical spreading over the two sites characteristic of thermal fluctuations, has been exhibited and commented on in ref. [25] for the high-temperature case when $r_{eq}$, which can generally be taken to equal tanh($V/kT$), vanishes. Exciting new behavior, which occurs when $r_{eq}$ does *not* vanish, will now be mentioned here briefly.

For vanishing $\alpha$, the probability difference $p$ oscillates and then tends to the stationary value which is 0 if the nonlinearity parameter is small enough and finite (corresponding to a localized state) if it is large enough. As $\alpha$ increases, the detrapping effect is seen: $p$ tends to 0 at larger times even for large nonlinearities. As $\alpha$ increases further, a surprising burst of $p$ occurs for

a short time, and the burst recurs after a time period. The bursts become more frequent with further increase of $\alpha$ and behavior that *appears* chaotic occurs. Phase space plots in the $p-q$ plane show that a limit cycle has been reached at this point, as $p$ (as well as $q$) oscillates steadily between two finite values (see (iv) and (v) in fig. 1). A further increase in $\alpha$ destroys the limit cycle, and stable dissipative behavior is recovered (case (vi)): $p$, $q$ tend to vanishing values while $r$ tends to $r_{eq}$. Stability analysis carried out

by Kenkre and Kuš [26] shows that the destruction of limit cycles occurs for $\alpha > r_{eq}\chi/\Gamma$, i.e., for temperatures exceeding that at which $V/kT$ becomes smaller than its hyperbolic cotangent if $\chi$ and $\Gamma$ are assumed to be independent of temperature.

This bifurcation behavior is reflected in exciting predicted features of several observable quantities [26]. An example is the degree of fluorescence polarization in poly-L-proline oligomers of variable length, which is given [6] by a
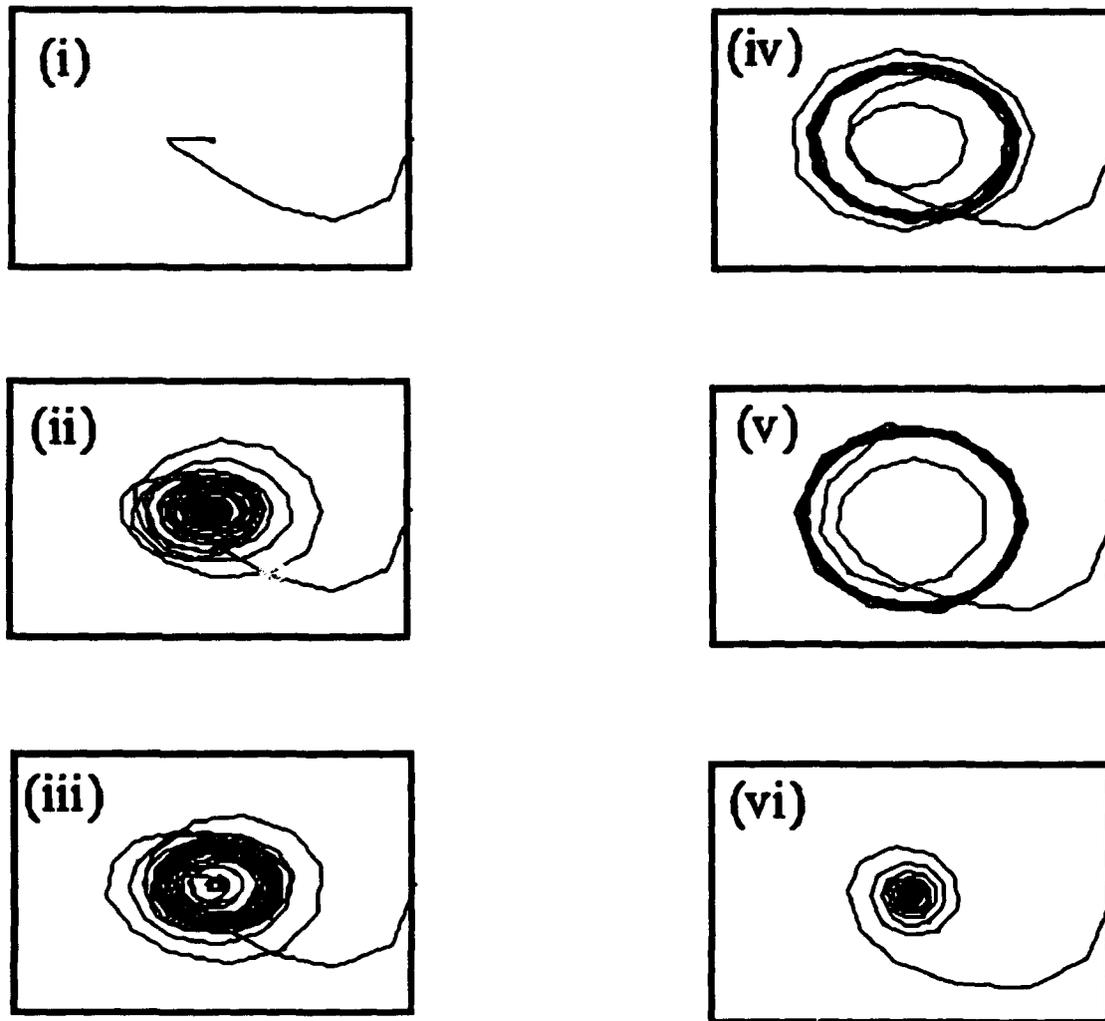


Fig. 1. $p-q$ phase space plots from (17), the dimer form of (1), showing bifurcaton behavior. Each frame extends from $-1$ to $1$ on the horizontal $p$ and the vertical $q$ axis, and the initial condition is of one-site occupation, i.e., $p = 1$, $q = 0$. The nonlinearity $\chi/2V$ is larger than 1: localized states exist. The values of $\alpha/2V$ are (i) 0.002, (ii) 0.00968, (iii) 0.04685, (iv) 0.22676, and (vi) 1.09752. In (i), one sees clear evolution to the self-trapped stationary state ($p < 0$) followed by symmetrization ($p = 0$) as a result of thermal fluctuation. Limit cycles are destroyed in (vi).

linear superposition of $p$ and $r$, the coefficients of the superposition being subject to experimental control.

## 5. Concluding remarks

The derivation of (1) has proceeded through a combination of exact analysis and physical arguments. The semiclassical approximation which yields the DNLSE, the reservoir interaction, also introduced classically, which results in the Fokker–Planck equation, the high damping limit, the contraction analysis from the Fokker–Planck equation, and the generalization from the dimer back to the extended system necessary to pass from (17) to (1), are the primary ingredients of that combination. We have seen the rich behavior that emerges from (1). It encompasses simple linear evolution, reflects the characteristic features of nonlinear dynamics, describes the settling of the system into stationary states as a result of dissipation and finite relaxation, and predicts fascinating bifurcations. What is required in future investigations, in the opinion of the present author, is not only a better understanding of the range of validity and applicability of this and similar transport instruments, but also (and especially) work on the *design of specific experiments* on the basis of the kind of theoretical predictions that have been described in the present article. The depolarization of fluorescence, the scattering of probe particles such as neutrons by light interstitials, the quenching of fluorescence via trapping, and the evolution of transient gratings are particular candidates for the experimental probe. It is hoped that the development of theoretical analysis and the design of experiments will occur in close collaboration in the near future.

## Acknowledgements

## References

[1] A.C. Scott, Phil. Trans. R. Soc. London A 315 (1985) 423.

[2] A.C. Scott, Phys. Rev. A 26 (1982) 578.

[3] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, Physica D 16 (1985) 318.

[4] A.C. Scott and P.L. Christiansen, Phys. Scr. 42 (1990) 257.

[5] V.M. Kenkre and D.K. Campbell, Phys. Rev. B 34 (1986) 4959;
V.M. Kenkre and G.P. Tsironis, Phys. Rev. B 35 (1987) 1473.

[6] V.M. Kenkre, in: Singular Behaviour and Nonlinear Dynamics, Vol. II, eds. St. Pnevmatikos, T. Bountis and Sp. Pnevmatikos (World Scientific, Singapore, 1989).

[7] V.M. Kenkre, in: Disorder and Nonlinearity in Condensed Matter, eds. A. Bishop, D.K. Campbell and S. Pnevmatikos (Springer, Berlin, 1989).

[8] V.M. Kenkre, in: Nonlinear and Coherent Structures, eds. M. Barthes and J. Léon (Springer, Berlin, 1990); in: Polarons and Applications, eds. A. Lakhno and G. Chuev (1992).

[9] D.W. Brown, K. Lindenberg and B. West, Phys. Rev. B 37 (1988) 2946, and references therein.

[10] W.C. Kerr and P. Lomdahl, Phys. Rev. B 35 (1987) 3629.

[11] D.H. Dunlap, V.M. Kenkre and P. Reineker, Phys. Rev. B 47 (1993) 14842.

[12] V.M. Kenkre and M. Kus´, Phys. Rev. B 46 (1992) 13792.

[13] J.D. Andersen and V.M. Kenkre, in: Europ. Sci. Found. Nonlin. Syst. Workshop on Transport Properties in Nonlinear Condensed Matter Physics (Lyngby, Denmark, August 1989); Phys. Stat. Solidi (b) 177 (1993) 397; Phys. Rev. B 47 (1993) 11134.

[14] J.D. Andersen, Bul. Am. Phys. Soc. 34 (1989) 967.

[15] V.M. Kenkre and G.P. Tsironis, Chem. Phys. 128 (1988) 219.

[16] V.M. Kenkre and G.P. Tsironis, Phys. Rev. B 35 (1987) 1473.

[17] H.-L. Wu, Ph.D. Thesis, University of New Mexico, 1988.

[18] V.M. Kenkre, H.-L. Wu and I. Howard, Bul. Am. Phys. Soc. 33 (1988) K22-4; Theory of rotational polarons, Phys. Rev. B, submitted.

[19] H.-L. Wu and V.M. Kenkre, Phys. Rev. B 39 (1989) 2664.

[20] K. Lindenberg, R.P. Hemenger and R.M. Pearlstein, J. Chem. Phys. 56 (1972) 4852.

[21] R.S. Knox, in: Bioenergetics of Photosynthesis, ed. Govindjee (Academic Press, New York, 1975) pp. 183.

[22] G.P. Tsironis, Ph.D. Thesis, University of Rochester, 1987, unpublished.

[23] M. Kus and V.M. Kenkre, Analytical solutions for the nonadiabatic nonlinear dimer in the absence of dissipation, Phys. Rev. B, submitted.

[24] P. Grigolini, H.-L. Wu and V.M. Kenkre, Phys. Rev. B 40 (1989) 7045.

[25] V.M. Kenkre and P. Grigolini, Z. Phys. B 90 (1993) 247.

[26] G.P. Tsironis, V.M. Kenkre and D. Finley, Phys. Rev. A 37 (1988) 4474.

[27] V.M. Kenkre and H.-L. Wu, Phys. Rev. B (1989) 6907; Phys. Lett. A 135 (1989) 120; H.-L. Wu, P. Grigolini and V.M. Kenkre, J. Phys. Cond. Matter 2 (1990) 4417.

PHYSICA D

# Quantum integrable systems

Miki Wadati, Taro Nagao and Kazuhiro Hikami

*Department of Physics, Faculty of Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan*

We report some recent results related to quantum integrable systems. The Baxter formula is generalized into the case of finite temperature. We reformulate the quantum inverse scattering method for 1D quantum systems with long-range interactions. Extensions of the Gaudin model and integrability of the Calogero model are discussed.

## 1. Introduction

In 1965, the soliton as a new concept in nonlinear dynamics was introduced by Zabusky and Kruskal [1]. Exciting findings in the pioneer period were collected in an excellent review article by Scott, Chu and McLaughlin [2]. During the last decade, there continued to be many important developments [3]. One of them is a unification of exactly solvable models in physics. The quantum inverse scattering method places the theory of exactly solvab! dels in a unified framework and provides ul method for studying those models. So models in (1 + 1)-dimensional quantum theory and in 2-dimensional classical statistical mechanics share a common property: to each model we can associate a family of commuting transfer matrices which are generators of an infinite number of conserved quantities. This property may correspond to the Liouville theorem for classical Hamiltonian systems. A sufficient condition for the commutability of the transfer matrices is called the Yang–Baxter relation [4]. The Yang–Baxter relation is a key of new ideas and new concepts in recent mathematical physics such as knot theory based on solvable models, and quantum groups.

This report presents some recent results related to the quantum integrable systems. It is known that there exists a correspondence between 1D quantum system and 2D classical system. The Baxter's formula relates a 2D solvable lattice model to a 1D integrable spin system. In section 2, we extend the Baxter's formula into the case of finite temperature. Combining this extension with the evaluation of finite size corrections, we obtain a systematic method to calculate low temperature expansions of thermodynamic quantities [5–7]. Sections 3 and 4 deal with integrable systems with long-range interactions. Some time ago, Gaudin introduced a class of such spin models using the Bethe ansatz method. In section 3, we give an algebraic formulation of the models. Introducing inhomogeneities into lattice models, we present a general method to construct quantum integrable spin Hamiltonians with long-range pairwise interactions. The last section is devoted to a reformulation of the quantum inverse scattering method for 1D quantum particle systems. In particular, construction of conserved operators for the Calogero–Moser system is explicitly shown. The expression in terms of the Lax operator is new. Further, we find that the Lax pair yields an interesting algebra.

## 2. Finite temperature Baxter's formula

We consider a square lattice with $M$ rows and $N$ columns under the periodic boundary condition $(M + 1 \equiv 1, N + 1 \equiv 1)$, and define a model of classical statistical mechanics on the lattice. We assume that the Boltzmann weights of the lattice model satisfy the Yang–Baxter relation. Let $T_N(u)$ denote the transfer matrix, where $u$ is the spectral parameter. The Yang–Baxter relation assumes the commutability of the transfer matrices:

$$[T_N(u), T_N(v)] = 0 . \tag{2.1}$$

It is known that there is a relationship between the row-to-row transfer matrix $T_N(u)$ and a 1D quantum Hamiltonian $H$:

$$H = -T_N(0)^{-1} \frac{\partial}{\partial u} T_N(u) \bigg|_{u=0} . \tag{2.2}$$

For instance, the Heisenberg $XYZ$ $(XXZ)$ model is related to the eight- (six-) vertex model by a formula (2.2). We call (2.2) Baxter's formula.

The relation (2.1) contains useful information. Expanding the transfer matrix in powers of the spectral parameter, we can easily verify that the transfer matrix is a generator of an infinite number of conserved quantities (operators). The first term in such an expansion is the shift operator:

$$T_N(0)^{b_1 b_2 \cdots b_N}_{a_1 a_2 \cdots a_N} = \delta(b_1, a_N) \cdots \delta(b_N, a_{N-1}) , \tag{2.3}$$

where $a = \{a_j\}$ and $b = \{b_j\}$ denote state variables respectively in lower and upper rows, and $\delta(b_i, a_j)$ is the Kronecker delta. A set of the conserved quantities, $\{I_k\}$, is given by

$$I_k = \frac{\partial^k}{\partial u^k} \log T_N(u) \bigg|_{u=0} . \tag{2.4}$$

Baxter's formula (2.2) corresponds to $I_1$ up to a sign.

The above discussion indicates that $[T_N(u), I_k] = 0$ and that $T_N(u)$ and $\{I_k\}$ have common eigenstates. Thus, Baxter's formula

(2.2) leads to the relationship between the ground-state energy $E_g$ of $H$ and the maximum eigenvalue $\Lambda_{max}(u)$ of $T_N(u)$:

$$E_g = -\frac{\partial}{\partial u} \log \Lambda_{max}(u) \bigg|_{u=0} . \tag{2.5}$$

We have observed that the equivalence through Baxter's formula between a 1D integrable quantum system and a 2D solvable lattice model is useful in particular when we discuss the zero-temperature properties (e.g. ground state energy) of the former. A challenging problem may be how we extend Baxter's formula to study the finite temperature properties.

We look for a new approach to analyze thermodynamic properties of the 1D quantum system. Baxter's formula (2.2) assures the expansion,

$$T_N(u) = T_N(0) [1 - uH + \mathcal{O}(u^2)] . \tag{2.6}$$

Using Trotter's formula, we have

$$\exp(-\beta H) = \lim_{M \to \infty} (T_N(0)^{-1} T_N(\beta/M))^M , \tag{2.7}$$

where $\beta = 1/k_B T$ as usual. Since $T_N(0)$ is the shift operator, we may interpret $T_N(0)^{-1} T_N(u)$ as a *diagonal-to-diagonal* transfer matrix and denote it by $T_N^{DTD}(u)$. Then, the free energy per site, $f$, of the 1D quantum system can be expressed in terms of the partition function $Z_{MN}$ of a 2D classical model of the lattice:

$$f = -\frac{1}{N\beta} \lim_{N \to \infty} \lim_{M \to \infty} \log Z_{MN}(u) , \tag{2.8a}$$

$$Z_{MN}(u) = \mathrm{Tr}(T_N^{DTD}(u))^M , \quad u \equiv \beta/M . \tag{2.8b}$$

The eigenvalues of the transfer matrix are infinitely degenerate as $u \to 0$. Therefore, (2.8) is not practical and remains to be formal. To avoid this situation, we introduce a trick: we look at the lattice from the crossing channel, that is, from a 90° rotated frame.

Because of the crossing symmetry of the model, this manipulation is made by the change in the spectral parameter $u$ into $\lambda - u$, where $\lambda$ is

called the crossing parameter [4,5]. Correspondingly, we introduce a notation $T^X_M(u) \equiv T^{DTD}_M(\lambda - u)$. Now, (2.8) is replaced by

$$f = -\frac{1}{\beta} \lim_{M \to \infty} \log \Lambda_M(u) , \qquad (2.9a)$$

$$Z_{MN}(u) = \mathrm{Tr}(T^X_M(u))^N , \quad u = \beta/M , \qquad (2.9b)$$

where $\Lambda_M(u)$ is the largest eigenvalue of the crossing transfer matrix $T^X_M(u)$. Equation (2.9) is a finite temperature extension of Baxter's formula (2.2), and has been named finite temperature Baxter's formula. Remark that the spectral parameter $u$, a fundamental quantity in soliton theory, now plays a role of intertwiner between a finite size system and a finite temperature system.

A caution is necessary for making use of formula (2.9). Since the temperature dependence enters through $u = \beta/M$, one should be careful in taking the limit $M \to \infty$. Otherwise, any information of finite temperature will be lost. Instead, a finite size calculation or a $1/M$-expansion gives a systematic method to obtain low temperature expansions of thermodynamic quantities. Figure 1 summarizes the significance of the limits $T \to 0$ and $M \to \infty$ for the 1D quantum system and the 2D classical system.

The theory of finite size corrections has attracted much attention of theoretical physicists since it gives central charge $c$ and scaling dimension $x$ which are essential quantities in the conformal field theory. To evaluate finite size corrections, we adopt a method by de Vega, Woynarovich and Eckle [8]. The calculation is
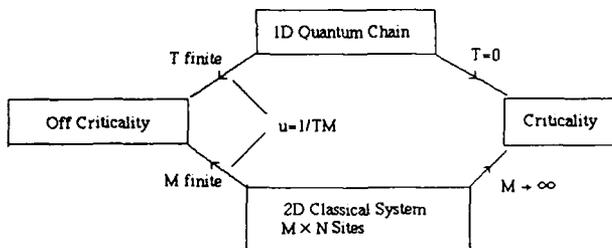


Fig. 1. Relations between finite size system and finite temperature system.

involved but straightforward. We apply the Euler–Mclaurin expansion formula and the solution technique of the Wiener–Hopf integral equation to the Bethe ansatz equation.

The new approach to finite temperature quantum systems presented above has been successfully applied [5–7]. It should be stressed that this approach is valid for any quantum chain as long as it is solvable by the Bethe ansatz method. We only mention here the results for the XXZ spin chain in a gapless phase:

$$H = J_T \sum_{l=1}^{N} (\sigma^x_l \sigma^x_{l+1} + \sigma^y_l \sigma^y_{l+1} + \Delta \cdot \sigma^z_l \sigma^z_{l+1}) , \qquad (2.10)$$

$$\Delta = \cos \lambda . \qquad (2.11)$$

The free energy $f$ with the first thermal correction is

$$f = f_\infty - \frac{\lambda(k_B T)^2}{12 \sin \lambda \cdot J_T} + \cdots , \qquad (2.12)$$

where $f_\infty$ is the ground state energy. The result (2.12) is the same as the one obtained by the Bethe ansatz method with the string hypothesis.

Calculation of the higher terms in $T$ can be done by making use of the perturbation theory of the Wiener–Hopf integral equation. In the limit $\lambda \to 0$ (the XXX model limit), the correction terms have rather simple expressions. The free energy $f$ and the correlation length $\xi$ are respectively given by

$$f = f_\infty - \frac{(k_B T)^2}{12 J_T} + \frac{\pi}{2 J_T} \frac{(k_B T)^2}{[\log(J_T/k_B T)]^3} + \cdots , \qquad (2.13)$$

$$\xi = \frac{a_1}{T} + \frac{a_2}{T \log(1/T)} + \cdots , \qquad (2.14)$$

where $a_1$ and $a_2$ are constants. Two remarks should be made. First, the appearances of logarithmic terms in (2.13) and (2.14) are noteworthy. This is the first analytical derivation of such terms and has been confirmed by numerical

calculations based on a different formalism. Second, we find that the calculation of the correlation length is quite straightforward, by making use of the result for the corresponding 2D classical models. This is remarkable since in string hypothesis approach we can obtain only the free energy and cannot get any information on the correlation length.

## 3. The Gaudin model and generalizations

Depending on the interaction ranges, completely integrable systems may be classified into two groups. One is a system with short-range interactions including the $\delta$-function gas, the Heisenberg $XYZ$ model and the Toda lattice. The other is a system with long-range interactions. Contrary to the short-range interaction models, there has not been a systematic work on the long-range interaction models. So far we know two integrable systems, the Calogero–Moser model and the Gaudin model [9]. Recently there has been a renewed interest in the long-range interactions models. It is believed that the models have the Jastrow-type wave functions which play a central role in the theory of the quantum Hall effects and the high-$T_c$ superconductors. In this section, we present a general method [10,11] to construct quantum integrable spin Hamiltonians with long-range pairwise interactions which extend the Gaudin model.

Let the matrix $R$ act in a tensor product $V \otimes V$ of the linear space $V$. The Yang–Baxter relation reads

$$R_{12}(u - v)\, R_{13}(u)\, R_{23}(v)$$
$$= R_{23}(v)\, R_{13}(u)\, R_{12}(u - v) \,. \tag{3.1}$$

Here $R_{ij}$ denotes the matrix on $V \otimes V \otimes V$, acting as $R$ on the $i$th and the $j$th spaces and as identity on the other space. A solution of (3.1) is referred to as $R$-matrix. The $R$-matrix depends on the spectral parameter $u$ (or $v$) and other parame-

ters, one of which we denote by $\eta$. It is sometime convenient to use the $\check{R}$-matrix defined by

$$\check{R}(u; \eta) = R(u; \eta)\, P \,, \tag{3.2}$$

where $P$ is a permutation operator, $P(x \otimes y) = y \otimes x$. For $R(u; \eta)$ and $\check{R}(u; \eta)$, we assume the following two conditions.

(I) regular condition,

$$R(u = 0; \eta) = P \,, \quad \check{R}(u = 0; \eta) = I \,, \tag{3.3}$$

(II) quasi-classical condition,

$$R(u; \eta = 0) = I \,, \quad \check{R}(u; \eta = 0) = P \,. \tag{3.4}$$

(3.3) is also called the standard initial condition.

In terms of $R$-matrices, we define the transition matrix $\tau_N(u)$ and the transfer matrix $T_N(u)$ by

$$\tau_N(u) = \tau_N(u; \{x_j\}; \eta)$$
$$= R_{0N}(u - x_N) \cdots R_{01}(u - x_1) \,, \tag{3.5}$$

$$T_N(u) = T_N(u; \{x_j\}; \eta)$$
$$= \mathrm{Tr}\, \tau_N(u) \,. \tag{3.6}$$

Remark the shift $x_j$ of the spectral parameter in $R_{0j}$. We regard $R_{jk}$ as a matrix acting on $V_j \otimes V_k$, where $V_0$ is the "auxiliary" space and $V_k$ denotes the $k$th site on a one-dimensional lattice. In (3.6), Tr means trace in the space $V_0$. Then, the system (3.5) is identical to an inhomogeneous 1D lattice with $N$ sites (fig. 2).

Because of the additive property of the spectral parameters, the Yang–Baxter relation remains valid for the inhomogeneous lattice. Then, the transfer matrices (3.6) with different spectral parameters commute,
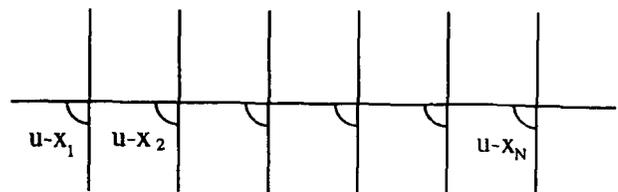


Fig. 2. Inhomogeneous lattice.

$[T_N(u), T_N(v)]$

$\equiv T_N(u)\,T_N(v) - T_N(v)\,T_N(u) = 0\,.$ (3.7)

This way of introducing the inhomogeneity is known as the $Z$-invariance.

Let $|\lambda\rangle$ be an eigenstate of $T_N(u)$ which is characterized by a set of "quasi-momenta" $\lambda_1$, $\lambda_2, \ldots, \lambda_M$. A set of mutually commuting operators and the eigenvalue problem are respectively expressed as

$Z_k = T_N(u = x_k; \{x_j\}; \eta)\,,\quad k = 1, 2, \ldots, N\,,$ (3.8)

$$Z_k|\lambda\rangle = \prod_{l=1}^{M} \frac{1}{c(\lambda_l - x_k)}\,|\lambda\rangle\,.$$ (3.9)

Of course, explicit forms of $c(\lambda)$ and $\lambda_j$ are model-dependent.

We differentiate both sides of (3.9) with respect to $\eta$. Then, setting $\eta = 0$ and using the quasi-classical condition (3.4), we get

$$H_k|\lambda\rangle\big|_{\eta=0} = \left(\frac{\partial}{\partial\eta}\prod_{l=1}^{M}\frac{1}{c(\lambda_l - x_k)}\right)|\lambda\rangle\bigg|_{\eta=0}\,,$$ (3.10)

where

$$H_k \equiv \frac{\partial}{\partial\eta}\,Z_k\bigg|_{\eta=0}\,.$$ (3.11)

We have found that $H_k$ in (3.11) can be expressed concisely in terms of the $R$-matrices. We denote by $T_{N\{a\}}^{\{b\}}$ the transfer matrix from lower states $\{a_j\}$ to upper states $\{b_j\}$. From the definitions (3.5), (3.6) and (3.8) and the regular condition (3.3), we have

$Z_{k\{a\}}^{\{b\}} = T_{N\{a\}}^{\{b\}}(u = x_k; \{x_j\}; \eta)$

$= \mathrm{Tr}\, R_{0N}(x_k - x_N) \cdots P_{0k} \cdots R_{01}(x_k - x_1)$

$= R_{kk-1}(x_k - x_{k-1}) \cdots R_{k1}(x_k - x_1)$

$\quad \times R_{kN}(x_k - x_N) \cdots R_{kk+1}(x_k - x_{k+1})\,.$

(3.12)

Differentiating the last expression and using the quasi-classical condition (3.4), we obtain

$$H_k = \sum_{j\neq k}\frac{\partial}{\partial\eta}\,R_{kj}(x_k - x_j; \eta)\bigg|_{\eta=0}\,,$$

$k = 1, 2, \ldots, N\,.$ (3.13)

We can regard $\{H_k\}$ as commuting Hamiltonians. In fact, substituting the expression $Z_k = I + \eta H_k + \mathcal{O}(\eta^2)$ into (3.7), we see that

$[H_k, H_l] = 0\,,\quad k, l = 1, 2, \ldots, N\,.$ (3.14)

This indicates that the Hamiltonian system

$$H = \sum_{k=1}^{N} a_k H_k\,,\quad a_k: \text{constants}\,,$$ (3.15)

is completely integrable.

It is interesting to compare formula (3.13) with Baxter's formula (2.2). While Baxter's formula gives a spin Hamiltonian with nearest-neighbor interactions, the formula (3.13) yields a spin Hamiltonian with long-range interactions.

The above discussion is general. We have used only the Yang–Baxter relation and the properties of the $R$-matrix such as the regular condition, the quasi-chemical condition and the inhomogeneity. Given an $R$-matrix, we have commuting Hamiltonians $\{H_k\}$ by formula (3.13). For instance, from the $R$-matrix for the spin-$\frac{1}{2}$ $XYZ$ model, we have

$$H_l = \tfrac{1}{2}\sum_{j\neq l}\frac{1}{\mathrm{sn}(x_l - x_j)}$$

$\times \{[1 + k\,\mathrm{sn}^2(x_l - x_j)]\sigma_l^x\sigma_j^x$

$+ [1 - k\,\mathrm{sn}^2(x_l - x_j)]\sigma_l^y\sigma_j^y$

$+ \mathrm{cn}(x_l - x_j)\,\mathrm{dn}(x_l - x_j)\,(\sigma_l^z\sigma_j^z - 1)\}\,,$ (3.16)

where $k$ is the modulus of Jacobi's elliptic functions. Applications to other models including the spin-1 model have been discussed [10,11].

## 4. Quantum integrable particle systems

The quantum inverse scattering method for $N$-particles on a line may be introduced as follows. Let $L$ and $M$ be $N \times N$ matrix

operators. We choose $L$, $M$ (Lax pair) such that the Lax equation

$$\dot{L} = i[H, L] = i[L, M] , \qquad (4.1)$$

is equivalent to the equation of motion generated by a Hamiltonian $H$ under consideration. We associate an evolution of "eigenstate" as

$$i\dot{U} = [U, H] = MU . \qquad (4.2)$$

Then, from (4.1) and (4.2), we have

$$[H, U^{-1}LU] = 0 . \qquad (4.3)$$

Equation (4.3) is a quantum version of the unitary equivalence by P. Lax. For classical systems, the condition that $U^{-1}LU$ does not depend on time leads to the existence of $N$ conserved quantities $\{I_n\}$,

$$I_n \equiv \frac{1}{n} \mathrm{Tr}(U^{-1}L^nU) = \frac{1}{n} \mathrm{Tr}\, L^n , \qquad (4.4)$$

where Tr means the trace of matrix. For quantum systems, since $L$ and $U$ are operators, the last equality in (4.4) is not guaranteed.

To be specific, we restrict our discussion to the Calogero–Moser system whose Hamiltonian is given by

$$H = \tfrac{1}{2} \sum_{j=1}^{N} p_j^2 + \tfrac{1}{2}g \sum_{j \neq k} \gamma(x_j - x_k) , \qquad (4.5)$$

$$p_j = -i\frac{\partial}{\partial x_j} , \quad \gamma(x) = \frac{1}{x^2} . \qquad (4.6)$$

The Lax pair is found to be

$$L_{jk} = p_j\delta_{jk} + ia(1 - \delta_{jk})\,\alpha(x_j - x_k) , \qquad (4.7)$$

$$M_{jk} = a(1 - \delta_{jk})\,\beta(x_j - x_k) + a\delta_{jk} \sum_{l \neq j} \gamma(x_j - x_l) , \qquad (4.8)$$

where $a$ is a constant related to the coupling constant $g$ by $g = a^2 - a$, and

$$\alpha(x) = 1/x , \quad \beta(x) = -1/x^2 . \qquad (4.9)$$

Calogero, Ragnisco and Marchioro [12] presented a set of conserved operators (rigorously speaking, their potential is $1/\sinh^2 x$ instead of

$1/x^2$, but the discussion is the same). They proved that with $g = a^2$,

$$[H, \det \Lambda] = 0 , \qquad (4.10)$$

where the matrix $\Lambda$ is an $N \times N$ matrix defined by $\Lambda_{jk} = L_{jk} - \lambda\delta_{jk}$. The expansion of $\det \Lambda$ in powers of $\lambda$ gives a set of conserved operators $\{J_n\}$:

$$\det[\Lambda_{jk}] = \det[L_{jk} - \lambda\delta_{jk}]$$

$$= (-\lambda)^N + \sum_{n=1}^{N} (-\lambda)^{N-n} J_n . \qquad (4.11)$$

Those conserved operators $\{J_n\}$ have the same functional forms as the classical ones. Only the difference is that $p_j$ is the operator in the quantum case. To carry out the proof of (4.10), they modified (4.1) into

$$[H, L_{jk}] = \tfrac{1}{2} \sum_{l} (L_{jl}M_{lk} + M_{lk}L_{jl}$$

$$- M_{jl}L_{lk} - L_{lk}M_{jl}) . \qquad (4.12)$$

However, the symmetrized commutator in the r.h.s. of (4.12) needs a justification from a viewpoint of the quantum inverse scattering method. We therefore believe that their proof is not satisfactory, although the conserved operators $\{J_n\}$ are widely accepted.

An alternative method to find conserved operators may be the following. From (4.1), we have

$$[H, (L^n)_{jk}] = \sum_{l} [(L^n)_{jl}M_{lk} - M_{jl}(L^n)_{lk}] . \qquad (4.13)$$

Explicit form of $M_{jk}$ in (4.8) gives

$$\sum_{j} M_{jk} = 0 , \quad \sum_{k} M_{jk} = 0 . \qquad (4.14)$$

Then, we find from (4.13) and (4.14) that

$$\left[H, \sum_{j,k} (L^n)_{jk}\right] = 0 , \qquad (4.15)$$

that is, conserved operators $\{I_n\}$ are given by

$$I_n = \frac{1}{n} \sum_{j,k} (L^n)_{jk} , \quad n = 1, 2, \ldots, N . \qquad (4.16)$$

Expression (4.16) is new. The first three of the conserved operators are

$$I_1 = \sum_{j,k} L_{jk} = \sum_j p_j ,$$

$$I_2 = \frac{1}{2} \sum_{j,k} (L^2)_{jk} = \frac{1}{2} \sum_j p_j^2 + \frac{1}{2} g \sum_{j \neq k} \frac{1}{(x_j - x_k)^2} ,$$

$$I_3 = \frac{1}{3} \sum_{j,k} (L^3)_{jk}$$

$$= \frac{1}{3} \sum_j p_j^3 + \frac{1}{3} g \sum_{j \neq k} \left( p_j \frac{1}{(x_j - x_k)^2} \right.$$

$$\left. + \frac{1}{x_j - x_k} p_k \frac{1}{x_j - x_k} + \frac{1}{(x_j - x_k)^2} p_k \right) . \qquad (4.17)$$

As we expected, $g$ appears in $\{I_n\}$. While $I_1$ and $I_2$ are the total momentum operator and the Hamiltonian, $I_3$ is a non-trivial conserved operator. In general, $I_n = 1/n \sum p_j^n + \cdots$ is a polynomial in $p_j$ and $1/(x_j - x_k)$.

We have found that the Lax pair gives an interesting algebra. Let us introduce a set of operators,

$$h_j^\dagger = \sum_k L_{kj} = p_j + ia \sum_{k \neq j} \frac{1}{x_k - x_j} , \qquad (4.18)$$

$$h_j = \sum_k L_{jk} = p_j + ia \sum_{k \neq j} \frac{1}{x_j - x_k} . \qquad (4.19)$$

The operators $h_j^\dagger$ and $h_j$ are Hermitian conjugated to each other. They satisfy the commutation relations

$$[h_l, h_m] = 0 , \quad [h_l^\dagger, h_m^\dagger] = 0 ,$$

$$[h_l, h_m^\dagger] = 2M_{lm} , \quad l, m = 1, 2, \ldots, N . \qquad (4.20)$$

A similarity between (3.14) and (4.20) is intriguing. In terms of these operators, the total momentum operator $I_1$ and the Hamiltonian $I_2$ are simply expressed as

$$I_1 = \sum_{k=1}^N h_k^\dagger = \sum_{k=1}^N h_k , \quad I_2 = \frac{1}{2} \sum_{l=1}^N h_l^\dagger h_l . \qquad (4.21)$$

Further detailed analysis of long-range integrable Hamiltonian systems, in particular, the Calogero–Moser system with spins will be discussed elsewhere [13].

## References

[1] N.J. Zabusky and M.D. Kruskal, Phys. Rev. Lett. 15 (1965) 240.

[2] A.C. Scott, F.Y.F. Chu and D.W. McLaughlin, Proc. IEEE 61 (1973) 1443.

[3] A.S. Fokas and V.E. Zakharov, eds., Important Developments in Soliton Theory (Springer, Heidelberg, 1993).

[4] M. Wadati, T. Deguchi and Y. Akutsu, Phys. Rep. 180 (1989) 247.

[5] M. Wadati and Y. Akutsu, Prog. Theor. Phys. Suppl. 94 (1988) 1.

[6] J. Suzuki, T. Nagao and M. Wadati, Int. J. Mod. Phys. B 6 (1992) 1119.

[7] J. Suzuki, Y. Akutsu and M. Wadati, J. Phys. Soc. Jpn. 59 (1990) 2667.

[8] H.J. de Vega and F. Woynarovich, Nucl. Phys. B 251 (1985) 439;
F. Woynarovich and H.P. Eckle, J. Phys. A ˋ (1987) L97.

[9] M. Gaudin, La fonction d'one de Bethe (Masson, Paris, 1983; J. Phys. (Paris) 37 (1976) 1087.

[10] K. Hikami, P.P. Kulish and M. Wadati, J. Phys. Soc. Jpn. 61 (1992) 3071.

[11] K. Hikami, P.P. Kulish and M. Wadati, Chaos, Solitons Fractals 2 (1992) 543.

[12] P. Calogero, O. Ragnisco and C. Marchioro, Lett. Nuovo Cimento 13 (1975) 383.

[13] K. Hikami and M. Wadati, J. Phys. Soc. Jpn. 62 (1993) 469; Phys. Lett. A 173 (1993) 263.

**PHYSICA** $\boxed{D}$

# On localization in the discrete nonlinear Schrödinger equation

O. Bang[a], J.J. Rasmussen[b] and P.L. Christiansen[a]

[a]*Laboratory of Applied Mathematical Physics, The Technical University of Denmark, DK-2800 Lyngby, Denmark*
[b]*Optics and Fluid Dynamics Department, Risø National Laboratory, DK-4000 Roskilde, Denmark*

For some values of the grid resolution, depending on the nonlinearity, the discrete nonlinear Schrödinger equation with arbitrary power nonlinearity can be approximated by the corresponding continuum version of the equation. When the discretization becomes too coarse, the discrete equation exhibits localization in regimes where blow-up cannot occur in the continuum system. This phenomenon is investigated numerically, and the grid resolution at which the transition occurs is determined.

## 1. Introduction

The one-dimensional discrete nonlinear Schrödinger equation with arbitrary power nonlinearity, $\sigma$,

$$i\dot{U}_j + h^{-2}(U_{j+1} - 2U_j + U_{j-1}) + |U_j|^{2\sigma}U_j = 0 , \tag{1}$$

describes a system of $f$ coupled anharmonic oscillators, $j = -\frac{1}{2}(f-1), \ldots, \frac{1}{2}(f-1)$ ($f$ odd). The dispersive coupling has the strength $h^{-2}$, where $h$ is the distance between adjacent sites. Eq. (1) is a discretization of the nonlinear Schrödinger equation with arbitrary power nonlinearity

$$i\psi_t + \psi_{xx} + |\psi|^{2\sigma}\psi = 0 . \tag{2}$$

For high grid resolution (high values of $f$) the solution to eq. (1) is expected to resemble the solution to eq. (2). It is the aim of this paper to study to what extent and in which areas of parameter space $(f, \sigma)$ this is true. In particular we concentrate on the phenomenon of localization or self-trapping [1], that is, the dynamically stable state in which most of the energy of the system gets concentrated at a single site. The continuous counterpart of localization is blow-up [2], which designates the situation where the maximum of $|\psi|$ tends to infinity in finite time. Blow-up has been studied extensively in the last decades, and has now been observed both experimentally, analytically and numerically in many physical systems. Most famous is presumably the Langmuir wave-collapse in plasma physics [3] and self-focussing of laser beams in nonlinear optics [4]. Furthermore it has been used to describe energy transfer in molecular thin films [5].

## 2. The discrete and the continuum model

The discrete model, eq. (1), has only the two conserved quantities the number, $N$, and the Hamiltonian, $H$,

$$N = h \sum_{j=-j'}^{+j'} |U_j|^2 ,$$

$$H = -h \sum_{j=-j'}^{+j'} h^{-2}U_j^*(U_{j+1} - 2U_j + U_{j-1})$$
$$+ (1 + \sigma)^{-1}|U_j|^{2\sigma+2} , \tag{3}$$

where $j' = \frac{1}{2}(f + 1)$. Thus, eq. (1) is not integra-

ble when the system has more than two degrees of freedom. As a measure of localization (or self-trapping), we introduce the mean width of a pulse, $W$, defined as $W = (I/N)^{1/2}$, where the virial, $I$, is given by [2]

$$I = h^3 \sum_{j=-j'}^{j'} j^2 |U_j|^2 \quad \text{corresponding to}$$

$$I = \int_{-\infty}^{+\infty} x^2 |\psi|^2 \, dx .\tag{4}$$

Here the pulse center is assumed to be at site $j = 0$, corresponding to $x = 0$. Thus for a complete localization of the energy at the center site, $W$ becomes zero. The virial theory [2] is a classical method to determine whether a given initial waveform will collapse in a finite time ($I \to 0$ for $t \to t_0$, $t_0 < \infty$), or not.

Corresponding to $N$ and $H$ given by eq. (3), the continuum eq. (2) has the conserved quantities

$$N = \int_{-\infty}^{+\infty} |\psi|^2 \, dx ,$$

$$H = \int_{-\infty}^{+\infty} \left( |\psi_x|^2 - (1 + \sigma)^{-1} |\psi|^{2\sigma+2} \right) dx .\tag{5}$$

As the ground state solitary solution to eq. (2), with zero velocity and centered at $x = 0$, we find

$$\psi_s(x, t; \omega) = \omega^{1/\sigma}(1 + \sigma)^{1/2\sigma} \operatorname{sech}^{1/\sigma}(\omega\sigma x) \, e^{i\omega^2 t} ,\tag{6}$$

where $\omega$ is a real parameter. From general studies it follows that the solution, given by eq. (6), is unstable and blows up at the center if $\sigma \geq 2$, and is unconditionally stable for $\sigma < 2$ [2]. The different regimes of $\sigma$ are denoted the subcritical ($\sigma < 2$), critical ($\sigma = 2$) and supercritical regime ($\sigma > 2$), respectively. In the critical case an arbitrary initial condition will blow-up if the requirement $N \geq N_s$ is fulfilled. Here $N_s \equiv N(\psi_s)$, where $N(\psi)$ is given by eq. (5).

## 3. Numerical results

Eq. (1) is integrated numerically with $\psi_s(x, t; \omega = 1)$, given by eq. (6), as initial condition

$$U_j(0) = \psi_s(jh, 0; \omega = 1) .\tag{7}$$

A fourth order Runge–Kutta scheme, with time step $dt = 10^{-5}$, is used for the integration throughout the paper. The value $10^{-5}$ of $dt$ assures the conservation of $N$ and $H$ within $10^{-9}$ of their initial values for all runs. The most important results found, have been checked by repeating them with $dt = 5 \times 10^{-6}$ and $dt = 5 \times 10^{-5}$, and by applying another integration scheme, the Burlirsch–Stoer method with Richardson extrapolation [6], to assure that they are independent of the time integration. $N, H, I$ etc. are checked at time intervals $\Delta t = 10^{-2}$ and the length of the $x$-interval, $L = (f - 1)h$, has the value 25.6 for results presented.

Fig. 1 shows the time evolution of the amplitude $A_j = |U_j|$, for different values of $\sigma$ and $f$. For $\sigma = 1.90$ and $f = 185$ (fig. 1a) the pulse is stable. This is also predicted from the continuum
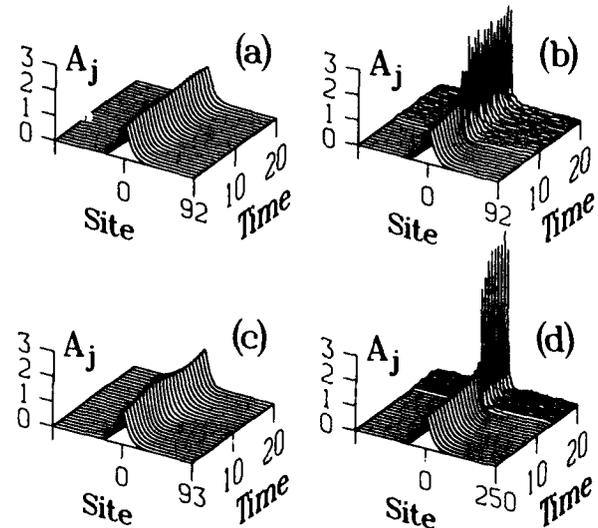


Fig. 1. Amplitude of the solution to eq. (1) and (7) with $L = 25.6$. (a) $\sigma = 1.90$, $f = 185$, (b) $\sigma = 1.95$, $f = 1.85$, (c) $\sigma = 1.95$, $f = 191$, (d) $\sigma = 2.00$, $f = 501$.

system, eq. (2), since $\sigma = 1.9$ is a subcritical value. However, a small oscillation of the amplitude and width of the pulse is noticed. Fig. 1b shows that most of the energy becomes localized at the center site when $\sigma$ is increased to 1.95. This contrasts the stability of the ground state solution for the continuum model, in the subcritical regime. The observed localization is not complete ($W \neq 0$) since not all of the energy is concentrated at the center site. Actually, for the initial conditions used in our numerical simulations, the state of complete localization cannot be reached. There will always be a certain amount of energy at the neighbouring sites. However, a final state as in fig. 1b will be characterized as localized in the following. In fig. 1c, $f$ is raised to 191 while $\sigma$ is still 1.95. It is seen that this decrease of $h$ prevents localization of the energy. Instead periodic oscillations occur. These are of the same type as in fig. 1c. We have checked that they continue, at least until $t = 100$ (further increase of $f$ leads to a decrease of the oscillations). In fig. 1d, $\sigma$ is raised to the critical value 2, where blow-up occurs in the continuum system. Correspondingly localization is observed, despite the fine resolution $f = 501$. We note that the localization for $\sigma = 1.95$ (fig. 1b) is due to the discreteness of the system, when the oscillation pulse is most narrow it cannot be resolved. In contrast, the localization for $\sigma = 2$ (fig. 1d) is a reflection of the blow-up phenomenon in the continuum system. The discreteness and the conservation of $N$ prevents the amplitude at a single site to reach infinity. While the localization for $\sigma = 1.95$ can be removed by increasing $f$, this is not the case with the localization for $\sigma = 2$. Here the localization is only delayed when $f$ is increased. The variation of the oscillations of pulse width and amplitude for different values of $\sigma$ are shown in more detail in fig. 2 in the subcritical regime, $\sigma < 2$. Here the time evolution of the mean width, $W$, and the center amplitude, $A_0$, are depicted. Clearly both the period and the amplitude of the oscillations are increasing with $\sigma$. Furthermore, the oscilla-
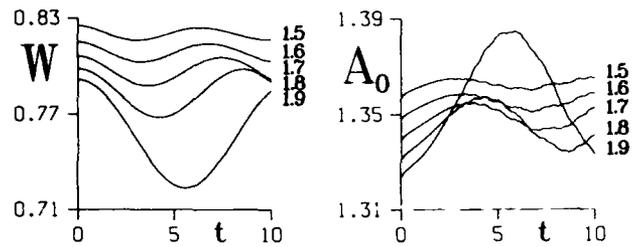


Fig. 2. Width, $W$, and center amplitude, $A_0$, as function of time for $f = 257$ and $L = 25.6$. $\sigma = 1.5$, 1.6, 1.7, 1.8, 1.9.

tions of $W$ and $A_0$ are 90° out of phase, due to conservation of $N$. Fig. 3 shows the variation of the oscillations for different values of $f$, the number of sites in the interval $L = 25.6$. The nonlinearity, $\sigma$, has the subcritical value 1.95. For $f \leq 1.87$ the time evolution has two distinguishable states: (i) First there is a monotonical decrease (increase) of $W$ ($A_0$), resembling the initial phase of a blow-up in the continuum system. The first minimum (maximum) of $W$ ($A_0$) is delayed as $f$ is increased. We note again that $W$ never reaches zero, but has a minimum value approximately equal to the distance between sites, $h$. (ii) Then follows sustained rapid oscillations. The mean value of this final oscillation increases with $f$, while the period decreases with $f$. For these $f$-values the time evolution is of the same type as shown in fig. 1b.

At $f = 191$ an abrupt transition has occurred. There is no longer localization. Instead the pulse performs a slow periodic oscillation, of the same type as seen in fig. 1a and fig. 1c. We shall denote this transition value of $f$, by $f_t$. Increasing $f$ to 501, the oscillation are still present, but much weaker.
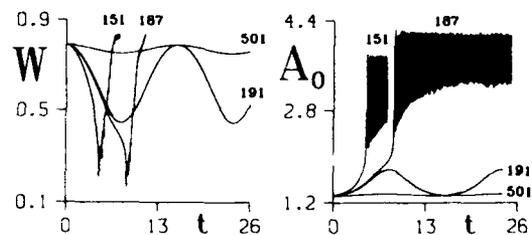


Fig. 3. $A_0$ as function of time for $\sigma = 1.95$ and $L = 25.6$. $f = 171$, 187, 191, 201.
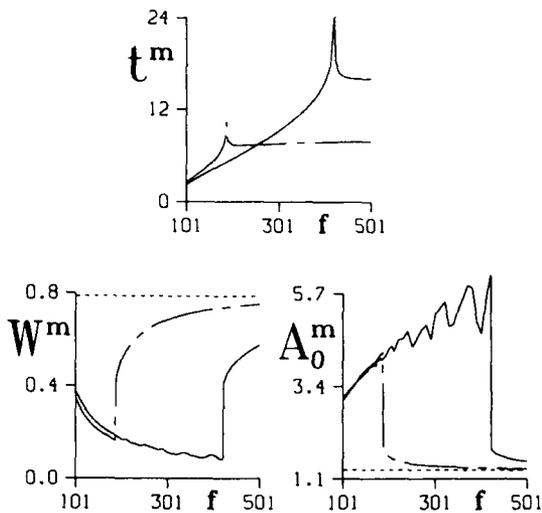
Fig. 4. $t^m$, $W^m$ and $A_0^m$ (defined in text) as function of $f$ with $L = 25.6$. (- · -) $\sigma = 1.95$ and (——) $\sigma = 1.99$. Initial values of $W^m$ and $A_0^m$ are shown as dashed lines.



Fig. 5. Amplitude of the solution to eq. (1) with $f = 107$, $\sigma = 2$ and $L = 25.6$. The amplitude of the initial condition, given by eq. (7), is reduced by a factor (a) 1.000, (b) 0.994, (c) 0.993.

These results are summarized in fig. 4. Here $t^m$ is defined as the value of $t$ at which $W$ reaches its first minimum. Correspondingly, $W^m \equiv W(t^m)$ and $A_0^m \equiv A_0(t^m)$. Results for two subcritical values of $\sigma$, $\sigma = 1.95$ (- · -) and $\sigma = 1.99$ (——), are shown.

For $f < f_t$, $t^m$ increases with $f$. Here $t^m$ is the time of localization. For $f \geq f_t$, $t^m$ decreases rapidly towards an asymptotic value. Here $2t^m$ is the period of the slow oscillations.

For $f < f_t$, $W^m$ decreases approximately as $1/f$, again implying that the width is bounded from below by the distance between sites, $h$. For $f \geq t_t$, $W^m$ increases asymptotically towards $W(0)$, indicated by dashed line. A corresponding behavior is observed for $A_0^m$. (The randomness for $f < f_t$ is due to limited time resolution in the plotting $\Delta t = 10^{-2}$).

In fig. 5 $\sigma = 2$ (the critical value) and $f = 107$. Fig. 5a shows the time evolution of the amplitude, with the initial condition given by eq. (7). Localization is seen to occur. In fig. 5b the initial amplitude is reduced by a factor 0.994. Localization still occurs, but delayed with respect to the localization in fig. 5a. In fig. 5c the initial

amplitude is reduced by a factor 0.993. Here the pulse no longer localizes, but disperses slowly.

In the continuum model the criterion for blow-up of an arbitrary initial condition is $N \geq N_s = \frac{1}{2}\pi\sqrt{3}$ in the critical case. Fig. 5 shows that in the discrete system, for $f = 107$, the limiting value of $N$ is reduced from $N_s$ to $0.988 \cdot N_s$. Decreasing $f$ further, this limiting value will decrease correspondingly.

## 4. Conclusion

In this paper we have studied numerically the relation between localization in the discrete nonlinear Schrödinger equation with arbitrary power nonlinearity, and blow-up in the corresponding continuum equation. It is found that localization may occur, even for subcritical values of the degree of nonlinearity. A sharp transition between localization and smooth oscillatory behavior occurs. A reduction of the initial amplitude is shown to be equivalent to a reduction of the nonlinearity.

## Acknowledgements

## References

[1] L.D. Landau, Phys. Z. Sowjetunion 3 (1933) 664.
[2] J.J. Rasmussen and K. Rypdal, Physica Scr. 33 (1986) 481.
[3] V.E. Zakharov, Sov. JETP 35 (1972) 908;
M.V. Goldman, Rev. Mod. Phys. 56 (1984) 709.
[4] V.N. Lugovoi, A.P. Sukhorukov and R.V. Khokhlov, Usp. Fiz. Nauk. 93 (1967) 19.
[5] P.L. Christiansen, S. Pagano and G. Vitello, Phys. Lett. A 154 (1991) 381.
[6] W.H. Pres, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes (FORTRAN Version) (Cambridge Univ. Press, 1989) p. 563.
[7] M. Desaix, D. Anderson and M. Lisak, J. Opt. Soc. Am. B 8 (1991) 2082.
[8] O. Bang, J.J. Rasmussen and P.L. Christiansen, Subcritical localization in the discrete nonlinear Schrödinger equation with arbitrary power nonlinearity, submitted.

**PHYSICA** Ⓓ

# Quantizing a self-trapping transition

## L.J. Bernstein

*Department of Chemistry 0340 and The Institute for Nonlinear Science, University of California,*
*San Diego, La Jolla, CA 92093-0340, USA*

Stationary state energy levels of the quantum discrete self-trapping equation (DST) dimer are computed numerically and estimated analytically. Quantum-mechanical manifestations of a self-trapping transition that occurs in the corresponding classical DST dimer are found in both quantum energy levels and quantum wavefunctions. The work is motivated by applications of the DST equation as a model for local vibrational modes in polyatomic molecules.

## 1. Introduction

Clarifying the influence of quantum-mechanical effects on the behavior of classically non-linear dynamical systems is an important goal for future nonlinear science research. A mathematical model particularly well-suited to this research area is the discrete self-trapping equation (DST), introduced in 1985 by Eilbeck, Lomdahl and Scott [1]. The DST is a nonlinear and generally non-integrable classical dynamical system of $f$ coupled, anharmonic oscillators and is uncommonly easy to quantize. Already the DST has been useful in studies of quantum manifestations of nonlinear classical phenomena such as soliton-like energy localization [2,3] and deterministic chaos [4,5]. The DST is also a useful model for molecular bond vibrations [6–8].

In the special case of just two degrees of freedom, that DST equation takes the form

$$i\frac{da_1}{dt} = \omega_0 a_1 - Ja_2 - \gamma|a_1|^2 a_1 ,$$

$$i\frac{da_2}{dt} = \omega_0 a_2 - Ja_1 - \gamma|a_2|^2 a_2 , \tag{1}$$

where $a_j$ is the complex amplitude of vibrations of the $j$th bond, $\omega_0$ is the frequency of uncoupled, low amplitude oscillations and the real parameters $J$ and $\gamma$ give the magnitudes of linear intersite coupling and on-site anharmonicity, respectively. The effective nonlinearity is

$$\zeta = \frac{\gamma N}{J} , \tag{2}$$

where $N = |a_1|^2 + |a_2|^2$. Both the norm $N$ and the Hamiltonian

$$H = \omega_0 N - \tfrac{1}{2}\gamma(|a_1|^4 + |a_2|^4) - J(a_1^* a_2 + a_2^* a_1)$$

(* denotes the complex conjugate) are conserved by eq. (1); for two freedoms the DST is an integrable system.

An important class of solutions of eq. (1) are the *stationary solutions*, for which

$$a_j = a_j(0)\,e^{i\omega t} , \tag{3}$$

with the same constant frequency $\omega$ for all $j$ [1]. For $0 \le \zeta < 2$, there are two stationary solutions: one is symmetric ($a_1 = a_2$) and the other is antisymmetric ($a_1 = -a_2$); each distributes energy equally between the two sites. However, at $\zeta = 2$ there is a bifurcation. For $\zeta > 2$ the symmetric stationary solution is unstable and two new stable solutions are found. As $\zeta$ increases, these two other stationary solutions become the lowest in total energy and each concentrates more and more energy on one of the two sites. For these two *local mode* stationary solutions [1],
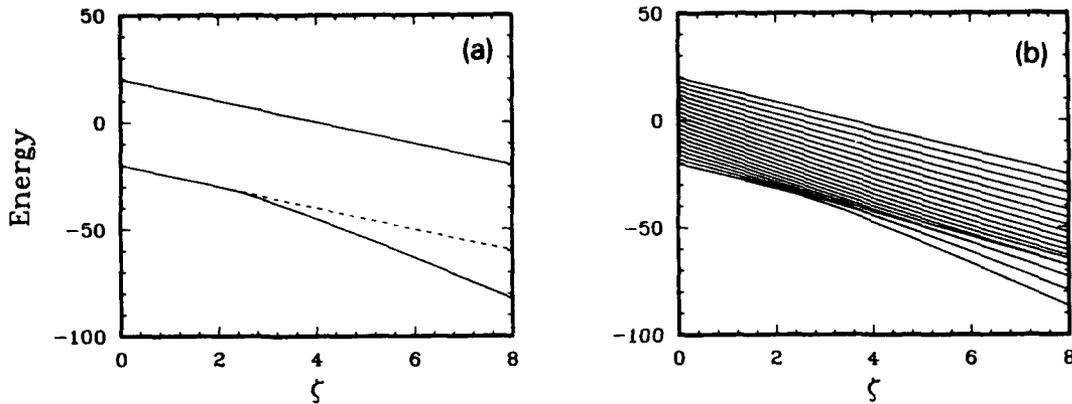
Fig. 1. Quantum and classical transitions. (a) The energies of stationary solutions of eq. (1) vs. $\zeta$ for $N = 20$, $J = 1$, and $\omega_0 = 0$. The dashed curve indicates that the symmetric stationary solution is unstable above $\zeta = 2$. (b) Quantum energy levels vs. $\zeta$ for the quantum level $n = 20$, $J = 1$ and $\omega_0 = 0$.

$$|a_1|^2 - |a_2|^2 = \pm N\sqrt{1 - \frac{4}{\zeta^2}}. \tag{4}$$

Although the two oscillators are identical, eq. (4) shows that above the $\zeta = 2$ self-trapping transition, vibrational amplitude – and hence energy – can be *localized*. Figure 1a shows how the total energies of all the stationary solutions of eq. (1) vary with $\zeta$.

This paper describes recent efforts to understand how the classical self-trapping transition at $\zeta = 2$ is reflected in the energy levels and stationary state wavefunctions of the corresponding quantum-mechanical problem (QDST).

## 2. Quantum energy levels

Because solutions of the classical DST equation conserve the norm $N$, solutions of the corresponding quantum-mechanical problem conserve the total number $n$ of vibrational quanta (bosons). This makes it easy to quantize eq. (1) [2]. The classical Hamiltonian $H$ becomes the quantum Hamiltonian operator

$$\hat{H} = (\omega_0 - \tfrac{1}{2}\gamma)(\hat{B}_1^\dagger \hat{B}_1 + \hat{B}_2^\dagger \hat{B}_2)$$

$$- \tfrac{1}{2}\gamma[(\hat{B}_1^\dagger \hat{B}_1)^2 + (\hat{B}_2^\dagger \hat{B}_2)^2] - J(\hat{B}_1^\dagger \hat{B}_2 + \hat{B}_2^\dagger \hat{B}_1),$$

where $\hat{B}_i^\dagger$ ($\hat{B}_i$) is a quantum raising (lowering) operator for bosons on site $i$.

For each non-negative integer boson number $n$ there are $n + 1$ stationary quantum states which we label $\psi^{(m)}$, $m = 0, \ldots, n$. These states are represented in the $(n + 1)$-dimensional basis of product states $|n_1\rangle|n_2\rangle$, where $|n_i\rangle$ is the eigenfunction of a harmonic oscillator at site $i$ with quantum number $n_i$, and $n_1 + n_2 = n$. We write

$$\psi^{(m)} = \sum_{j=0}^{n} c_j^{(m)}|n - j\rangle|j\rangle. \tag{5}$$

Requiring that $\psi^{(m)}$ solves Schrödinger's linear wave equation $\hat{H}\psi^{(m)} = E^{(m)}\psi^{(m)}$ for a particular value of $n$, we get an $(n + 1) \times (n + 1)$ matrix eigenvalue problem for the stationary state energies $E^{(m)}$ and the corresponding coefficient vectors $\vec{c}^{(m)} = [c_0^{(m)}, \; c_1^{(m)}, \; c_2^{(m)}, \; \ldots, \; c_n^{(m)}]^{\mathrm{T}}$, $m = 0, \ldots, n$.

Computed energy levels for $n = 20$ are plotted vs. $\zeta$ in fig. 1b. Near the transition point $\zeta = 2$, the curves for the lowest levels change slope and combine into nearly-degenerate pairs. The classical bifurcation at $\zeta = 2$ is clearly manifested in changes in the corresponding quantum energy spectrum.

Analytical estimates for the $n + 1$ energy levels and eigenfunctions at the $n$th quantum level in the large-$\zeta$ limit can be obtained by writing the

quantum Hamiltonian operator $\hat{H} = \hat{H}_0 + J\hat{H}_1$. Degenerate regular perturbation theory gives

$$E^{(m)} = (\omega_0 - \tfrac{1}{2}\gamma)n - \tfrac{1}{2}\gamma[(n-m)^2 + m^2]$$

$$+ \frac{J^2}{\gamma}\left(\frac{m(n-m+1)}{n-2m+1} - \frac{(n-m)(m+1)}{n-2m-1}\right)$$

$$+ \mathcal{O}(J^3) \qquad (6)$$

for the energy levels, $m = 0, \ldots, n$. The corresponding zero-order wavefunctions are also found up to $\mathcal{O}(J)$.

Perturbation calculations for small $\zeta$ can also be carried out, by representing the problem in a new basis that diagonalizes $\hat{H}$ in the limit $\zeta = 0$. Writing $\hat{H} = \hat{H}_0 + \gamma\hat{H}_1$ and using non-degenerate regular perturbation theory, we obtain for $m = 0, \ldots, n$ energies

$$E^{(m)} = \omega_0 n + J(2m - n) - \tfrac{3}{4}\gamma n$$

$$- \tfrac{1}{4}\gamma(n^2 + 2nm - 2m^2) + \frac{\gamma^2(2m-n)}{32J}$$

$$\times [2m(n-m) + n - 1] + \mathcal{O}(\gamma^3) \qquad (7)$$

along with the $n + 1$ corresponding wavefunctions up to $\mathcal{O}(\gamma)$.

In using coupled anharmonic oscillator models such as the QDST to describe local modes in molecules, the lowest energy levels for a given value of $n$ are especially significant; these levels correspond to the predominant peaks in experimental absorption spectra [9]. In our notation for both large- and small-$\zeta$ perturbation theories, the lowest level for a given $n$ is $E^{(0)}$. In fig. 2 the small $\gamma$ perturbation result, the small $J$ perturbation result, and the exact (numerical) value of $E^{(0)}$ are plotted vs. $\zeta$ for $n = 5$. For $\zeta < 2$, the small-$\gamma$ perturbation result and the exact result are so close that they appear as a single line; the same is true for the small-$J$ perturbation result and the exact $E^{(0)}$ for $\zeta > 2$. Patching together the two second-order perturbation theory results appears to give a good approximation for the $E^{(0)}$ level for all $\zeta$.
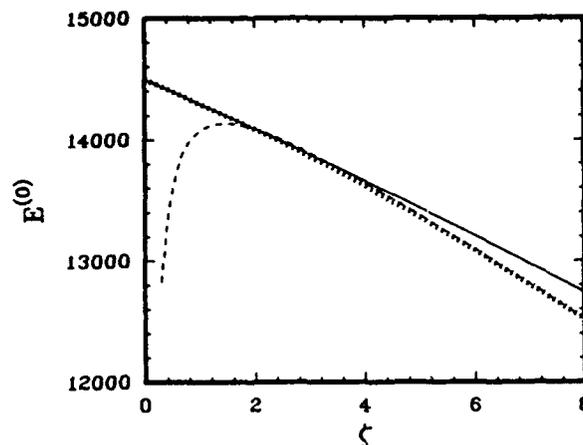


Fig. 2. $E^{(0)}$ vs. $\zeta$: exact value ($\cdots$), small $\gamma$ perturbation theory (—) and small $J$ perturbation theory (-----). The parameters $\omega_0 = 3000$ and $J = 100$ and the quantum level $n = 5$ used here are realistic for modeling carbon–hydrogen stretch vibrations in molecules.

## 3. Energy level splittings

For $\zeta \geqslant 2$ and $n > 2$, the two lowest-energy quantum eigenstates for a given $n$ are nearly degenerate and can combine to form a long-lived, energy-localizing wavepacket [3]. Here we consider more general values of $\zeta$. In fig. 3, the energy splitting $\Delta E$ between the lowest two energy levels of the QDST dimer is given as a function of $\zeta$ for several $n$ values. These numerical results suggest that in the classical limit
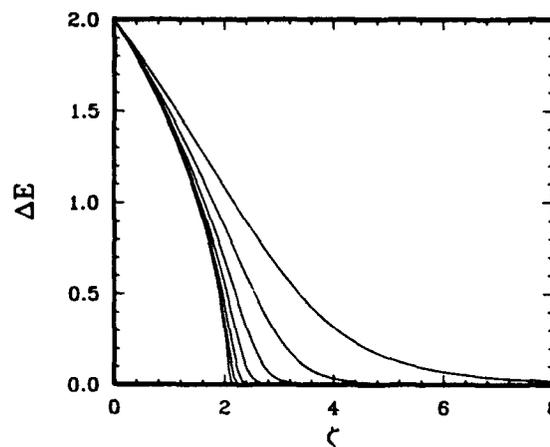


Fig. 3. Computed splitting $\Delta E$ between the two lowest QDST energy levels for a given $n$ vs. $\zeta$ for (from right to left) $n = 6$, 12, 25, 50, 100, 200 and 400, $J = 1$ and $\omega_0 = 0$.

$n \to \infty$, the two lowest levels become degenerate precisely at the classical bifurcation point $\zeta = 2$.

In the large $\zeta$ limit, a useful formula for the energy splitting $\Delta E$ is [3]

$$\Delta E \approx \frac{2n^2 \gamma n^n}{n! \zeta^n} . \tag{8}$$

Numerical tests [3] have shown that for fixed $J \ll \gamma$, the exact $\Delta E$ converges to the expression in eq. (8) for large $n$. It has been conjectured that eq. (8) is good for $\zeta > 2$.

Using Stirling's approximation for $n!$ in eq. (8), one can show that for large $n$, the exact $\Delta E$ diverges from the right side of eq. (8) for $\zeta \le e \approx 2.718$. Thus eq. (8) should apply to the QDST model for quantum levels $n > eJ/\gamma$. This has been verified numerically. Thus, the approximation in eq. (8) breaks down just above the classical bifurcation point.

## 4. Wavefunctions

Further insight into the quantum-mechanical dynamics of the QDST dimer can be obtained by considering the quantum wavefunctions. However, the problem of finding a quantum-mechanical analog of classical local modes is complicated by a quantum-mechanical symmetry condition: the nondegenerate stationary quantum states for systems of two identical oscillators must be symmetric or anti-symmetric with respect to exchanges of the two sites. Here we handle this complication in two ways: (i) by comparing classical local mode stationary solutions to non-stationary quantum wavepackets and (ii) by examining a certain property of the properly-symmetrized stationary quantum states that is related to classical energy localization.

### 4.1. Local mode wavepackets

Nonstationary energy-localizing quantum *wavepackets* can be created from superpositions

of the stationary states. For example, the lowest- and second-lowest-energy stationary states for a given value of $n$ are symmetric and anti-symmetric, respectively. A non-stationary wavepacket created by adding these two stationary states can initially concentrate energy on one of the two sites [3]. The degree of localization in such a non-stationary wavepacket can be characterized by the value of

$$\langle \hat{B}_1^\dagger \hat{B}_1 - \hat{B}_2^\dagger \hat{B}_2 \rangle |_{t=0} , \tag{9}$$

where $t = 0$ is the initial time. This is the quantum analog of the classical quantity $|a_1|^2 - |a_2|^2$. Numerical calculations for the quantum problem suggest that one can create for $\zeta \ge 2$ quantum wavepackets for which the value of $\langle \hat{B}_1^\dagger \hat{B}_1 - \hat{B}_2^\dagger \hat{B}_2 \rangle |_{t=0}$ approaches the right side of eq. (4) as $n \to \infty$. In this way the initial degree of localization in such quantum wavepackets approaches the degree of localization in the corresponding classical local mode stationary solutions.

Creating a wavepacket that initially concentrates energy on site 1 requires that symmetric and anti-symmetric stationary states be added with a specific relative phase at $t = 0$. Because the two stationary states have slightly different energies and hence evolve at different characteristic frequencies, the phase relationship established at $t = 0$ gradually changes and energy is no longer concentrated on the original site. For $\zeta$ above the self-trapping transition, the time $\Delta t$ for which the energy remains concentrated on the initial site increases with increasing $n$, because the energy splitting $\Delta E$ between the two stationary state components decreases (see fig. 3). For large $n$ and $\zeta \ge 2$, the initial phase relation takes a long time to decay. Such quantum wavepackets are the analogs of the classical local modes.

### 4.2. Stationary state wavefunctions

There are a number of reasons to consider properties of QDST wavefunctions other than the characteristics of the local mode wavepackets

discussed above. First, it is desirable to under-
stand how all types of classical solutions (not just
local mode stationary solutions) of eq. (1) are
connected to exact solutions of the corre-
sponding quantum problem. Also, it is important
from the point of view of applications to consider
the case where a system remains in a superposi-
tion of stationary states for a relatively long time
but external forces lead to a rapid randomization
of the phase relations between the wavepacket
components [10]; in such cases the properties of
the stationary states are more significant than are
wavepacket properties (such as (9)) that depend
on the initial phase relations. This motivates an
examination of how the individual quantum
stationary states change as the anharmonicity
parameter $\zeta$ is tuned through the self-trapping
transition.

It is useful to characterize stationary quantum
states $\psi^{(m)}$ using the quantum analog of the
classical time-average of $||a_1|^2 - |a_2|^2|$. The ana-
log for stationary quantum states is the diagnos-
tic

$$D^{(m)} = \frac{1}{n} \sum_{j=0}^{n} |c_j^{(m)}|^2 |n - 2j| ,$$  (10)

a measure of the average difference between the
excitation levels of the two sites in the dimer that
does not distinguish which site is more highly
excited. The values of $D^{(m)}$ are easily calculated.
The number $c_j^{(m)}$ in eq. (10) is the coefficient of
$|n - j\rangle|j\rangle$ in the basis-state expansion of the
stationary wavefunction $\psi^{(m)}$. The factor $|n - 2j|$
gives the magnitude of the difference between
the quantum level at site 1 and the quantum
level at site 2 in that basis state. A value of $D^{(m)}$
near 1 indicates that when the system is in the
stationary state $\psi^{(m)}$, there is a high probability
that one site is more excited than the other.

In fig. 4, $D^{(m)}$ is plotted vs. $\zeta$ for each of the
21 stationary quantum states $\psi^{(m)}$ for the case
$n = 20$. Note that $D^{(m)}$ can be nonzero even
when $\zeta = 0$; the stationary quantum states with
the smallest values of $D^{(m)}$ for $\zeta = 0$ correspond
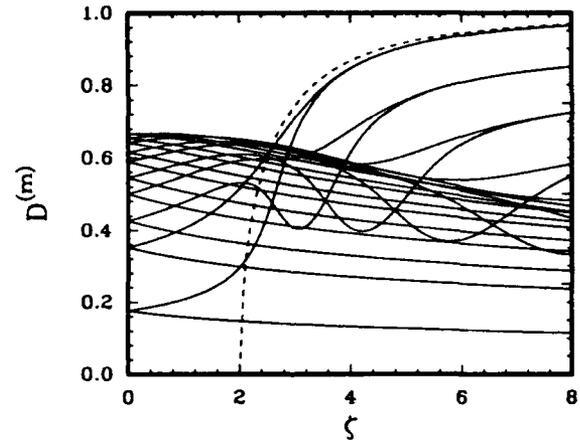


Fig. 4. Solid lines: the localization measure $D^{(m)}$ vs. $\zeta$ for the
21 stationary quantum states at the quantum level $n = 20$,
$J = 1$, $\omega_0 = 0$. Dashed line: $||a_1| - |a_2||$ for classical local
mode stationary solutions.

to the classical symmetric and anti-symmetric
normal modes. The stationary quantum states
with the largest values of $D^{(m)}$ when $\zeta = 0$
correspond to non-stationary classical solutions
in which 100% of the energy oscillates between
the two sites.

As $\zeta$ increases above 2, pairs of stationary
states develop large values of $D^{(m)}$. These pairs
of stationary state wavefunctions are those with
which one can build an energy-localizing quan-
tum wavepacket; they correspond to the pairs of
nearly-degenerate energy levels at the bottom of
fig. 1b. For large values of $n$, the uppermost
solid curve in fig. 4 for $\zeta \geq 2$ is well-approxi-
mated by the absolute value of the right side of
eq. (4) with $N = n$. The classical expression is
plotted with a dashed line in fig. 4 for com-
parison.

## 5. Discussion

Solutions of the fully quantum-mechanical
version of the DST dimer reflect the classical
self-trapping transition in a number of different
ways. Numerically-calculated quantum energy
levels manifest the classical self-trapping bifurca-
tion most notably in the quasi-degeneracy that

occurs above the transition point between the two lowest levels for a given boson number $n$. Carefully constructed quantum wavepackets and even some properly-symmetrized stationary quantum states (if characterized by the quantity $D^{(m)}$) echo the behavior of the classical local modes for $\zeta \geq 2$. The results presented here illustrate connections between quantum and classical behavior that are likely to exist for other anharmonic oscillator models as well.

A main motivation for studying the quantum mechanics of coupled anharmonic oscillator systems is to obtain a better understanding of the bond vibrations in polyatomic molecules. Further knowledge about modeling bond vibrations is needed for theoretical chemistry problems such as interpreting molecular spectra [11], understanding chemical reactions involving vibrationally-excited molecules [12] and developing techniques for bond-specific laser photochemistry [13,14].

The DST equation has a useful role in the progress of such research because, unlike some other commonly-studied anharmonic oscillator systems used in physical chemistry, the DST can be quantized without approximation. Comparisons between solutions of the classical and quantum versions of the DST are thus not clouded by questions about the approximate nature of the quantum solutions. Further, the exact-quantization property holds not only for the DST dimer but also for DST equations representing a great variety of coupled-oscillator systems such as $f$-freedom oscillator chains [15] and systems of non-identical oscillators [16]. As a simple model that captures some fundamental physics without sacrificing analytical tractability, the DST will continue to provide insight into the nonlinear and quantum-mechanical dynamics of molecules.

## References

[1] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, Physica D 16 (1985) 318.

[2] A.C. Scott and J.C. Eilbeck, Phys. Lett. A 119 (1986) 60.

[3] L. Bernstein, J.C. Eilbeck and A.C. Scott, Nonlinearity 3 (1990) 293.

[4] L. Cruzeiro-Hansson, H. Feddersen, R. Flesch, P.L. Christiansen, M. Salerno and A.C. Scott, Phys. Rev. B 42 (1990) 522.

[5] S. DeFilippo and M. Salerno, Phys. Lett. A 142 (1989) 479.

[6] A.C. Scott, P.S. Lomdahl and J.C. Eilbeck, Chem. Phys. Lett. 113 (1985) 29.

[7] A.C. Scott and J.C. Eilbeck, Chem. Phys. Lett. 132 (1986) 23.

[8] A.C. Scott, L. Bernstein and J.C. Eilbeck, J. Biol. Phys. 17 (1989) 1.

[9] I.A. Watson, B.R. Henry and I.G. Ross, Spectrochim. Acta A 37 (1981) 857.

[10] S. Stenholm, Foundations of Laser Spectroscopy (Wiley, New York, 1984) p. 36ff.

[11] M.E. Kellman and L. Xiao, Chem. Phys. Lett. 162 (1989) 486.

[12] T. Uzer and W.H. Miller, Phys. Rep. 199 (1991) 73.

[13] C.D. Schwieters and H. Rabitz, Phys. Rev. A 44 (1991) 5224.

[14] B. Hartke, A.E. Janza, W. Karrlein, J. Manz, V. Mohan and H.-J. Schreier, J. Chem. Phys. 96 (1992) 3569.

[15] A.C. Scott, L. Bernstein and J.C. Eilbeck, J. Biol. Phys. 17 (1989) 1.

[16] A.C. Scott, Phys. Scr. 42 (1990) 14.

## Acknowledgements

PHYSICA D

# On a modified discrete self-trapping dimer

Michael F. Jørgensen, Peter L. Christiansen and Imad Abou-Hayt

*Laboratory of Applied Mathematical Physics, The Technical University of Denmark, DK-2800 Lyngby, Denmark*

Conditions for integrability of a modified discrete self-trapping dimer are derived. Dynamical properties of three special cases are investigated. Blow-up in finite time may occur in a non-compact system.

## 1. Introduction

The discrete self-trapping (DST) equation was introduced in [1] as a general model for energy self-trapping in molecules and molecular crystals. The DST equation is only integrable for two degrees of freedoms.

The DST equation includes as a special case the discrete nonlinear Schrödinger equation which is integrable in the continuum limit too [2]. An alternative discretisation of the nonlinear Schrödinger (NLS) equation is the Ablowitz–Ladik (AL) equation [3], which is integrable for any number of freedoms.

Several generalisations of the DST equation have been studied in the literature, cf. [4]. It has recently been shown in [5] that a Hamiltonian and a non-standard Poisson bracket, depending on a deformation parameter, can reproduce the discrete NLS equation and the AL equation in two different limits.

In this article we consider the following combination of the DST equation and the AL equation with two degrees of freedom and periodic boundary conditions:

$$i\dot{A}_1 = \alpha_1 |A_2|^2 A_1 + \beta_1 |A_2|^2 A_2 + \gamma_1 |A_1|^2 A_1$$
$$+ \delta_1 |A_1|^2 A_2 + \epsilon_1 A_2 \,,$$
$$i\dot{A}_2 = \alpha_2 |A_1|^2 A_2 + \beta_2 |A_1|^2 A_1 + \gamma_2 |A_2|^2 A_2$$
$$+ \delta_2 |A_2|^2 A_1 + \epsilon_2 A_1 \,, \qquad (1)$$

where $A_1$ and $A_2$ are complex functions of time, dot denotes differentiation with respect to time, and the coefficients $\alpha_1, \beta_1, \ldots, \epsilon_2$ are a priori complex constants. Terms of the form $\gamma_1 |A_1|^2 A_1$, $\gamma_2 |A_2|^2 A_2$, $\epsilon_1 A_2$ and $\epsilon_2 A_1$ occur in the DST system [1], terms of the form $\delta_1 |A_1|^2 A_2$, $\delta_2 |A_2|^2 A_1$, $\epsilon_1 A_2$ and $\epsilon_2 A_1$ are found in the AL system [3], while terms of the form $\alpha_1 |A_2|^2 A_1$ and $\alpha_2 |A_1|^2 A_2$ arise in the coupling between two pulses in optical fibres [6,7]. The remaining terms are new.

The aim of this paper is to study the integrability of various special cases of the general system (1). In the next section we derive three special cases which are Hamiltonian and completely integrable. In section three we investigate the solutions in each of the three cases. We have explicitly shown that one of the cases possesses solutions which are not globally defined, i.e., they tend to infinity in finite time (blow-up).

## 2. Deriving the special cases

The general system (1) is only integrable when it has two conserved quantities. We already know two special cases which are integrable, the DST and the AL case. The DST dimer has two conserved quantities, one of which is the norm (or number)

$$N = |A_1|^2 + |A_2|^2 . \tag{2}$$

Equivalently, one of the conserved quantities for the AL dimer is

$$r_1 = A_1 A_2^* + A_1^* A_2 . \tag{3}$$

This last quantity is one of the three Feynman variables introduced in [8], hence the notation $r_1$.

Since the system (1) is more general than the DST and the AL, we consider the remaining two Feynman variables too. These are

$$r_2 = i(A_1 A_2^* - A_1^* A_2) , \tag{4}$$

$$r_3 = |A_1|^2 - |A_2|^2 . \tag{5}$$

We then demand one of these four quantities to be conserved and thereby derive the four sets of conditions on the coefficients of (1) given in table 1. The four quantities $N$, $r_1$, $r_2$ and $r_3$ satisfy the following fundamental relation:

$$N^2 = r_1^2 + r_2^2 + r_3^2 . \tag{6}$$

Using this relation it is possible, in each of the four cases, to rewrite the dynamical system (1) in the following way ("the Feynman top formalism")

$$\dot{L} = L \times \omega , \tag{7}$$

where $L$ is a three-component vector constructed from the three remaining (non-conserved) quan-

tities, and $\omega$ is a three-component vector related to $L$ through an affine map

$$\omega = C + \mathbf{D} \cdot L , \tag{8}$$

where $C$ is a constant vector and $\mathbf{D}$ is a constant matrix. The vector $L$ in the four cases is given in table 2. In each case $L_x^2 + L_y^2 + L_z^2$ is constant due to the fundamental relation (6), in agreement with eq. (7). Having rewritten the dynamical system in the form of (7) and (8) it is easy to construct a second conserved quantity using an analogy with classical mechanics [9]. Generalising the kinetic energy we obtain the quantity

$$T = L^T \cdot C + \tfrac{1}{2} L^T \cdot \mathbf{D} \cdot L , \tag{9}$$

where $^T$ denotes transposition. This quantity is indeed conserved, provided the matrix $\mathbf{D}$ is symmetric. The conditions for symmetry of the matrix $\mathbf{D}$ in each case is reproduced in table 3. We now have four different special cases of (1), each with two conserved quantities. The final step is to apply a gauge transform in order to simplify the equations. The gauge transform used is of the form

Table 1
Conditions on coefficients.

| Conserved quantity | Conditions on coefficients |
|---|---|
| $N$ | $\gamma_1$, $\gamma_2$ real; $\mathrm{Im}\,\alpha_1 = -\mathrm{Im}\,\alpha_2$; $\delta_1 = \beta_2^*$, $\beta_1 = \delta_2^*$; $\epsilon_1 = \epsilon_2^*$ |
| $r_1$ | $\gamma_1 = \alpha_2^*$, $\alpha_1 = \gamma_2^*$; $\mathrm{Im}\,\delta_1 = -\mathrm{Im}\,\delta_2$; $\beta_1$, $\beta_2$ real; $\epsilon_1$, $\epsilon_2$ real |
| $r_2$ | $\gamma_1 = \alpha_2^*$, $\alpha_1 = \gamma_2^*$; $\mathrm{Re}\,\delta_1 = \mathrm{Re}\,\delta_2$; $\beta_1$, $\beta_2$ imaginary; $\epsilon_1$, $\epsilon_2$ imaginary |
| $r_3$ | $\gamma_1$, $\gamma_2$ real; $\mathrm{Im}\,\alpha_1 = \mathrm{Im}\,\alpha_2$; $\delta_1 = -\beta_2^*$, $\beta_1 = -\delta_2^*$; $\epsilon_1 = -\epsilon_2^*$ |

Table 2
The vector $L$.

| Conserved quantity | $L = (L_x, L_y, L_z)$ |
|---|---|
| $N$ | $(r_1, r_2, r_3)$ |
| $r_1$ | $(N, ir_2, ir_3)$ |
| $r_2$ | $(N, ir_1, ir_3)$ |
| $r_3$ | $(N, ir_1, ir_2)$ |

Table 3
Conditions for two conserved quantities.

| Conserved quantity | Conditions for symmetry of $\mathbf{D}$ |
|---|---|
| $N$ | $\alpha_1$ real, $\beta_1 = \delta_1$ |
| $r_1$ | $\gamma_1$, $\gamma_2$ real, $\mathrm{Re}\,\delta_1 = \mathrm{Re}\,\delta_2$ |
| $r_2$ | $\alpha_1$, $\gamma_1$ real, $\mathrm{Im}\,\delta_1 = -\mathrm{Im}\,\delta_2$ |
| $r_3$ | $\alpha_1$ real, $\beta_1 = -\delta_1$ |

$$A_j(t) \mapsto A_j(t)$$

$$\times \exp\left(-i \int_{t_0}^{t} (\kappa_1 |A_1|^2 + \kappa_2 |A_2|^2)\, dt\right), \qquad j = 1, 2,$$

where $t_0$, $\kappa_1$ and $\kappa_2$ are real constants. Note that this transformation preserves the form of the four quantities $N$, $r_1$, $r_2$ and $r_3$.

By examining the simplified equations it is easily shown that the case $\dot{r}_2 = 0$ can be mapped linearly to the case $\dot{r}_1 = 0$. We are finally left with the three cases $\dot{N} = 0$, $\dot{r}_1 = 0$ and $\dot{r}_3 = 0$ [10]. In the following we list them in their simplified form together with the two conserved quantities:

## Case 1
The dynamical equations are

$$i\dot{A}_1 = \gamma_1 |A_1|^2 A_1 + \epsilon A_2,$$

$$i\dot{A}_2 = \gamma_2 |A_2|^2 A_2 + \epsilon^* A_1,$$

where $\gamma_1$ and $\gamma_2$ are real constants and $\epsilon$ is a complex constant. The two conserved quantities are

$$N = |A_1|^2 + |A_2|^2,$$

$$H = \tfrac{1}{2}\gamma_1 |A_1|^4 + \tfrac{1}{2}\gamma_2 |A_2|^4 + \epsilon A_1^* A_2 + \epsilon^* A_1 A_2^*.$$

This system reduces to the DST dimer [1] for $\gamma_1 = \gamma_2$ and $\epsilon$ real. For $\gamma_1 \neq \gamma_2$ the system is equivalent [11] with the non-resonant DST dimer studied by Scott in [12]. The quantum problem for this equation is studied in [13].

## Case 2
The dynamical equations are

$$i\dot{A}_1 = \gamma_1 |A_1|^2 A_1 + \epsilon A_2,$$

$$i\dot{A}_2 = \gamma_2 |A_2|^2 A_2 - \epsilon^* A_1,$$

where $\gamma_1$ and $\gamma_2$ are real constants and $\epsilon$ is a complex constant. The two conserved quantities are

$$r_3 = |A_1|^2 - |A_2|^2,$$

$$H = \tfrac{1}{2}\gamma_1 |A_1|^4 - \tfrac{1}{2}\gamma_2 |A_2|^4 + \epsilon A_1^* A_2 + \epsilon^* A_1 A_2^*.$$

It can be shown [11] that this system is a specialisation of case 3. In [14] a Lax operator satisfying the quadratic Sklyanin algebra [15] is given.

## Case 3
The dynamical equations are

$$i\dot{A}_1 = \gamma |A_1|^2 A_2 + \beta_1 |A_2|^2 A_2 + \epsilon_1 A_2 + i\alpha A_1,$$

$$i\dot{A}_2 = \gamma |A_2|^2 A_1 + \beta_2 |A_1|^2 A_1 + \epsilon_2 A_1 - i\alpha A_2,$$

where $\gamma$, $\beta_1$, $\beta_2$, $\epsilon_1$, $\epsilon_2$ and $\alpha$ are real constants. The two conserved quantities are

$$r_1 = A_1 A_2^* + A_1^* A_2,$$

$$H = \tfrac{1}{2}\beta_2 |A_1|^4 + \tfrac{1}{2}\beta_1 |A_2|^4 + \epsilon_2 |A_1|^2 + \epsilon_1 |A_2|^2$$
$$+ \gamma |A_1|^2 |A_2|^2 + i\alpha (A_1 A_2^* - A_1^* A_2).$$

For $\beta_1 = \beta_2 = 0$, $\epsilon_1 = \epsilon_2$ and $\alpha = 0$ this system reduces to the AL case [3]. Choosing instead $\gamma = \beta_1 = \beta_2$ the reduced system is equivalent to case 2. However, where the general system, case 3, possesses blow-up solutions, this is not so for case 2.

This concludes the derivation of the three special cases of (1).

## 3. Properties of the three dynamical systems

In this section we will examine the integrability of the three cases and investigate the properties of the solutions.

All three cases are Hamiltonian with the second conserved quantity $H$ as the Hamiltonian. The Poisson structure is standard and the canonical variables are $(iA_1^*, iA_2^*, A_1, A_2)$, $(iA_1^*, A_2, A_1, iA_2^*)$ and $(iA_1^*, iA_2^*, A_2, A_1)$ respectively.

The three cases are all completely integrable [16]. Indeed, defining $N_1 \equiv |A_1|^2$ we get in case 1:

$$(\dot{N}_1)^2 = -[H - \tfrac{1}{2}\gamma_1 N_1^2 - \tfrac{1}{2}\gamma_2(N - N_1)^2]^2$$
$$+ 4|\epsilon|^2 N_1(N - N_1) , \qquad (10)$$

and in case 2:

$$(\dot{N}_1)^2 = -[H - \tfrac{1}{2}\gamma_1 N_1^2 + \tfrac{1}{2}\gamma_2(N - N_1)^2]^2$$
$$+ 4|\epsilon|^2 N_1(N - N_1) . \qquad (11)$$

So in both cases $(\dot{N}_1)^2$ is a fourth degree polynomial in $N_1$, and hence the solution $N_1(t)$ may be expressed in terms of Jacobi elliptic functions [17].

In case 1 all solutions are bounded in infinite time, since $N = |A_1|^2 + |A_2|^2$ is a constant. In case 2 the coefficient to $N_1^4$ in (11) is $-\tfrac{1}{4}(\gamma_1 - \gamma_2)^2$, i.e., negative. Hence all solutions are bounded for $\gamma_1 \neq \gamma_2$. For $\gamma_1 = \gamma_2$ the right-hand side of (11) becomes a second degree polynomial in $N_1$ with the highest coefficient $4|\epsilon|^2 - \gamma_1^2 r_3^2$. The solutions to (11) are in this case either sinusoidal or exponential, and hence they are bounded in finite time.

However, in case 3 the solutions can generally not be written in terms of Jacobi elliptic functions. Nor are the solutions globally defined, i.e., bounded in all finite time; some solutions exhibit blow-up. In the following we investigate case 3 in more detail.

By examining the non-compact manifolds $r_1 = $ const. and $H = $ const. we quickly derive sufficient conditions for bounded solutions. These are

– either $\gamma$, $\beta_1$ and $\beta_2$ have the same sign ,    (12)

– or $\beta_1\beta_2 > \gamma^2$ .    (13)

For $\beta_1 = \beta_2 \equiv \beta$ and $\epsilon_1 = \epsilon_2 \equiv \epsilon$, the dynamical equations for case 3 can be written as

$$\dot{N} = (\gamma - \beta)r_2 r_3 + 2\alpha r_3 ,$$

$$\dot{r}_2 = -2\epsilon r_3 - 2\beta N r_3 ,$$

$$\dot{r}_3 = (\gamma + \beta)N r_2 + 2\alpha N + 2\epsilon r_2 ,  \qquad (14)$$

where we have reintroduced the Feynman variables (2)–(5). Recalling that $N^2 = r_1^2 + r_2^2 + r_3^2$, we find that

$$(\dot{N})^2 = (N^2 - r_1^2 - r_2^2)[(\gamma - \beta)r_2 + 2\alpha]^2 , \qquad (15)$$

where $r_2$ is related to Hamiltonian through

$$H = \tfrac{1}{2}\beta N^2 + \epsilon N + \tfrac{1}{4}(\delta - \beta)r_2^2 + \alpha r_2$$
$$+ \tfrac{1}{4}(\delta - \beta) r_1^2 ,$$

with $r_1$ being the first conserved quantity.

Simple calculations show that $(\dot{N})^2$ can be expressed as

$$(\dot{N}) = P_4(N) + R_2(N)\sqrt{Q_2(N)} , \qquad (16)$$

where $P_4(N)$ is a fourth-degree polynomial in $N$, and $R_2(N)$ and $Q_2(N)$ are second degree polynomials. For $\alpha = 0$, $(\dot{N})^2$ becomes a fourth-degree polynomial in $N$. Thus, in this special case the Feynman variable $N$ can be written in terms of Jacobi elliptic functions [17]. In the general case eq. (16) must be solved numerically.

Using Mathematica [18] we get the bounded
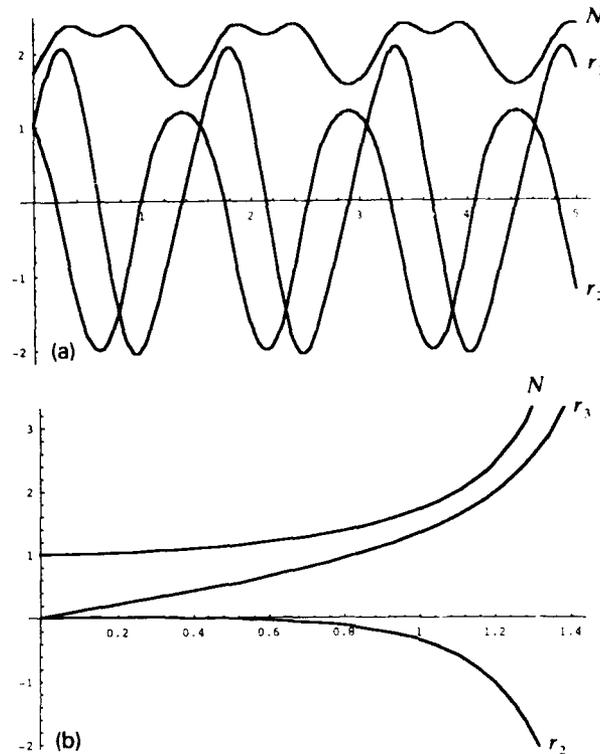


Fig 1. Solution of eqs. (14) with (a) $\gamma = 2$, $\beta = 1$, $\epsilon = 0.5$, $N(0) = 1.732$, $r_2(0) = 1$, $r_3(0) = 1$, $r_1 = 1$; (b) $\gamma = -1$, $\beta = 1$, $\epsilon = 1$, $\alpha = 0.5$, $N(0) = 1$, $r_2(0) = 0$, $r_3(0) = 0$, $r_1 = 1$.

solution shown in fig. 1a for parameter values and initial conditions as indicated in agreement with (12). Alternatively, fig. 1b illustrates an unbounded solution. Numerically, the blow-up was found to occur at time $t = 1.684$.

For $\epsilon = 0$ and $\alpha = 0$, eq. (15) reduces to

$$(\dot{N})^2 = (C_1 - 2\beta N^2)[(\gamma + \beta)N^2 - C_2],\qquad (17)$$

where the conserved quantities are given by

$$C_1 = 2\beta N^2 + (\gamma - \beta)r_2^2,$$

$$C_2 = (\gamma + \beta)N^2 - (\gamma - \beta)r_3^2.$$

Equation (17) may be integrated analytically and the solution is

$$N = N(t) = \sqrt{\frac{C_2}{\beta + \gamma}}$$

$$\times \mathrm{ns}\left(\sqrt{-2\beta C_2}\,(t_0 - t), \sqrt{\frac{C_1}{C_2}\frac{\beta + \gamma}{2\beta}}\right),\qquad (18)$$

where $\mathrm{ns}(u, k)$ is the reciprocal of the Jacobi elliptic function $\mathrm{sn}(u, k)$ of argument $u$ and modulus $k$ [17]. In (18), $t_0$ is the blow-up time determined by the initial conditions as

$$t_0 = \frac{1}{\sqrt{-2\beta(\beta + \gamma)}}$$

$$\times \int_{N(0)}^{\infty} \frac{\mathrm{d}N}{\sqrt{[N^2 - C_2/(\beta + \gamma)](N^2 - C_1/2\beta)}}.$$

$$(19)$$

This quantity is real and finite for $\gamma > -\beta > 0$ in agreement with the failure of the two conditions (12) and (13).

In the limit $\beta \to 0^-$, which corresponds to the AL case, the blow-up time $t_0$ approaches infinity logarithmically [16].

## Acknowledgement

## References

[1] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, Physica D 16 (1985) 318–358.
[2] V.E. Zakharov and A.B. Shabat, Sov. Phys. JETP 34 (1972) 62.
[3] M.J. Ablowitz and J.F. Ladik, J. Math. Phys. 17 (1976) 1011; Stud. Appl. Math. 55 (1976) 213.
[4] A.C. Scott and P.L. Christiansen, Phys. Scri. 42 (1990) 257.
[5] M. Salerno, Quantum Deformations of the Discrete Nonlinear Schrödinger Equation, preprint (1992).
[6] D.F. Parker and C. Sophocleous, Pulse Collisions in Bimodal Wave-guides, in: Proc. on Future Directions of Nonlinear Dynamics in Physical and Biological Systems (Plenum, 1993).
[7] P.S. Spencer and K.A. Shore, Dynamic Response of Semiconductor Nonlinear Optical Waveguides, Proceedings on Future Directions of Nonlinear Dynamics in Physical and Biological Systems (Plenum, 1993).
[8] R.P. Feynman, F.L. Vernon and R.W. Hellwarth, J. Appl. Phys. 28 (1957) 49.
[9] H. Goldstein, Classical Mechanics, 2nd Ed. (Addison-Wesley, MA, 1980) [ISBN 0-201-02969-3], Formulae (5-9) and (5-17).
[10] M.F. Jørgensen and P.L. Christiansen, Hamiltonian structure for a modified discrete self-trapping dimer, Nonlinearity, submitted (1992).
[11] V.B. Kuznetsov, Private communication (December 1992).
[12] A.C. Scott, Phys. Scri. 42 (1990) 14–18.
[13] V.Z. Enol'skii, V.B. Kuznetsov and M. Salerno, On the Quantum Inverse Scattering Method for the DST Dimer, in: Proc. on Future Directions of Nonlinear Dynamics in Physical and Biological Systems, Physica D 68 (1993) 138–152, these proceedings.
[14] P.L. Christiansen, M.F. Jørgensen and V.B. Kuznetsov, On integrable system close to the Toda lattice, Lett. Math. Phys., submitted (1993).
[15] P.P. Kulish and E.K. Sklyanin, Lecture Notes in Physics, Vol. 151 (1982) pp. 61–119.
[16] M.F. Jørgensen, Integrability and Blow-up of Nonlinear Differential Equations, Master's Thesis, Laboratory of Applied Mathematical Physics, The Technical University of Denmark, DK-2800 Lyngby, Denmark (1992).
[17] P.F. Byrd and M.D. Friedman, Handbook of Elliptic Integrals for Engineers and Scientists, 2nd Ed. (Springer, New York, 1971) [ISBN 0-387-05318-2].
[18] S. Wolfram, Mathematica, 2nd Ed. (Addison-Wesley, CA, 1991) [ISBN 0-201-51502-4].

# List of contributors

Abou-Hayt, I., 180

Bang, O., 169
Barbara, P., 35
Barthes, M., 45
Bernstein, L.J., 174
Bullough, R.K., 83

Chen, Y.-z, 83
Christiansen, P.L., 93, 169, 180
Cruzeiro-Hansson, L., 65

Dauxois, T., 104
Davidson, A., 35
Deering, W.D., 135
Doderer, T., 41
Duncan, D.B., 1

Eckert, J., 45
Eilbeck, J.C., 1
Enol'skii, V.Z., 138

Feddersen, H., 1
Filatrella, G., 35
Förner, W., 68

Grønbech-Jensen, N., 18

Hansen, P.J., 12

Hikami, K., 162
Holm, J., 35
Hoyet, H., 104
Huebener, R.P., 41

Jakobsen, P.K., 127
Johnson, S.W., 45
Jørgensen, M.F., 180

Kellouai, H., 45
Kenkre, V.M., 153
Krumhansl, J.A., 97
Kuznetsov, V.B., 138

Lega, J., 127
Lomdahl, P.S., 18
Lonngren, K.E., 12

Malomed, B.A., 38
McLaughlin, D.W., 116
Molina, M.I., 135
Moloney, J.V., 127
Moret, J., 45
Muraki, D., 116
Muto, V., 93
Mygind, J., 35

Nagao, T., 162
Newell, A.C., 127

Oboznov, V.A., 41

Page, G., 45
Pedersen, N.F., 27, 35, 38, 41
Peyrard, M., 104
Pierce, B.M., 51

Rasmussen, J.J., 169

Salerno, M., 138
Savin, A.V., 59
Shelley, M.J., 116
Sørensen, M.P., 38

Timonen, J.T., 83
Tsironis, G.P., 135

Ustinov, A.V., 38, 41

Vernik, I.V., 41

Wadati, M., 162
Wattis, J.A.D., 1
Wenden, S.G., 127
Willis, C.R., 104

Zolotaryuk, A.V., 59