Lecture Notes in Physics

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Physics and Astronomy ONLINE LIBRARY
http://www.springer.de/phys/

The Editorial Policy for Proceedings

The series Lecture Notes in Physics reports new developments in physical research and teaching – quickly, informally, and at a high level. The proceedings to be considered for publication in this series should be limited to only a few areas of research, and these should be closely related to each other. The contributions should be of a high standard and should avoid lengthy redraftings of papers already published or about to be published elsewhere. As a whole, the proceedings should aim for a balanced presentation of the theme of the conference including a description of the techniques used and enough motivation for a broad readership. It should not be assumed that the published proceedings must reflect the conference in its entirety. (A listing or abstracts of papers presented at the meeting but not included in the proceedings could be added as an appendix.)

When applying for publication in the series Lecture Notes in Physics the volume's editor(s) should submit sufficient material to enable the series editors and their referees to make a fairly accurate evaluation (e.g. a complete list of speakers and titles of papers to be presented and abstracts). If, based on this information, the proceedings are (tentatively) accepted, the volume's editor(s), whose name(s) will appear on the title pages, should select the papers suitable for publication and have them refereed (as for a journal) when appropriate. As a rule discussions will not be accepted. The series editors and Springer-Verlag will normally not interfere with the detailed editing except in fairly obvious cases or on technical matters.

Final acceptance is expressed by the series editor in charge, in consultation with Springer-Verlag only after receiving the complete manuscript. It might help to send a copy of the authors' manuscripts in advance to the editor in charge to discuss possible revisions with him. As a general rule, the series editor will confirm his tentative acceptance if the final manuscript corresponds to the original concept discussed, if the quality of the contribution meets the requirements of the series, and if the final size of the manuscript does not greatly exceed the number of pages originally agreed upon. The manuscript should be forwarded to Springer-Verlag shortly after the meeting. In cases of extreme delay (more than six months after the conference) the series editors will check once more the timeliness of the papers. Therefore, the volume's editor(s) should establish strict deadlines, or collect the articles during the conference and have them revised on the spot. If a delay is unavoidable, one should encourage the authors to update their contributions if appropriate. The editors of proceedings are strongly advised to inform contributors about these points at an early stage.

The final manuscript should contain a table of contents and an informative introduction accessible also to readers not particularly familiar with the topic of the conference. The contributions should be in English. The volume's editor(s) should check the contributions for the correct use of language. At Springer-Verlag only the prefaces will be checked by a copy-editor for language and style. Grave linguistic or technical shortcomings may lead to the rejection of contributions by the series editors. A conference report should not exceed a total of 500 pages. Keeping the size within this bound should be achieved by a stricter selection of articles and not by imposing an upper limit to the length of the individual papers. Editors receive jointly 30 complimentary copies of their book. They are entitled to purchase further copies of their book at a reduced rate. As a rule no reprints of individual contributions can be supplied. No royalty is paid on Lecture Notes in Physics volumes. Commitment to publish is made by letter of interest rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume.

The Production Process

The books are hardbound, and the publisher will select quality paper appropriate to the needs of the author(s). Publication time is about ten weeks. More than twenty years of experience guarantee authors the best possible service. To reach the goal of rapid publication at a low price the technique of photographic reproduction from a camera-ready manuscript was chosen. This process shifts the main responsibility for the technical quality considerably from the publisher to the authors. We therefore urge all authors and editors of proceedings to observe very carefully the essentials for the preparation of camera-ready manuscripts, which we will supply on request. This applies especially to the quality of figures and halftones submitted for publication. In addition, it might be useful to look at some of the volumes already published. As a special service, we offer free of charge LaTeX and TeX macro packages to format the text according to Springer-Verlag's quality requirements. We strongly recommend that you make use of this offer, since the result will be a book of considerably improved technical quality. To avoid mistakes and time-consuming correspondence during the production period the conference editors should request special instructions from the publisher well before the beginning of the conference. Manuscripts not meeting the technical standard of the series will have to be returned for improvement.

For further information please contact Springer-Verlag, Physics Editorial Department II, Tiergartenstrasse 17, D-69121 Heidelberg, Germany

Series homepage – http://www.springer.de/phys/books/lnpp

P. L. Christiansen M. P. Sørensen A. C. Scott (Eds.)

Nonlinear Science at the Dawn of the 21st Century

Springer

Editors

P. L. Christiansen
M. P. Sørensen
A. C. Scott
Department of Mathematical Modelling
The Technical University of Denmark
Building 321
2800 Kgs. Lyngby, Denmark

Library of Congress Cataloging-in-Publication Data applied for.

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Nonlinear science at the dawn of the 21st century / P. L. Christiansen ... (ed.). -
Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ;
Milan ; Paris ; Singapore ; Tokyo : Springer, 2000
(Lecture notes in physics ; 542)
ISBN 3-540-66918-3

ISSN 0075-8450
ISBN 3-540-66918-3 Springer-Verlag Berlin Heidelberg New York

Typesetting: Camera-ready by the authors/editors
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
SPIN:10720717 55/3144/du - 5 4 3 2 1 0

Remembering Bob

Alwyn Scott

In the summer of 1962, young Robert Dana Parmentier was finishing a master's thesis in the Department of Electrical Engineering at the University of Wisconsin, where it had been decided to support a major expansion of laboratory facilities in the rapidly developing area of solid state electronics. Jim Nordman and I—both spanking new PhDs—were put in charge of this effort, and we soon found ourselves involved in a variety of unfamiliar activities, including the slicing, polishing, cleaning, and doping of semiconductor crystals prior to the formation of *p*-*n* junctions by liquid and vapor phase epitaxy in addition to the more conventional process of dot alloying. We had much to learn, and welcomed Bob as a collaborator as he worked toward his doctorate in the area.

It was an exciting time, with research opportunities beckoning to us from several directions. From a more general perspective than had been originally contemplated by the Department, we began studying—both experimentally and theoretically—nonlinear electromagnetic wave propagation on semiconductor junctions with transverse dimensions large compared to a wave length. And there were many interesting nonlinear effects to consider.

Using ordinary reverse biased semiconductor diodes, the nonlinear capacitance of the junction causes shock waves, suggesting a means for generation of short pulses. At high doping levels, the junctions emit light to become semiconductor lasers, and at yet higher doping levels the negative conductance discovered by Leo Esaki appears, leading to a family of traveling wave amplifiers and oscillators. In 1966, this latter effect was also realized on insulating junctions between superconducting metals, rendered nonlinear through Ivar Giaever's tunneling of normal electrons.

As a basis for our theoretical work, we started with John Scott Russell's classic *Report on Waves*, a massive work that had been resting on a shelf of the University Library for well over a century, and in 1963 two events occurred that were to have decisive influences on Bob's professional life. The first of these was a Nobel Prize award to the British electrophysiologists Alan Hodgkin and Andrew Huxley for their masterful experimental, theoretical and numerical investigations of nonlinear wave propagation on a nerve fiber. This seminal work—to which applied mathematicians made no contributions whatsoever—pointed the way to Bob's doctoral research on the *neuristor*, a term recently coined for an electronic analog of a nerve axon.

The other event of 1963 was the experimental verification of Brian Josephson's prediction of tunneling by coupled electron pairs between superconducting metals, leading to an unusual sort of nonlinear inductor for which

current is a periodic function of the magnetic flux. From this effect, the relevant nonlinear wave equation for transverse electromagnetic waves on a strip-line structure takes the form

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{\partial^2 \phi}{\partial t^2} = \sin \phi \,, \tag{0.1}$$

where ϕ is a normalized measure of the magnetic flux trapped between the two superconducting strips.

Originally proposed in 1938 to describe dislocation dynamics in crystals and later to become widely known as the *sine-Gordon* equation, this is a nonlinear wave equation that conserves energy (which nerves and neuristors do not), and by the spring of 1966 we were aware that it carries little lumps of magnetic flux very much as Scott Russell's Great Wave of Translation transported lumps of water on the Union Canal near Edinburgh.

Just as Equation (0.1) can be viewed as a nonlinear augmentation of the standard wave equation, the system

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} = u(u-a)(u-1) \,, \tag{0.2}$$

is a nonlinear augmentation of the linear diffusion equation. Originally proposed in 1937 to describe the diffusion of genetic variations in spatially distributed populations, Equation (0.2) is the basic equation of excitable media, now known to have a variety of applications in chemistry and biology. Since it has a nonlinear traveling wave solution that represents the leading edge of a Hodgkin–Huxley nerve impulse, this equation is of central interest in the theory of a neuristor.

From a broader perspective, Equation (0.1) describes basic features of nonlinear wave propagation on *closed* (or energy conserving) systems, while Equation (0.2) plays the same role for *open* (or energy dissipating) systems; thus the two equations are fundamentally different and their traveling wave solutions have quite different behaviors. Equation (0.1) can be realized through Josephson tunneling and Equation (0.2) through both Esaki and Giaever tunneling. Interestingly, these three young researchers shared the Nobel Prize in physics in 1973.

Bob's doctoral research was concerned with both theoretical and experimental studies of these two equations, and his thesis was characterized by two unique features: it was entirely his own work and it was easily the shortest thesis that I have ever approved. Looking through *The Superconductive Tunnel Junction Neuristor* today, I am impressed by his simple and direct prose, and filled again with the delicious sense of how exciting was nonlinear science in those early days. So much was sitting just in front of us, waiting to be discovered.

This thesis was a tour de force, consisting of five distinct contributions.

- On the theoretical side, he introduced the idea of studying traveling

wave stability in a moving frame, using this concept to establish the stability of step (or level changing) solutions of Equation (0.2).

- Again theoretically, he considered an augmentation of Equation (0.2) with a realistic description of superconducting surface impedance, leading to the hitherto unexpected possibility of a pulse-shaped traveling wave. The existence of such a solution is important if the superconducting transmission line is to be employed as a neuristor; a fact recognized in US Patent Number 3,717,773 "Neuristor transmission line for actively propagating pulses," which was awarded on February 20, 1973.

- On the experimental side of his research, Bob constructed an electronic transmission line model of the superconducting neuristor—using Esaki tunnel diodes—demonstrating that his neuristor does indeed have pulse-like solutions. Nowadays, this sort of check would be done on a digital computer, but in the 1960s electronic modeling was an effective, if tedious, approach.

- Extending fabrication procedures previously developed in our laboratory, he constructed tin–tin oxide–lead superconducting tunnel transmission lines of the Giaever type, showing that they could function as neuristors by propagating traveling pulses as predicted by his theory. This part of the research was a major effort, involving the making of 80 superconducting transmission lines, of which only 8 (all constructed during winter months when the air in the laboratory was very dry) were usable.

- Finally, Bob fabricated several superconducting transmission lines of the Josephson type—by reducing the thickness of the oxide layer—and showed that they could support pulse-like solutions of varying speeds, in agreement with the properties of Equation (0.1). These were the first such systems ever constructed.

All of this work was clearly presented in 94 double spaced pages—to which I do not recall making a single editorial correction—leading me to suspect (only half in jest) that the worth of a thesis is inversely proportional to its weight.

But it would be incorrect to leave the impression that Bob occupied himself only with scientific matters, for his social conscience was keenly developed. As the folly of the Vietnam War unfolded throughout the 1960s and the city of Madison became polarized into flocks of "hawks" and "doves," he was in the vanguard of Americans working for an end to the killing and a peaceful resolution of the conflict. Although those were difficult years for the University of Wisconsin, the activities of concerned and committed students like Bob showed it to be a truly great educational institution.

Having completed his thesis in September of 1967, he spent the 1967–68 academic year as a postdoctoral assistant in the Electronics Department of Professor Georg Bruun at the Technical University of Denmark, where a group was then engaged in a substantial program of neuristor research. It was during this period that Bob took the opportunity to visit Prague and share the euphoria of that beautiful city in its short-lived release from foreign domination, an experience that left a strong impression, deepening his suspicion of the motivations behind many official actions.

In the fall of 1968 Bob was recruited by Wisconsin's Electrical Engineering Department as a tenure track assistant professor, a signal honor for the department then had a firm policy against hiring its own graduates in order to avoid "inbreeding." The reasons for this departure from standard procedure was that integrated circuit technology was becoming an important aspect of solid state electronics, and both Jim Nordman and I were fully occupied with our own research activities. As the most competent person we knew, Bob was brought on board and charged with developing an integrated circuits laboratory.

Not surprisingly, he was also caught up by the general feeling of student unrest that characterized those days, eagerly embracing novel approaches to teaching that would supersede the dull habits of the past. Following his lead, we presented some courses together on the relationships between modern technology and national politics that attracted both graduate and undergraduate students from a wide spectrum of university departments. One such class, I recall, met by an evening campfire in a wooded park on Madison's Lake Mendota, where we would sit in a circle discussing philosophy, science, technology, and politics as the twilight deepened. The *circle* is important. Under Bob's inspiration, we were all students—the highest status of an academic—striving together to understand.

So two salient characteristics of Bob's nature become evident: a sure-footed and independent approach to his professional work, and a deeply rooted concern for the spiritual health of his society. But there was more.

Bob had a way of quietly influencing events, of deftly intervening at the critical moment without worrying about taking credit for the results. From Denmark in the spring of 1968, he wrote that I should look at the papers of one E R Caianiello, who was doing interesting work on the theory of the brain, a vast subject toward which Bob's neuristor studies beckoned. Upon being contacted, Professor Caianiello responded that he would be pleased to deliver some reprints in person, as he was soon to be visiting in Chicago. Over a lunch by the lake, I vividly recall, he sketched plans for the *Laboratorio di Cibernetica*, a new sort of research institution that was then being launched in the village of Arco Felice, near Naples.

Following ideas that had been advanced a decade before by the American mathematician Norbert Wiener, the *Laboratorio* staff would comprise mathematicians, physicists, engineers, chemists, computer scientists, electrophysiologists, and neurobiologists—working in a collaborative effort to

understand the dynamic nature of a brain. As Wayne Johnson (who was just completing an experimental doctorate in superconductive devices) and I marveled at the scope of this scheme, Eduardo paused, looking thoughtfully at Wayne, and said: "I want you to come to Arco Felice and make Josephson junctions." In that moment, the Naples–Madison axis began.

Bob was the fourth Madisonian to trek to the *Laboratorio*, and the experience took hold of his psyche to an unanticipated degree. Encouraged by some subtle cultural chords, it seems, this Wisconsin boy felt immediately at home. There was something in the air of the *mezzogiorno* that resonated with deeper aspects of his spirit. Was it the haunting presence of Homer's "wine dark sea" or the glow of afternoon sunlight on Vesuvio's gorse? Or the exuberant dance of the olive trees in an autumn breeze, their silver underskirts flashing in the sun? Contributing perhaps to Bob's sense of belonging to Campania was the marvelous *cucina napoletana* and the fierce humor and independence of a people who have endured centuries of foreign domination. All of these reasons and more, I suspect, drew Bob into the bosom of Southern Italy.

Madison's loss was the gain of Naples as Bob carried his talent and experience in integrated circuit technology into this new environment, deftly wedding the new photo-lithographic fabrication techniques to emerging studies of nonlinear wave propagation on long Josephson junctions. Throughout the 1970s, theoretical, numerical and experimental research in the nonlinear science of Josephson transmission lines—described by physically motivated perturbations of the sine-Gordon equation—began to grow and prosper under the leadership of Bob and Antonio Barone and their students and colleagues, now far too many to list.

Although our personal and professional lives were entwined over more than three decades, Bob and I published very little together. One exception, of which I am particularly proud, was a paper that emerged from a famous soliton workshop that he organized in the summer of 1977 at the University of Salerno, to which he had moved a couple of years earlier. Held at the old quarters of the Physics Department in the middle of the city, this meeting attracted several stars of nonlinear science and provided unusual opportunities for real scientific and personal interactions. One formal talk in the morning was followed by lunch at a local restaurant that would have pleased Ernest Hemingway, lasting for a minimum of three hours and boasting unbounded conversation. Then in the late afternoon we would gather for another formal talk, after which smaller groups would carry on into the evening. It was from this inspired disorganization—perhaps only possible in the *mezzogiorno*—that it became clear how to solve Equation (0.1) with boundary conditions, making possible the analytic calculation of zero field steps in long (but finite) Josephson junctions.

In the mid-seventies, Bob's bent for subtly influencing events was exercised again. Having become friends with Niels Falsig Pedersen through meetings at international conferences, Bob encouraged the initiation of

studies on Josephson junction solitons among physicists and applied mathematicians at DTU, anticipating the advantages that could be gained from a collaboration between those near the top and bottom (geographically speaking) of Europe. During the 1980s, as is evident from several chapters of this book, such research came of age. In the best traditions of nonlinear science, a remarkable *ménage à trois* of experimental, theoretical and numerical work emerged, relating the deep insights of soliton theory to a growing spectrum of experimental observations on long Josephson junctions. Reflecting the earlier Hodgkin-Huxley work on nerves, this international effort serves as a paradigm of how nonlinear science should be conducted.

Throughout these developments, Bob's steady influence was ever present, leading the group mind away from the abrasive competition that is all too common in many areas of modern science. Much of the civilized tone characterizing current investigations of superconductive devices stems from Bob's guiding hand.

Looking wistfully back over these fleeting years, I see a paradox in Bob's nature. Although ever tolerant of human foibles, sensitive to cultural imperatives, and ready to seek an intelligent compromise among conflicting personalities, he remained wary to the end of petty bureaucrats and mean spirited power games. Indeed, the last email messages we exchanged in December of 1996 were about codes for protecting internet users against prying officials of government.

Heading into the twenty-first century, practitioners of nonlinear science will miss Bob's wise and gentle counsel. While discussing his tragic death, Antonio Barone mentioned that in such cases, one often remarks that the departed person was a "good guy."

"But, you know," said Antonio, "Bob, he really *was* a good guy."

PREFACE

The world of science has seen many successes over the past century, but none has been more striking than the recent flowering of nonlinear research. Largely ignored in the realms of physics until some three decades ago, studies of the *emergence of coherent structures* from the underlying nonlinear dynamics is now a vital facet of applied and theoretical science, providing ample evidence—for those who still need it—that

The whole is more than the sum of its parts.

In this book, twenty-eight distinguished contributors describe these developments from the perspectives of their individual interests, paying particular attention to those aspects that seem to be of importance for the the coming century. Although the chapters included here comprise but a small portion of the current activities, we expect the readers to be impressed by its diversity and challenge.

The story opens with two fundamental chapters, underlying all of the others. The first of these presents a general description of *coherent phenomena* in a variety of experimental settings, including plasma physics, fluid dynamics and nonlinear optics. The second is a review of developments in *perturbation theory* that have been profoundly influenced by research in nonlinear science since the mid 1960s.

The next four chapters describe various studies of *Josephson junction superconductive devices,* which have both stimulated and been encouraged by corresponding developments in nonlinear science. Undreamed of 40 years ago, these devices have increased the sensitivities of magnetometers and voltmeters by several orders of magnitude, and they promise corresponding advances in submillimeter wave oscillators and in the speed of digital computations. Not unrelated to recent progress in the development of superconductors with higher operating temperatures are the quasi-two-dimensional magnets that support *vortex structures* as described in Chapter 7. This is an exciting field of theoretical study that stems directly from recent advances in condensed matter physics.

Without doubt, the most important technical application of the ubiquitous and hardy soliton is as a carrier of digital information along optical fibers. In recognition of this, five chapters are included on various aspects

of modern optical research, ranging from general studies of basic properties to more detailed considerations of current design objectives. We believe these chapters will provide the reader with an unusually clear exposure to both the theoretical and the practical implications of *optical solitons* for the coming century.

Another significant branch of present day nonlinear science is that of *nonlinear lattices.* Going back to the early 1980s, this work is introducing the revolutionary concept of *local modes* into the study of molecular crystals. Of the six chapters in this area, the first deals with *dislocation dynamics* in crystals, and the second suggests the key role that two-dimensional *breathers* may have played in the formation of crystal structures, such as muscovite mica. Other chapters deal with novel phenomena arising from more than one length scale, mechanical models for lattice solitons, and the quantum theory of solitons in real lattices. From such work, we believe, may emerge basic elements for coherent information processing in the terahertz (far infra-red) region of the electromagnetic spectrum. The final chapter in this nonlinear lattice segment of the book describes ways in which "colored" thermal noise can give rise to *molecular motors* at the scale of nanometers. This idea has important implications for transport mechanisms that may operate within living cells, setting the stage for the final four chapters which address the *nonlinear science of life.*

Just as the past 100 years have been called the "century of physics," we expect that the next will be recognized as the "century of biology." Since almost every aspect of biology is nonlinear, this is the area in which we see the new ideas having their greatest impact. Thus the last four chapters are devoted to physical aspects of biological research.

The first of these describes various attempts to understand the dynamics of DNA in the context of modern biophysics. This survey provides the reader with a hierarchy of mathematical models, each gaining in accuracy as the computational difficulties correspondingly increase. Related to this chapter is the following one, describing exact numerical solutions for the dynamics of certain helical biomolecules that are components of natural protein.

The penultimate chapter—on exploratory investigations of the nonlinear dynamics of bacterial populations—is intended to draw physical scientists into the study of *life.* Similarly, the final chapter attempts to encourage young experimentalists and theorists to consider the most intricate dynamical system in the known universe: the *human brain.*

It is our hope that the readers of this book will make a significant contribution to research activities in the "century of biology."

Lyngby and Tucson THE EDITORS
1999

Contents

I Nonlinear Science **1**

1 Nonlinear Coherent Phenomena in Continuous Media **3**
E.A. Kuznetsov and V.E. Zakharov
1 Introduction 3
2 Phase randomization in nonlinear media 4
3 Nonlinear Schrödinger equation 9
4 Solitons in the focusing NSLE 11
5 Collapses in the NLSE 15
6 Weak, strong and superstrong collapses 18
7 Anisotropic black holes 21
8 Structure in media with weak dispersion 26
9 Singularities on a fluid surface 32
10 Solitons and collapses in the generalized KP equation . . . 34
11 Self-focusing in the boundary layer 38
12 References 42

2 Perturbation Theories for Nonlinear Waves **47**
L. Ostrovsky and K. Gorshkov
1 Introduction 47
2 Modulated waves 49
3 Direct perturbation method 51
4 Perturbation theories for solitary waves 52
4.1 Direct perturbation method for solitons: quasistationary approach 52
4.2 Nonstationary approach 54
4.3 Inverse-scattering perturbation scheme 55
4.4 "Equivalence principle" 57
4.5 Example: soliton interaction in Lagrangian systems . 58
4.6 Radiation from solitons 59
5 Asymptotic reduction of nonlinear wave equations 60
6 Conclusions 61
7 References 62

II Superconductivity and Magnetism 67

3 Josephson Devices 69
A. Barone and S. Pagano
1 Introduction . . . 69
2 Elements of the Josephson effect . . . 70
3 SQUIDs . . . 72
4 Digital devices . . . 75
5 Detectors . . . 77
6 Voltage standard . . . 78
7 Microwave oscillators . . . 80
8 Conclusions . . . 83
9 References . . . 83

4 Josephson Flux-Flow Oscillators in Microwave Fields 87
M. Salerno and M. Samuelsen
1 Introduction . . . 87
2 Flux-flow oscillators in uniform microwave fields . . . 88
3 Flux-flow oscillators in non-uniform microwave fields . . . 91
4 Numerical experiment . . . 94
5 Conclusions . . . 98
6 Appendix . . . 98
7 References . . . 100

5 Coupled Structures of Long Josephson Junctions 103
G. Carapella and G. Costabile
1 Stacks of two long Josephson junctions . . . 103
1.1 The physical system and its model . . . 103
1.2 Experiments on stacks of two long Josephson junctions . . . 106
2 Parallel arrays of Josephson junctions . . . 109
2.1 The physical system and its model . . . 109
2.2 Numerical and experimental results on five-junctions parallel arrays . . . 112
3 Triangular cells of long Josephson junctions . . . 113
3.1 The model . . . 114
3.2 Numerical and experimental results . . . 115
4 References . . . 118

6 Stacked Josephson Junctions 121
N.F. Pedersen
1 Introduction . . . 121
2 Short summary of fluxon properties . . . 121
3 Stacked junctions . . . 123
4 Fluxon solutions, selected examples . . . 125

4.1 The coherent 2-fluxon mode . . . 125
4.2 The two modes of the two fluxon case . . . 127
5 Stacked junction plasma oscillation solutions . . . 128
6 Conclusion . . . 135
7 References . . . 135

7 Dynamics of Vortices in Two-Dimensional Magnets 137
F.G. Mertens and A.R. Bishop
1 Introduction . . . 137
2 Collective variable theories at zero temperature . . . 141
2.1 Thiele equation . . . 141
2.2 Vortex mass . . . 143
2.3 Hierarchy of equations of motion . . . 146
2.4 An alternative approach: coupling to magnons . . . 152
3 Effects of thermal noise on vortex dynamics . . . 155
3.1 Equilibrium and non-equilibrium situations . . . 155
3.2 Collective variable theory and Langevin dynamics simulations . . . 155
3.3 Noise-induced transitions between opposite polarizations . . . 161
4 Dynamics above the Kosterlitz-Thouless transition . . . 163
4.1 Vortex-gas approach . . . 163
4.2 Comparison with simulations and experiments . . . 164
4.3 Vortex motion in Monte Carlo simulations . . . 165
5 Conclusion . . . 166
6 References . . . 167

III Nonlinear Optics 171

8 Spatial Optical Solitons 173
Yu.S. Kivshar
1 Introduction . . . 173
2 Spatial vs. temporal solitons . . . 175
3 Basic equations . . . 176
4 Stability of solitary waves . . . 178
4.1 One-parameter solitary waves . . . 179
4.2 Two-parameter solitary waves . . . 180
5 Experiments on self-focusing . . . 182
6 Soliton spiralling . . . 184
7 Multi-hump solitons and solitonic gluons . . . 186
8 Discrete spatial optical solitons . . . 189
9 References . . . 191

9 Nonlinear Fiber Optics **195**
G.P. Agrawal
1 Introduction . . . 195
2 Fiber characteristics . . . 196
2.1 Single-mode fibers . . . 196
2.2 Fiber nonlinearities . . . 196
2.3 Group-velocity dispersion . . . 197
3 Pulse propagation in fibers . . . 198
3.1 Nonlinear Schrödinger equation . . . 198
3.2 Modulation instability . . . 198
4 Optical solitons . . . 199
4.1 Bright solitons . . . 200
4.2 Dark solitons . . . 201
4.3 Loss-managed solitons . . . 202
4.4 Dispersion-managed solitons . . . 203
5 Nonlinear optical switching . . . 205
5.1 SPM-based optical switching . . . 205
5.2 XPM-based optical switching . . . 207
6 Concluding remarks . . . 209
7 References . . . 209

10 Self-Focusing and Collapse of Light Beams in Nonlinear Dispersive Media **213**
L. Bergé and J. Juul Rasmussen
1 Introduction . . . 213
2 General properties of self-focusing with anomalous group velocity dispersion . . . 214
2.1 Basic properties . . . 214
2.2 Self-similar wave collapses . . . 216
3 Self-focusing with normal group velocity dispersion . . . 220
4 Discussion of the general properties, outlook . . . 224
5 References . . . 226

11 Coherent Structures in Dissipative Nonlinear Optical Systems **229**
J.V. Moloney
1 Introduction . . . 229
2 Nonlinear waveguide channeling in air . . . 234
2.1 Dynamic spatial replenishment of femtosecond pulses propagating in air . . . 236
3 Control of optical turbulence in semiconductor lasers . . . 239
3.1 The control . . . 240
4 Summary and conclusions . . . 243
5 References . . . 245

12 Solitons in Optical Media with Quadratic Nonlinearity 247
B.A. Malomed
1 Introduction . . . 247
2 The basic theoretical models . . . 249
3 The solitons . . . 253
4 Conclusion . . . 260
5 References . . . 261

IV Lattice Dynamics and Applications 263

13 Nonlinear Models for the Dynamics of Topological Defects in Solids 265
Yu.S. Kivshar, H. Benner and O.M. Braun
1 Introduction . . . 265
2 The FK model and the SG equation . . . 266
3 Physical models . . . 269
3.1 Dislocations in solids . . . 269
3.2 Magnetic chains . . . 270
3.3 Josephson junctions . . . 271
3.4 Hydrogen-bonded chains . . . 273
3.5 Surface physics and adsorbed atomic layers . . . 275
3.6 Models of the DNA dynamics . . . 276
4 Properties of kinks . . . 277
4.1 On-site potential of general form . . . 277
4.2 Discreteness effects . . . 279
4.3 Kinks in external fields . . . 280
4.4 Compacton kinks . . . 281
5 Experimental verifications . . . 281
5.1 Josephson junctions . . . 281
5.2 Magnetic systems . . . 283
6 Concluding remarks . . . 285
7 References . . . 286

14 2-D Breathers and Applications 293
J.L. Marín, J.C. Eilbeck and F.M. Russell
1 Introduction . . . 293
2 Deciphering the lines in mica . . . 294
3 Numerical and analogue studies . . . 296
4 Longitudinal moving breathers in 2D lattices . . . 299
5 Breather collisions . . . 302
6 Conclusions and further applications . . . 303
6.1 Application to sputtering . . . 303
6.2 Application to layered HTSC materials . . . 303
7 References . . . 304

15 Scale Competition in Nonlinear Schrödinger Models **307**
Yu. B. Gaididei, P.L. Christiansen and S.F. Mingaleev
1 Introduction 307
2 Excitations in nonlinear Kronig-Penney models 308
3 Discrete NLS models with long-range dispersive interactions 311
4 Stabilization of nonlinear excitations by disorder 316
5 Summary 319
6 References 320

16 Demonstration Systems for Kink-Solitons **323**
M. Remoissenet
1 Introduction 323
2 Mechanical chains with double-well potential 325
2.1 Chain with torsion and gravity 325
2.2 Chain with flexion and gravity 329
2.3 Numerical simulations 330
3 Experiments 331
3.1 Chain with torsion and gravity 331
3.2 Chain with flexion and gravity 332
4 Lattice effects 333
5 Conclusion 334
6 Appendix 335
7 References 336

17 Quantum Lattice Solitons **339**
A.C. Scott
1 Introduction 339
2 Local modes in the dihalomethanes 339
2.1 Classical analysis 340
2.2 Quantum analysis 341
2.3 Comparison with experiments 344
3 A lattice nonlinear Schrödinger equation 344
4 Local modes in crystalline acetanilide 348
5 Conclusions 354
6 References 355

18 Noise in Molecular Systems **357**
G.P. Tsironis
1 Introduction 357
2 Additive correlated ratchets 358
3 Current reversal 363
4 Synthetic motor protein motion 364
5 Targeted energy transfer and nonequilibrium fluctuations in bioenergetics 368
6 References 369

V Biomolecular Dynamics and Biology 371

19 Nonlinear Dynamics of DNA 373
L.V. Yakushevich
1 Introduction . . . 373
2 General description of DNA dynamics. Classification of the internal motions . . . 375
3 Mathematical modeling of DNA dynamics. Hierarchy of the models . . . 376
3.1 Principles of modeling . . . 376
3.2 Structural hierarchy . . . 376
3.3 Dynamical hierarchy . . . 377
4 Nonlinear mathematical models. Solved and unsolved problems . . . 379
4.1 Ideal models . . . 379
4.2 Nonideal models . . . 380
4.3 Statistics of solitons in DNA . . . 380
5 Nonlinear DNA models and experiment . . . 381
5.1 Hydrogen-tritium (or hydrogen-deuterium) exchange 381
5.2 Resonant microwave absorption . . . 382
5.3 Scattering of neutrons and light . . . 383
5.4 Fluorescence depolarization . . . 384
6 Nonlinear conception and mechanisms of DNA functioning . 385
6.1 Nonlinear mechanism of conformational transitions . 385
6.2 Nonlinear conformational waves and long-range effects 385
6.3 Direction of transcription process . . . 386
7 References . . . 387

20 From the FPU Chain to Biomolecular Dynamics 393
A.V. Zolotaryuk, A.V. Savin and P.L. Christiansen
1 Introduction . . . 393
2 Helices in two and three dimensions . . . 394
3 Equations of motion for a helix backbone . . . 397
4 Small-amplitude limit . . . 398
5 Three-component soliton solutions . . . 400
5.1 3D case: solitons of longitudinal compression . . . 402
5.2 2D case: other types of solutions . . . 404
6 Conclusions . . . 405
7 References . . . 407

21 Mutual Dynamics of Swimming Microorganisms and Their Fluid Habitat 409
J.O. Kessler, G.D. Burnett and K.E. Remick
1 Introduction . . . 409
2 Bioconvection (I) . . . 411

2.1 Observations 411
2.2 Continuum theory 412
3 Bacteria in constraining environments (II) 415
4 Possibilities for computer simulation 417
5 Conclusion 421
6 Appendix I: Statistical methods 423
7 Appendix II 424
8 References 425

22 Nonlinearities in Biology: The Brain as an Example 427
H. Haken
1 Introduction 427
2 Some salient features of neurons 427
3 The noisy lighthouse model of a neural network 429
4 The special case of two neurons 430
5 Time-averages 433
6 The averaged neural equations 434
7 How to make contact with experimental data? Synergetics as a guide 440
8 Concluding remarks and outlook 443
9 References 444

Index 447

List of Authors

Agrawal, Govind P.
The Institute of Optics
University of Rochester
PO Box 270186
Rochester, NY 14627
USA

Barone, Antonio
Istituto di Cibernetica
Via Toiano 6
I-80072 Arco Felice
Napoli
Italy

Benner, Hartmut
Institut für Festkörperphysik
and SFB 185
Technische Hochschule Darmstadt
Hochschulstrasse 6
D-64289 Darmstadt
Germany

Bergé, Luc
Commissariat à l'Energie Atomique
CEA-DAM/Île-de-France
B.P. 12
F-91680 Bruyères-le-Châtel
France

Bishop, Alan R.
Theoretical Division and CNLS
Los Alamos National Laboratory
MS B262
Los Alamos, NM 87545
U. S. A.

Braun, Oleg M.
Institute of Physics
Ukrainian Academy of Sciences
46 Science Avenue
UA-252022 Kiev
Ukraine

Burnett, G. David
Physics Department, Building 81
University of Arizona
Tucson, AZ 85721
USA

Carapella, G.
Dipartimento de Fisica
Università di Salerno
I-84081 Baronissi
Italy

Christiansen, Peter L.
Department of Mathematical Modelling
Technical University of Denmark
DK-2800 Lyngby
Denmark

Costabile, Giovanni
Dipartimento de Fisica
Università di Salerno
I-84081 Baronissi
Italy

Eilbeck, J. Chris
Department of Mathematics
Heriot-Watt University
Riccarton
Edinburgh EH15 4AS
United Kingdom

Gaididei, Yuri B.
Institute for Theoretical Physics
Academy of Sciences of Ukraine
Metrologicheskaya Street 14-B
UA-252143 Kiev
Ukraine

Gorshkov, Konstantin A.
Institute of Applied Physics
Russian Acadamy of Sciences
46 Ulyanova Street
R-603006 Nizhny Novgorod
Russia

Haken, Herman
Institute for Theoretical Physics 1
Center of Synergetics
University of Stuttgart
Pfaffenwaldring 57/IV
D-70550 Stuttgart
Germany

Kessler, John O.
Physics Department, Building 81
University of Arizona
Tucson, AZ 85721
USA

Kivshar, Yuri S.
Optical Sciences Center
The Australian National University
Canberra ACT 0200
Australia

Kuznetsov, E. A.
L.D. Landau Institute
for Theoretical Physics
2 Kosygina Street
R-117334 Moscow
Russia

Malomed, Boris A.
Department of
Interdisciplinary Studies
Faculty of Engineering
Tel Aviv University
Tel Aviv 69978
Israel

Marín, J. L.
Department of Mathematics
Heriot-Watt University
Riccarton
Edinburgh EH15 4AS
United Kingdom

Mertens, Franz G.
Physikalisches Institut
Lehrstuhl für Theoretische Physik I
Universität Bayreuth
Postfach 101251, D-8580 Bayreuth
Germany

Mingaleev, Sergei F.
Institute for Theoretical Physics
Academy of Sciences of Ukraine
Metrologicheskaya Street 14-B
UA-252143 Kiev
Ukraine

Moloney, Jerry V.
Department of Mathematics
and Optical Science
University of Arizona
Tucson, Arizona 85721
USA

Ostrovsky, Lev
University of Colorado
Cooperative Institute for Research
in Environmental Sciences/NOAA
Environmental Technology Laboratory
325 Broadway
Boulder, Colorado 80303
USA

and
Institute of Applied Physics
Russian Acadamy of Sciences
46 Ulyanova Street
R-603006 Nizhny Novgorod
Russia

Pagano, Sergio
Istituto de Cibernetica del CNR
Via Toiano 6
Arco Felice, NA Napoli
Italy

Pedersen, Niels Falsig
Department of Electric
Power Engineering
Technical University of Denmark
DK-2800 Lyngby
Denmark

Rasmussen, Jens Juul
Risø National Laboratory
Optics and Fluid Dynamics Department
P.O. Box 49
DK-4000 Roskilde
Denmark

Remick, Katherine E.
Neuroscience
University of Texas Medical Branch
Galveston, TX 77555
USA

Remoissenet, Michel
Faculté des Sciences et Techniques
Université de Bourgogne
9 Av. A. Savary BP400
F-21011 Dijon
France

Russell, F. M.
Department of Mathematics
Heriot-Watt University
Riccarton
Edinburgh EH15 4AS
United Kingdom

Salerno, Mario
Dipartimento di Fisica Teorica
Università di Salerno
Via S. Allende
I-84081 Baronissi
Italy

Samuelsen, Mogens
Department of Physics
Technical University of Denmark
DK-2800 Lyngby
Denmark

Savin, Alexander V.
Institute for Physico-Technical Problems
Prechistenka 13/7
R-119034 Moscow
Russia

and
Department of Mathematical Modelling
Technical University of Denmark
DK-2800 Lyngby
Denmark

Scott, Alwyn C.
Department of Mathematics
University of Arizona
Tucson, AZ 85721
USA

Tsironis, George P.
Research Center of Crete
and Physics Department
University of Crete
P.O. Box 2208
G-71003 Iraklion
Greece

Yakushevich, Ludmilla V.
Institute of Cell Biophysics
Academy of Sciences
R-142292 Pushchino
Russia

Zakharov, V. E.
L.D. Landau Institute for
Theoretical Physics
2 Kosygina Street
R-117344 Moscow
Russia

Zolotaryuk, Alexander V.
Bogolyubov Institute for
Theoretical Physics
Ukranian Academy of Sciences
Metrologicheskaya Street 14-B
UA-252143 Kiev
Ukraine

1

Nonlinear Coherent Phenomena in Continuous Media

E.A. Kuznetsov
V.E. Zakharov

ABSTRACT This review is devoted to description of coherent nonlinear phenomena in almost conservative media with applications to plasma physics, fluid dynamics and nonlinear optics. The main attention in the review is paid to consideration of solitons, collapses, and black holes. The latter is a quasi-stationary singular object which appear after the formation of a singularity in nonlinear wave systems. We discuss in details the qualitative reasons of the wave collapse and a difference between solitons and collapses, and apply to their analysis exact methods based on the integral estimates and the Hamiltonian formalism. These approaches are demonstrated mainly on the basic nonlinear models, i.e. on the nonlinear Schrödinger equation and the Kadomtsev-Petviashvili equation and their generalizations.

1 Introduction

All real continuous media, including vacuum, are nonlinear. Nonlinearity might be a cause of quite opposite physical effects. One of them is phase randomization leading to formation of a chaotic state - weak or strong wave turbulence. Wind-driven waves on the ocean surface is the classical example of that sort. Another group of effects is spontaneous generation of coherent structures. These structures may be localized in space or both in space and in time. Phases of Fourier harmonics, forming the structures, are strongly correlated.

Very often coherent structures coexist with wave turbulence. A simple example of the coherent structure is 'white caps' on the crest of gravity wave of high amplitude. Elementary visual observation shows that just before breaking, a wave crest takes the universal, wedge-type shape. Apparently, the harmonics composing this shape have correlated phases. The wave breaking is an important mechanism of energy and momentum dissipation on the ocean. A satisfactory theory of this basic effect is not yet developed.

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 3−45, 2000.

A more standard example of a coherent structure is a solitary wave on the surface of shallow water. These examples present two major types of coherent structures - collapses and solitons. Solitons are stationary, spatially localized wave packets, which are very common in nonlinear media. Collapses are almost as wide-spread phenomena as solitons. These are catastrophic processes of concentration of wave energy in localized space domains leading to absorption of at least part of this energy. Collapses are an important mechanism of the wave energy dissipation in almost conservative media, in particular, they play essential roles for many methods of fusion plasma heating.

Collapses and solitons are not all the coherent structures that can be found in nonlinear media. Rich families of coherent structures exist in active media, providing the balance between pumping and dissipation. Among them there are patterns described by the Ginsburg-Landau type equations and spiral waves in reaction-diffusion systems. Rolls and hexagons in the Benard convection are such examples. But even in almost conservative media one can find coherent structures different from solitons and collapses. One can mention, for instance, "black holes", which are persistent localized regions of the wave energy dissipation arising in some cases after the act of wave collapse resulting in the formation of a singularity.

In this paper we shall discuss coherent structures in almost conservative media only. We concentrate our attention mostly on collapses and solitons, which are, in our opinion, closely related phenomena. In many important physical situations, collapse is a result of the soliton instability (for more details, see two reviews [1, 2] and references therein). We shall briefly discuss also the theory of black holes in the models describing by the nonlinear Schrödinger equation (NLSE). Using the Hamiltonian formalism gives us an opportunity to study the problem of coherent structures in its maximum generality (see also our recent review [3] devoted to this subject). Physical examples used in the paper are taken mostly from hydrodynamics, nonlinear optics, and plasma physics.

2 Phase randomization in nonlinear media

Let as consider wave propagation in a uniform boundless conservative medium. The wave field will be described by the complex normal variable $a_k(t)$, satisfying the equation of motion

$$\frac{\partial a_k}{\partial t} = -i\frac{\delta H}{\delta a_k^*}\,. \tag{1.1}$$

In the linear approximation

$$H = H_0 = \int \omega(k)|a_k|^2 dk, \tag{1.2}$$

where $\omega(k)$ is the dispersion law. In this case equation (1.1) is trivially integrated

$$\frac{\partial a_k(t)}{\partial t} + i\omega(k)a_k(t) = 0, \tag{1.3}$$

$$a_k(t) = C(k)e^{-i\omega(k)t}. \tag{1.4}$$

For a localized wave packet one should require $\int |C(k)|^2 dk < \infty$. For the system of monochromatic waves the distribution of a_k is a set of δ functions

$$C(k) = \sum C_n \delta(k - k_n). \tag{1.5}$$

In the linear approximation phases of each waves $\arg a_k = \phi_k$ grow linearly in time

$$\phi_k(t) = \phi_k(0) + \omega(k)t$$

and, respectively, the trajectory of the system winds on the infinitely-dimensional torus. The phase ϕ_k is defined modulo 2π. Therefore for two waves with incommensurable frequencies $\omega(k)$ and $\omega(k_1)$ difference (or sum) in phases $\phi_k(t) \mp \phi_{k_1}(t) = \phi_k\ (0) \mp \phi_{k_1}(0) + (\omega(k) \mp \omega(k_1))t$ with time becomes random function on the interval 2π. Thus, for continuous dependence $\omega = \omega(k)$ (except $\omega(k)$ =const), the linear dispersion leads to complete phase randomness for the wave distribution.

Now let us introduce into (1.1) a quadratic nonlinearity. It is enough to replace

$$H \to H_0 + H_1, \tag{1.6}$$

$$H_1 = -\frac{1}{2}\int V_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2}\left(a_{\boldsymbol{k}}{}^* a_{\boldsymbol{k}_1} a_{\boldsymbol{k}_2} + a_{\boldsymbol{k}} a_{\boldsymbol{k}_1}^* a_{\boldsymbol{k}_2}^*\right)\delta_{\boldsymbol{k}-\boldsymbol{k}_1-\boldsymbol{k}_2} d\boldsymbol{k} d\boldsymbol{k}_1 d\boldsymbol{k}_2\ . \tag{1.7}$$

Here $V_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2}$ are coupling coefficients for three-wave interaction. The equation of motion (1.1) takes now the form

$$\frac{\partial a_{\boldsymbol{k}}}{\partial t} + i\omega(\boldsymbol{k})a_{\boldsymbol{k}} + \frac{i}{2}\int \big\{V_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2} a_{\boldsymbol{k}_1} a_{\boldsymbol{k}_2}\delta_{\boldsymbol{k}-\boldsymbol{k}_1-\boldsymbol{k}_2} + 2V_{\boldsymbol{k}_1\boldsymbol{k}\boldsymbol{k}_2} a_{\boldsymbol{k}_1} a_{\boldsymbol{k}_2}^*\ \delta_{\boldsymbol{k}-\boldsymbol{k}_1+\boldsymbol{k}_2}\big\} d\boldsymbol{k}_1 d\boldsymbol{k}_2 = 0. \tag{1.8}$$

The Equation (1.8) describes several nonlinear effects. Suppose that the equations

$$\boldsymbol{k}_1 = \boldsymbol{k}_2 + \boldsymbol{k}_3, \qquad \omega(\boldsymbol{k}_1) = \omega(\boldsymbol{k}_2) + \omega(\boldsymbol{k}_3) \tag{1.9}$$

have nontrivial real solutions, as for instance, if $\omega(0) = 0$ and $\omega'' > 0$. Suppose further that at $t = 0$

$$a_{\boldsymbol{k}} = C_1^{(0)}\delta(\boldsymbol{k} - \boldsymbol{k}_1) + C_2^{(0)}\delta(\boldsymbol{k} - \boldsymbol{k}_2) + C_3^{(0)}\delta(\boldsymbol{k} - \boldsymbol{k}_3), \tag{1.10}$$

where $\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3$ satisfy the equations (1.9). Then at $t > 0$, in the limit of small enough intensities of waves, the complex amplitude $a_{\boldsymbol{k}}(t)$ can be sought in the form

$$a_{\boldsymbol{k}} = \sum_{i=1}^{3} C_i(t) e^{-i\omega(\boldsymbol{k}_i)t} \delta(\boldsymbol{k} - \boldsymbol{k}_i) \tag{1.11}$$

where $C_i(t)$ obey the system of ordinary differential equations, the so-called three-waves system [4]

$$\frac{\partial C_1}{\partial t} = iVC_2C_3, \quad \frac{\partial C_2}{\partial t} = iVC_1C_3^* \quad \frac{\partial C_3}{\partial t} = iVC_1C_2^*. \tag{1.12}$$

Here the coupling coefficient $V = V_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2}$.

Equation (1.12) can be easily solved in elliptic functions. The initial data

$$C_1 = 0, \qquad C_2 = C_2^{(0)}, \qquad C_3 = C_3^{(0)}$$

separate the solution describing growth of C_1. In particular, at small time

$$C_1 \simeq iVC_2^{(0)}C_3^{(0)}t.$$

This is the simplest nonlinear process - resonant "mixing" of two monochromatic waves.

The equations (1.12) describe also another very important nonlinear process, namely, the decay instability of the monochromatic waves. Let at $t = 0$

$$C_1 = Ae^{i\phi}, \qquad C_2 = q, \qquad C_3 = iq^*e^{i\phi}, \qquad |q| \ll A. \tag{1.13}$$

Now for small times

$$C_2 \simeq qe^{\gamma t} \tag{1.14}$$

where $\gamma = |v||A|$ is the growth rate of the so-called decay instability. This solution describes exponential growth of the waves C_2, C_3. Their phases ($C_2 = |C_2|e^{i\phi_2},\ C_3 = |C_3|e^{i\varphi_3}$) satisfy the condition

$$\phi_2 + \phi_3 = \phi + \pi/2. \tag{1.15}$$

Thus, the sum of phases ϕ_2 and ϕ_3 is fixed. But a phase of one of the waves in this pair (phase of q) is quite arbitrary. We found that in the most idealized case (when due to the decay instability only one pair of monochromatic waves is excited) this process yields the correlation for sum of phases of the excited waves and simultaneously introduces to the wave system an element of randomness, namely, the phase of q. In more realistic case the instability excites a whole ensemble of wave pairs satisfying the conditions (1.9) up to the accuracy of γ. Each exited pair adds one random phase. The exited waves are also unstable. Multiplication of the process of instability has to create in the system a lot of new waves with random

phases and to cause finally complete turbulization of the wave field. We must stress that this scenario is just a very plausible conjecture. It would be very important to check it by a direct numerical experiment. The point of common belief is the following. As a result of multiple events of the wave mixing and decay instability, after some time phases become completely random. In this case the wave field can be described statistically by the correlation function

$$\langle a_{\boldsymbol{k}} a^*{}_{\boldsymbol{k}'} \rangle = n_{\boldsymbol{k}} \delta_{\boldsymbol{k}\boldsymbol{k}'}. \tag{1.16}$$

Here $n_{\boldsymbol{k}}$ is the quasi-particle density (or the wave action). This quantity for sufficiently small wave intensity satisfies the kinetic equation (for details see [5])

$$\frac{\partial n_{\boldsymbol{k}}}{\partial t} = St(n,n), \tag{1.17}$$

$$\begin{aligned} St(n,n) &= \int \left\{ A_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2} - A_{\boldsymbol{k}_1\boldsymbol{k}\boldsymbol{k}_2} - A_{\boldsymbol{k}_2\boldsymbol{k}\boldsymbol{k}_1} \right\} d\boldsymbol{k}_1 d\boldsymbol{k}_2, \\ A_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2} &= 4\pi |V_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2}|^2 \left(n_{\boldsymbol{k}_1} n_{\boldsymbol{k}_2} - n_{\boldsymbol{k}} n_{\boldsymbol{k}_1} - n_{\boldsymbol{k}} n_{\boldsymbol{k}_2} \right) \cdot \\ &\quad \delta_{\boldsymbol{k}-\boldsymbol{k}_1-\boldsymbol{k}_2} \delta_{\omega(\boldsymbol{k})-\omega(\boldsymbol{k}_1)-\omega(\boldsymbol{k}_2)}. \end{aligned}$$

The kinetic equation accounts for the correlation in wave phases (1.15) in the first order with respect to the matrix element $V_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2}$ that, in particular, provides a nonzero three wave correlation function $\langle a_{\boldsymbol{k}} a^*{}_{\boldsymbol{k}_1} a^*{}_{\boldsymbol{k}_2} \rangle = J_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2} \delta_{\boldsymbol{k}-\boldsymbol{k}_1-\boldsymbol{k}_2}$.

The state of the wave field described by the kinetic equation (1.17) is called weak turbulence. Direct numerical examination of the theory of weak turbulence is one of the most interesting problem in computational physics at the time.

It might happen that equations (1.9) have no real solutions. In this case the first interacting term in the Hamiltonian has to be taken in the form

$$H_1 = \frac{1}{2} \int T_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{k}_3} a_{\boldsymbol{k}}{}^* a^*_{\boldsymbol{k}_1} a_{\boldsymbol{k}_2} a_{\boldsymbol{k}_3} \delta_{\boldsymbol{k}+\boldsymbol{k}_1-\boldsymbol{k}_2-\boldsymbol{k}_3} d\boldsymbol{k} d\boldsymbol{k}_1 d\boldsymbol{k}_2 d\boldsymbol{k}_3. \tag{1.18}$$

Equation (1.8) transforms now into the form

$$\begin{aligned} &\frac{\partial a_{\boldsymbol{k}}}{\partial t} + i\omega(\boldsymbol{k}) a_{\boldsymbol{k}} = \\ &-i \int T_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{k}_3} a^*_{\boldsymbol{k}_1} a_{\boldsymbol{k}_2} a_{\boldsymbol{k}_3} \delta_{\boldsymbol{k}+\boldsymbol{k}_1-\boldsymbol{k}_2-\boldsymbol{k}_3} d\boldsymbol{k}_1 d\boldsymbol{k}_2 d\boldsymbol{k}_3. \end{aligned} \tag{1.19}$$

Equation (1.8) has the natural constants of motion, i.e., the Hamiltonian and the momentum

$$H = H_0 + H_1 \quad \text{and} \quad \boldsymbol{P} = \int \boldsymbol{k} a_{\boldsymbol{k}} a_{\boldsymbol{k}}{}^* d\boldsymbol{k}. \tag{1.20}$$

The Equation (1.19) has the additional invariant $N = \int a_{\boldsymbol{k}} a_{\boldsymbol{k}}{}^* d\boldsymbol{k}$. If $H_1 \ll H_0$ the equation (1.19) can be considered as linear and again has a solution (1.4), (1.5), but elementary process of the nonlinear wave interaction are now different.

If wave vectors $\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{k}_3$ of three monochromatic waves satisfy the condition

$$-\omega(\boldsymbol{k}_1) + \omega(\boldsymbol{k}_2) + \omega(\boldsymbol{k}_3) = \omega(-\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3), \tag{1.21}$$

they pump a new wave with the wave vector

$$\boldsymbol{k} = -\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3. \tag{1.22}$$

This is a "resonant mixing" of wave triads. Another type of nonlinear interaction is an instability of monochromatic waves. As in the previous case they lead to excitation of wave pairs. In the case of instability of an individual wave with the wave vector $\boldsymbol{k}_0$ there excites a pair with the wave vectors $\boldsymbol{k}_2, \boldsymbol{k}_3$, satisfying the conditions

$$\boldsymbol{k}_2 + \boldsymbol{k}_3 = 2\boldsymbol{k}_0, \qquad \omega(\boldsymbol{k}_2) + \omega(\boldsymbol{k}_3) = 2\omega(\boldsymbol{k}_0). \tag{1.23}$$

Phases of new waves ϕ_2, ϕ_3 are connected with the phase of the initial wave ϕ_0 by the relation

$$\phi_2 + \phi_3 = 2\phi_0 + \pi/2. \tag{1.24}$$

Their difference $\phi_2 - \phi_3$ is again arbitrary. Hence this instability introduces an element of chaos to the system.

Another instability taking place in the system (1.19) is instability of wave pairs. If initially the wave field consists of two monochromatic waves with wave vectors $\boldsymbol{k}_0, \boldsymbol{k}_1$, two other waves grow exponentially, if their wave vectors $\boldsymbol{k}_2, \boldsymbol{k}_3$ satisfy the resonant conditions

$$\boldsymbol{k}_2 + \boldsymbol{k}_3 = \boldsymbol{k}_0 + \boldsymbol{k}_1, \quad \omega(\boldsymbol{k}_2) + \omega(\boldsymbol{k}_3) = \omega(\boldsymbol{k}_0) + \omega(\boldsymbol{k}_1). \tag{1.25}$$

Now

$$\phi_2 + \phi_3 = \phi_0 + \phi_1 + \pi/2.$$

The phase difference $\phi_2 - \phi_3$ is arbitrary again.

Combination of instability and wave mixing causes complete stochastization of phases. Weak turbulence in the framework of the model (1.19) is described by the kinetic equation

$$\frac{\partial n_{\boldsymbol{k}}}{\partial t} = St(n, n, n), \tag{1.26}$$

$$St(n,n,n) = 4\pi \int |T_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{k}_3}|^2 \delta_{\boldsymbol{k}+\boldsymbol{k}_1-\boldsymbol{k}_2-\boldsymbol{k}_3} \delta_{\omega(\boldsymbol{k})+\omega(\boldsymbol{k}_1)-\omega(\boldsymbol{k}_2)-\omega(\boldsymbol{k}_3)}.$$

$$[n_{\boldsymbol{k}_1} n_{\boldsymbol{k}_2} n_{\boldsymbol{k}_3} + n_{\boldsymbol{k}} n_{\boldsymbol{k}_2} n_{\boldsymbol{k}_3} - n_{\boldsymbol{k}} n_{\boldsymbol{k}_1} n_{\boldsymbol{k}_2} n_{\boldsymbol{k}} n_{\boldsymbol{k}_1} n_{\boldsymbol{k}_3}] d\boldsymbol{k}_1 d\boldsymbol{k}_2 d\boldsymbol{k}_3.$$

It should be noted that the equations (1.23) not necessarily have real solutions. In an isotropic medium $\omega = \omega(|\mathbf{k}|)$, the sufficient condition for their existence is $\omega' > 0;\ \omega'' < 0$. If $\omega(0) = 0$, $\omega'_k > 0$, $\omega'' > 0$, the only solution of (1.22) is $\boldsymbol{k}_2 = \boldsymbol{k}_3 = \boldsymbol{k}_0$. In this case stochastization is less obvious and one has to expect formation of coherent structures. We study them in the next section.

3 Nonlinear Schrödinger equation

In some important physical situation, for instance, for waves on the surface of ideal fluid of finite depth $T_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{k}_3}$ has indeterminacies at $\boldsymbol{k}_1 = \boldsymbol{k}_2 = \boldsymbol{k}_3 = \boldsymbol{k}$. We will study only the simplest case when T is a continuous function on this submanifold. Denote $T(\boldsymbol{k}) = T_{\boldsymbol{k}\boldsymbol{k}\boldsymbol{k}\boldsymbol{k}}$. Then the equation (1.19) has the exact solution

$$a_{\boldsymbol{k}} = Ae^{-i\tilde{\omega}(\boldsymbol{k}_0)t}, \qquad \tilde{\omega}_{\boldsymbol{k}_0} = \omega_{\boldsymbol{k}_0} + T(\boldsymbol{k}_0)|A|^2. \tag{1.27}$$

Here, due to the obvious symmetry relation $T^*{}_{\boldsymbol{k}\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{k}_3} = T_{\boldsymbol{k}_2\boldsymbol{k}_3\boldsymbol{k}\boldsymbol{k}_1}$, $T(\boldsymbol{k})$ is a real function.

Let us consider a solution of (1.19) that is close to the exact nonlinear monochromatic wave (1.27). Now

$$a_{\boldsymbol{k}}(t) = C(\boldsymbol{\kappa}, t)e^{-i\omega(\boldsymbol{k}_0)t}, \qquad \boldsymbol{\kappa} = \boldsymbol{k} - \boldsymbol{k}_0, \tag{1.28}$$

$$C(\boldsymbol{\kappa}) = 0 \quad \text{if} \quad |\boldsymbol{\kappa}| \ll |k_0|.$$

Expanding $\omega(\boldsymbol{k})$ in the Taylor series

$$\omega(\boldsymbol{k}) = \omega(\boldsymbol{k}_0 + \boldsymbol{\kappa}) = \omega(\boldsymbol{k}_0) + \kappa_p \frac{\partial \omega}{\partial k_p} + \frac{1}{2}\kappa_p \kappa_q \frac{\partial \omega}{\partial k_p \partial k_q} + ...,$$

one can find that the Fourier transform from $C(\boldsymbol{\kappa}, t)$,

$$\psi(\boldsymbol{r}, t) = \int C(\boldsymbol{\kappa}, t)e^{i\boldsymbol{\kappa}\boldsymbol{r}} d\boldsymbol{\kappa},$$

satisfies the nonlinear Schrödinger equation (NLSE)

$$\frac{\partial \psi}{\partial t} + (\boldsymbol{v}\nabla)\psi - i\omega_{\alpha\beta}\frac{\partial^2 \psi}{\partial x_\alpha \partial x_\beta} = iT|\psi|^2\psi. \tag{1.29}$$

Here $\boldsymbol{v} = \partial\omega/\partial\boldsymbol{k}\big|_{\boldsymbol{k}=\boldsymbol{k}_0}$ is the group velocity,

$$\omega_{\alpha\beta} = \frac{1}{2}\frac{\partial \omega}{\partial k_\alpha \partial k_\beta}\Big|_{\boldsymbol{k}=\boldsymbol{k}_0}, \qquad T = -T(\boldsymbol{k}_0).$$

Going to the frame of reference moving with the group velocity one can eliminate the first space derivative. We now obtain

$$\frac{\partial \psi}{\partial t} - i\omega_{\alpha\beta}\frac{\partial^2 \psi}{\partial x_\alpha \partial x_\beta} = iT|\psi|^2\psi. \tag{1.30}$$

The monochromatic wave is described now by the solution of (1.28)

$$\psi = Ae^{iT|A|^2 t}. \tag{1.31}$$

One can study the stability of this solution, assuming

$$\psi = Ae^{iT|A|^2 t}(1 + \delta\psi e^{i(\Omega t - \boldsymbol{pr})}).$$

In the linear approximation $|\delta\psi| \ll 1$, one can obtain

$$\Omega^2 = (\omega_{\alpha\beta}p_\alpha p_\beta)^2 - 2T|A|^2\omega_{\alpha\beta}p_\alpha p_\beta. \tag{1.32}$$

If eigenvalues of the tensor $\omega_{\alpha\beta}$ have different signs, equation (1.31) yields instability at any sign of T. The domain of the instability in the p-space is concentrated along the cone

$$\omega_{\alpha\beta}p_\alpha p_\beta = 0. \tag{1.33}$$

If $p \simeq k_0$, this instability goes to the "second order decay instability" obeying the resonant conditions (1.23). If all eigenvalues of $\omega_{\alpha\beta}$ are of the same sign, instability takes place if

$$T\omega_{\alpha\beta}p_\alpha p_\beta > 0. \tag{1.34}$$

This instability is called the modulation instability (for details, see [1, 6, 7]). In this case the NLSE can be reduced to the form

$$i\psi_t + \Delta\psi + 2|\psi|^2\psi = 0. \tag{1.35}$$

We will call this equation the compact focusing NLSE. The domain of the instability of monochromatic wave is bounded now by the condition

$$|\omega_{\alpha\beta}p_\alpha p_\beta + \beta| < |T|A^2. \tag{1.36}$$

If $T\omega_{\alpha\beta}p_\alpha p_\beta < 0$, the monochromatic wave is stable and the NLSE can be simplified to the canonical form

$$i\psi_t + \Delta\psi - 2|\psi|^2\psi = 0. \tag{1.37}$$

This is the compact defocusing NLSE. Among non-compact NLSE the most interesting ones have the following canonical forms

$$i\psi_t + \psi_{xx} - \psi_{yy} + 2|\psi|^2\psi = 0, \tag{1.38}$$

$$i\psi_t + \Delta_\perp \psi - \psi_{xx} + 2|\psi|^2\psi = 0. \tag{1.39}$$

Here $\Delta_\perp = \partial^2/\partial y^2 + \partial^2/\partial z^2$. Equation (1.37) describes nonlinear modulations of gravity waves on a surface of deep water, while (1.38) is applicable to propagation of electromagnetic wave packets in media with negative (normal) dispersion. All species of the NLSE describe some coherent structures. Only for the compact cases (1.35) and (1.37) they are studied in a proper degree.

4 Solitons in the focusing NSLE

Development of instability of the monochromatic wave (condensate) in the framework of the compact focusing NLSE (1.35) does not lead to formation of weak-turbulent state directly. It leads first to formation of the coherent structures - solitons or collapses. In the quantum mechanical analogy, the NLSE (1.35) describes the motion of a particle in a self-consistent potential with attraction, where the attraction is the main cause of existence of the localized coherent structures. The nature of these structures depends essentially on the spatial dimension D. The most important coherent structure in (1.35) has maximal spatial symmetry. We will discuss only these structures.

Equation (1.35) can be rewritten as follows

$$i\psi_t + \psi_{rr} + \frac{D-1}{r}\psi_r + 2|\psi|^2\psi = 0, \qquad 0 < r < \infty. \tag{1.40}$$

This equation preserves two basic constant of motion: number of particles

$$N = 2^{D-1}\pi \int_0^\infty r^{D-1}|\psi|^2 dr \tag{1.41}$$

and the Hamiltonian

$$H = 2^{D-1}\pi \int_0^\infty r^{D-1}(|\psi_r|^2 - |\psi|^4)dr = X - Y, \tag{1.42}$$

where we denote

$$X = 2^{D-1}\pi \int_0^\infty r^{D-1}|\psi_r|^4 dr, \qquad Y = 2^{D-1}\pi \int r^{D-1}|\psi|^4 dr.$$

The Equation (1.40) has stationary solutions of the form

$$\psi = \varphi(r)e^{i\lambda^2 t} \tag{1.43}$$

where $\varphi(r)$ satisfies the equation

$$-\lambda^2\varphi + \Delta\varphi + 2\varphi^3 = 0 \tag{1.44}$$

Here $\Delta\varphi = \varphi_{rr} + (D-1)\varphi_r/r$. The solution (1.43) is a soliton if $\varphi(r) \to 0$ at $r \to \infty$ and integrals N, X, Y are finite. It is possible to show that the solutions of equation (1.44) for $D \leq 4$ decrease exponentially at infinity and this provides finiteness of the integrals N, X, Y.

The solution of equation (1.44) is a stationary point of the Hamiltonian for fixed number of particles N

$$\delta(H + \lambda^2 N) = 0. \tag{1.45}$$

The solution of (1.44) can be rescaled: $\varphi(r,\lambda) = \lambda\varphi_0(\lambda r)$, where $\varphi_0(\xi)$ satisfies the equation

$$-\varphi_0 + \Delta\varphi_0 + 2\varphi_0^3 = 0. \tag{1.46}$$

Hence $N = \lambda^{2-D} N_0$ with $N_0 = 2^{D-1}\pi \int\limits_0^\infty r^{D-1}\varphi_0^2(r)dr$. Let us perform the transform

$$\psi(r) \to a^{-D/2}\psi\left(\frac{r}{a}\right) \tag{1.47}$$

preserving the number of particles. As a result, the Hamiltonian takes a dependence on the parameter a

$$H(a) = \frac{X}{a^2} - \frac{Y}{a^D}. \tag{1.48}$$

According to (1.45) at the soliton solution $\partial H/\partial a|_{a=1} = 0$. Using (1.46) it is easy to get that at these solutions [1, 6]

$$X_s = \frac{D}{4-D}N_0^2/N_s, \quad Y_s = \frac{2}{4-D}N_0^2/N_s, \quad H_s = \frac{D-2}{4-D}N_0^2/N_s. \tag{1.49}$$

Here the index s denotes values of the integrals on the soliton solution. Since X, Y are positive, soliton solutions exist only if $D < 4$. Formulas (1.49), (1.58) make it possible to solve easily the question of soliton stability. If $D < 2$, $H_s < 0$, and the value $a = 1$ realizes the *minimum* of the Hamiltonian (1.48). Hence in this case one can assume that the soliton is stable. This result occurs to be true not only for scaling perturbations but also for the general ones that can be proved rigorously (see, for instance, [1],[8]).

This proof is based on the integral estimates of the Sobolev type. These inequalities arise as sequences of the general imbedding theorems between the spaces L_p and W_2^1 with the norms,

$$\|\psi\|_p = \left[\int |\psi|^p d\boldsymbol{r}\right]^{1/p}, \ (p > 0), \ \ \|\psi\|_{W_2^1} = \left[\int (|\psi|^2 + |\nabla\psi|^2) d\boldsymbol{r}\right]^{1/2}.$$

respectively. Namely, there exists such a constant $B > 0$ so that the following inequality between norms is valid (see, e.g., [9, 10]):

$$\left[\int |\psi|^p d\boldsymbol{r}\right]^{1/p} \leq B\left[\int (|\psi|^2 + |\nabla\psi|^2) d\boldsymbol{r}\right]^{1/2} \text{ if } D < \frac{2}{p}(p{+}4) \text{ and } \boldsymbol{r} \in \Re^D. \tag{1.50}$$

In this formula each integral is assumed to be convergent. Making in (1.50) the transform $\boldsymbol{r} \to \alpha\boldsymbol{r}$, it becomes

$$\int |\psi|^p d\boldsymbol{r} \leq B_1\left[\alpha^q \int |\psi|^2 d\boldsymbol{r} + \alpha^{q-2} \int |\nabla\psi|^2 d\boldsymbol{r}\right]^{p/2} \text{ with } q = d\left(1 - \frac{2}{p}\right).$$

Calculation of the minimum of the r.h.s. of this inequality with respect to scaling parameter α gives the multiplicative variant of the Sobolev inequality [9, 10],

$$\int |\psi|^p d\boldsymbol{r} \leq C\left(\int |\psi|^2 d\boldsymbol{r}\right)^{(2-q)p/4} \left(\int |\nabla\psi|^2 d\boldsymbol{r}\right)^{qp/4}, \tag{1.51}$$

where C is a new constant.

In particular, for $p = 4$ we have (compare with [11])

$$\int |\psi|^4 d\boldsymbol{r} \leq C\left(\int |\psi|^2 d\boldsymbol{r}\right)^{(4-D)/2} \left(\int |\nabla\psi|^2 d\boldsymbol{r}\right)^{D/2} \tag{1.52}$$

This inequality can be improved by finding the best constant C in (1.52). For this aim consider the functional

$$J\{\psi\} = \frac{N^{(4-D)/2} X^{D/2}}{Y}, \tag{1.53}$$

so that

$$C^{-1} = \min J\{\psi\}. \tag{1.54}$$

To find C consider all extremals of the functional $J\{\psi\}$ and take among these the one which gives a minimal value for J. Note, this functional is invariant with respect to two independent dilatations: $\psi \to \alpha\psi$ and $\boldsymbol{r} \to \beta\boldsymbol{r}$. Therefore the corresponding Euler-Lagrange equation for the functional extremum leads to

$$-\psi + \nabla^2\psi + 2|\psi|^2\psi = 0,$$

coinciding with Eq. (1.46) for the soliton solutions. A minimal value of $J\{u\}$ is attained on radically-symmetric distribution without nodes and simultaneously satisfied by Eq.(1.46). This distribution is the ground state soliton for the stationary NLSE. Hence, with account of (1.49) the best constant is equal to

$$C = \frac{2}{N_{0d}(4-D)}\left(\frac{4-D}{D}\right)^{d/2}. \tag{1.55}$$

Here N_{0D} is the number of particles in the ground state soliton N_s depending on the dimension D. For example, in the 1D case $N_{01} = 2$, at $D = 2$, according to [14] $N_{02} = 5.84$ and for $D = 3$, $N_{03} = 9.47$ [2]. As a result, the inequality (1.52) reads (see, for instance, [1] and [12])

$$Y \leq CN^{(4-d)/2} X^{d/2}. \tag{1.56}$$

This inequality allows immediately to get a proof of 1D soliton stability. Substituting (1.56) at $D = 1$ into expression (1.42) for the Hamiltonian and taking into account relations (1.49) we arrive at the following estimate (see, for instance, [13])

$$H \geq X - CX^{1/2} N^{3/2} = H_s + (X^{1/2} - X_s^{1/2})^2. \tag{1.57}$$

Thus, a 1D soliton realizes the global minimum (in the given class!) of the Hamiltonian and therefore is stable not only with respect to small perturbations but also against finite ones[1].

If $D > 2$, the stationary point yields a positive value of H, so that, instead of being a minimum in the one-dimensional case, solitons realize the maximum of the Hamiltonian, which is now unbounded from below and can take (at $a \to 0$) arbitrary large negative values. On the other hand, transformations of the type

$$\psi(r) \to \psi(r) e^{isr^2} \tag{1.58}$$

increases the integral X, leaving integrals N and Y unchanged. Hence the soliton solution is a *saddle point*, leading to the conjecture about instability of the soliton for $2 < D < 4$. This fact can be proved rigorously too.

The case $D = 2$ is special. Now $N = N_0$ and $H \equiv 0$. This result should be discussed separately. The parameter λ characterizes an inverse spatial size (width) of the soliton. Independence of λ of the basic constants of motion for $D = 2$ means that the soliton is "soft" - it can be compressed or inflated without changing its energy and number of particles. In the linear approximation the soliton is marginally stable [15]. More detailed study shows that the soliton is unstable with respect to perturbations of finite amplitude.

The application of the procedure (1.57) at $D = 2$ gives

$$H \geq X \left(1 - \frac{N}{N_{02}}\right).$$

[1]These inequalities were first used for the stability study of ion-acoustic solitons in magnetized plasma [73] based on the proof of the boundedness of the Hamiltonian. Later this approach was widely applied for the stability proof of the different kinds of solitons (see, for instance, [1]). The acknowledged best use of these inequalities for the collapse problem was presented by Weinstein [12]. Later more general results are reviewed in the paper [2].

From this estimate one can conclude that the Hamiltonian is bounded from below, taking non-negative values if the number of particles does not exceed the number of particles N_s at the ground state soliton solution. Its minimal value, equal to zero, is retained for distributions with vanishing mean square value of the wave number,

$$\langle k^2 \rangle = \frac{X}{N} \to 0.$$

Thus in this region of the phase space asymptotic states of any initial condition will be dispersively spreading distribution, i.e., asymptotically free fields.

In the three-dimensional case the analogous integral estimate for H [19],

$$H \geq X - CX^{3/2}N^{1/2}, \tag{1.59}$$

does not allow us to make any conclusion about soliton stability (note that maximum of the r.h.s. of (1.59) corresponds to a 3D soliton). Recall that the linear stability analysis predicts the instability of three-dimensional solitons [1, 15].

For $D = 1$ equation (1.40) is integrable [16] and has infinite number of extra constants of motion [16]. In this case a soliton can be found in the explicit form

$$\varphi_0(x) = \frac{1}{\cosh x}. \tag{1.60}$$

For $D = 4$ at $\lambda = 0$ the equation (1.46),

$$\varphi_0'' + \frac{3\varphi_0'}{r} + 2\varphi_0^3 = 0, \tag{1.61}$$

has the exact solution

$$\varphi_0 = \frac{2}{r^2 + 1}. \tag{1.62}$$

This is a limiting case for the soliton solutions. Now φ_0 vanishes powerfully at $r \to \infty$, that results in the logarithmical divergence of N_0 at $r \to \infty$.

5 Collapses in the NLSE

For $D \geq 2$ solitons are either unstable or do not exist. In this case the major coherent structure is a collapsing cavity (the region of higher wave intensity) leading to the formation of localized singularities of wave amplitude in a finite time.

One of the main reasons for the wave collapse existence is the Hamiltonian unboundedness. In such systems, like the NLSE, collapse can be represented as a process of falling down of some "particle" in a self-consistent

unbounded potential. Indeed, the picture is more complicated than considered above. From the very beginning we have a spatially-distributed system with infinite number of degrees of freedom and therefore, rigorously speaking, it is hard to describe such a system by its reduction to a system of ODEs. The NLSE is a wave system and wave radiation plays a very essential role for blow-up.

Let Ω be an arbitrary region with a negative Hamiltonian H_Ω. Then using the mean value theorem for the integral Y_Ω,

$$\int_\Omega |\psi|^4 \, d\boldsymbol{r} \leq \max_{r\in\Omega} |\psi|^2 \int_\Omega |\psi|^2 d\boldsymbol{r},$$

one can get (compare with [6] and [72]) the following inequality

$$\max_{x\in\Omega} |\psi|^2 \geq \frac{|H_\Omega|}{N_\Omega}. \tag{1.63}$$

Here the expression in the r.h.s. of the inequality has the meaning of the mean energy per one quasi-particle. From this inequality, valid also when $\Omega = \Re^3$, it follows that $\max |\psi|^2$ as a function of t always is majorized by the conservative value. So, vanishing or yet some sufficient decreasing of the initially existed maximum of $|\psi|^2$ are impossible.

Let the Hamiltonian be negative initially in some separate region Ω, $H_\Omega < 0$, and the radiation emerge from this region. In the outer region, far from Ω, radiative waves will have small amplitudes. Consequently, their nonlinear interaction will be negligible with respect to their dispersion and they will have a positive Hamiltonian. Therefore due to the wave radiation, the Hamiltonian of the region H_Ω will become more and more negative increasing its absolute value, that is possible only due to the *unboundedness* of the Hamiltonian. Simultaneously, N_Ω as a positive value will decrease so that the ratio in the r.h.s. of the inequality (1.63) will increase. It automatically leads to the growth of the maximal value of $|\psi|^2$. Thus, radiation, as a dissipative process promotes the wave collapse.

The occurrence of wave collapses can be proved by use of the virial theorem. From (1.40) one can derive the relation

$$\frac{d^2}{dt^2} \int r^2 |\psi|^2 d\boldsymbol{r} = 4[2H - (D-2)Y]. \tag{1.64}$$

At $D = 2$ this relation can be integrated twice

$$\langle r^2 \rangle = \frac{\int r^2 |\psi|^2 d\boldsymbol{r}}{\int |\psi|^2 d\boldsymbol{r}} = 4\frac{H}{N} t^2 + C_1 t + C_2, \tag{1.65}$$

where the constants C_1, C_2 are defined from the initial conditions. Hence it is seen that for $H < 0$, in spite of the values $C_{1,2}$ there always exists a finite time when the right hand side of (1.65) vanishes. Thus, $H < 0$ represent a

sufficient condition for the collapse, which was found by Vlasov, Petrishchev and Talanov (the VPT criterion [17]). For $D = 3$ the equality (1.65) can be replaced by the inequality

$$\langle r^2 \rangle < 4\frac{H}{N}t^2 + C_1 t + C_2. \tag{1.66}$$

from which follows the same sufficient criterion $H < 0$ [6]. This estimate, however, is rather rough and can be improved. As was shown in a recent paper [19], the collapse threshold is defined by the unstable ground state soliton solution which in some sense plays the role of separatrix between collapsing and noncollapsing solution. It was proved in [19] that at $D = 3$ the equality (1.64) can be changed to the inequality

$$\frac{d^2}{dt^2}\int r^2|\psi|^2 dr < 8(H - H_N). \tag{1.67}$$

Here $H_N = N_0^2/N$ is the value of the Hamiltonian of the ground state soliton (compare with (1.49)). Hence the equation (1.67) gives the sharper criterion for collapse [18, 19]

$$H \le H_N. \tag{1.68}$$

What is the scenario of the collapse? For $D > 2$ NLSE (1.40) has the self-similar solution

$$\psi(r,t) = \frac{1}{(t_0 - t)^{1/2 + i\kappa(D)}} g(\xi), \qquad \xi = \frac{r}{\sqrt{t_0 - t}} \tag{1.69}$$

where $g(\xi)$ satisfies the equation

$$g_{\xi\xi} + \left(\frac{D-1}{\xi} + \frac{i\xi}{2}\right) g_\xi + \left(\frac{i}{2} - \kappa\right) g + 2|g|^2 g = 0, \tag{1.70}$$

$$g_\xi|_{\xi=0} = 0, \qquad g(\infty) = 0.$$

Here $\kappa = \kappa(D)$ is the eigenvalue of the nonlinear boundary problem (1.70). It is easy to show that as $\xi \to \infty$

$$g(\xi) \simeq \xi^{-(1+2i\kappa)}, \tag{1.71}$$

hence $|\psi|^2 \to 1/r^2$ as $t \to t_0$. For $D > 2$ the singularity (1.71) is integrable.

There is a plausible hypothesis: the self-similar solution (1.69) describes collapse in a general position. So far, the only way to check this conjecture is by a numerical experiment. Two series of experiments performed by two independent groups confirmed the hypothesis with a very high accuracy [20, 23]. This problem was also discussed in many other papers. We would like to draw attention to some of them - [74] and [30].

A generic case $D > 2$ can be called supercritical. The case $D = 2$ is critical. This case is especially interesting because it describes stationary self-focusing of electromagnetic waves in a nonlinear Kerr dielectric.

For $D \leq 2$ the singularity (1.71) is non-integrable, and the boundary problem (1.70) cannot have regular solutions. In the critical case $D = 2$, $N_s = N_0$, $H_s = 0$, and one can guess that the collapse is the compressing soliton [24]

$$|\psi|^2 = \frac{1}{f^2}\varphi_0^2\left(\frac{r}{f}\right) + ..., \qquad f = f(t_0 - t), \quad f(0) = 0. \tag{1.72}$$

In the strictly self-similar case $f(\xi) = \sqrt{\xi}$. As far the divergence at $D = 2$ is very weak (logarithmic) one can conjecture that now

$$f(\xi) = \sqrt{\frac{\xi}{b(\xi)}}.$$

Here $b(\xi)$ is a "slow" function and $b(0) = \infty$. It was shown [27, 28] that

$$b(\xi) \simeq \ln\ln\left(\frac{1}{\xi}\right).$$

This result is confirmed by numerical experiments with satisfactory accuracy [21, 34, 33, 22] and also analytically [29].

At the end of this section we want to discuss the possibility of collapse in the non-compact NLSE (1.37) and (1.38). First of all, it is easy to show that solitons are absent in this case. The explanation of this fact is very simple. In the transverse plane, equation (1.38) describes attraction between particles, but, in contrast, along the x-axis repulsion. Moreover, from the virial identities for mean transverse size and mean longitudinal size one can show that collapse of the wave packet as a whole is impossible at the stage of the compression of the wave packet in all directions ([35]). Numerical integration of these equations (as it was published in the first paper [36], devoted to this subject, as well as in the recent one [37]), demonstrates the fractal behavior of the system. The initial distribution with sufficiently large amplitude at the beginning demonstrates compression in the transverse plane; at the later stage the wave packet undergoes waving instability that results in splitting of the packet into two packets. At the next stage dynamics of each secondary packet repeats the fate of the original one.

6 Weak, strong and superstrong collapses

The central problem of the physical theory of collapse is the estimate of the efficiency of collapse as a nonlinear mechanism of wave energy dissipation. To achieve that we must include the nonlinear dissipative terms into

equations describing the collapsing medium. The nonlinear Schrödinger equation could be modified as follows

$$i(\psi_t + \beta|\psi|^m\psi) + \psi_{rr} + \frac{D-1}{r}\psi_r + 2|\psi|^2\psi = 0. \tag{1.73}$$

Here for $\beta > 0$ the second term is responsible for the nonlinear dissipation. For a sufficiently large degree of nonlinearity m, the equation (1.73) has regular solutions for some small β. The amount of absorbed energy during the collapse is characterized by the integral

$$I \equiv \frac{dN}{dt} = \beta\pi \int dt \int_0^\infty |\psi|^{m+2}(2r)^{D-1}dr. \tag{1.74}$$

Our aim now will be to estimate integral (1.74) at $\beta \to 0$. When approaching the collapse, there are two possibilities. In the critical case, $D = 2$, a strong collapse occurs when a finite amount of energy is accumulated at a collapse point and, as a result, the δ-type singularity is formed. The direct numerical solution of the equation (1.73) confirmed that idea. It was found that the part of energy absorbed during the collapse is about 15% to 25% from the value of N_{cr} [21, 22]. This part practically does not change with the decreasing of β, and slightly reduces when m is increased. In the supercritical case, the integrable singularity of wave energy density is formed in the collapse point. We have

$$|\psi|^2 \sim 1/r^2. \tag{1.75}$$

Let the characteristic size of the collapsing cavity be of the order of r_0 then the characteristic formation time of that scale is $\Delta t \sim r_0^2$. Substituting into (1.74), we obtain

$$I \sim \beta(\Delta t)^{(D-m)/2} \sim \beta r_0^{D-m}. \tag{1.76}$$

From (1.76) it is clear that the nonlinear damping is efficient if $m \geq D$. If we agree that all energy in the collapse zone, $\Delta N \sim r_0^{D-2}(D-2)$, is absorbed, we have

$$r_0 \sim [\beta(D-2)]^{\frac{1}{m-2}}, \quad I \sim (D-2)^{-\frac{m-D}{m-2}}\beta^{\frac{D-2}{m-2}}. \tag{1.77}$$

So, $I \to 0$ at $\beta \to 0$. Such a collapse can be called weak [8]. From (1.77) we can see that for $D \to 2$ the weak collapse becomes a strong one.

The previous considerations were based on the assumption that only the energy arrived at the collapse moment of time $t = t_0$ dissipates in the collapse point. That is not always true. In the point of collapse there could be formed a zone of energy dissipation that absorbs the energy from the surrounding area. In this case the life time of the collapse $\tau \gg \Delta t$, and one must solve the problem of the entire wave packet to estimate the

absorbed energy. We suggest calling such a black hole regime a "superstrong collapse", because for a sufficiently large τ the full absorbed energy can exceed the absorption energy for the strong collapse regime. To describe the superstrong collapse it is necessary to obtain solutions of the stationary equation

$$i\beta|\psi|^m + \psi_{rr} + \frac{D-1}{r}\psi_r + 2|\psi|^2\psi = 2|\psi_0|^2\psi \tag{1.78}$$

with boundary conditions

$$\psi_r|_{r=0} = 0, \quad \psi \to \psi_0 \text{ as } r \to \infty.$$

The existence of the black hole also means that at the limit $\beta \to 0$, the equation

$$\psi_{rr} + \frac{D-1}{r}\psi_r + 2|\psi|^2\psi = 2|\psi_0|^2\psi \tag{1.79}$$

has a singular solution with a constant energy flux to the collapse point at $r = 0$

$$P = \lim_{r\to 0} \pi(2r)^{D-1} \operatorname{Im}(\psi\psi_r^*). \tag{1.80}$$

Let $D > 2$. Then the equation

$$\psi_{rr} + \frac{D-1}{r}\psi_r + 2|\psi|^2\psi = 0 \tag{1.81}$$

has the exact solution [23]

$$\psi = A_0/r, \qquad A_0 = \left(\frac{D-3}{2}\right)^{1/2}. \tag{1.82}$$

This solution can be used as a first step to construct a singular solution of the equation (1.81). Indeed, the solution near zero could be found as

$$|\psi| = \frac{A_0}{r}(1 + A_1 r^\mu + ...), \qquad \psi \to \psi_0, \qquad \mu = 2(4-D) > 0. \tag{1.83}$$

Here A_1 is an arbitrary constant, and $A_1 = qP^2$, where q is some multiplier. By the selection of P one can obtain the asymptotic solution of the equation (1.83) for $r \to \infty$. The numerical integration of (1.78) showed that in the interval $3 < D < 4$ the solution describing the superstrong collapse really can be constructed. Such solutions exist in a rather important physical case at $D = 3$, which corresponds to a $3-D$ nonstationary self-focusing. In this case the main asymptotic term at zero is a stationary solution [31, 32]

$$|\psi| = \frac{1}{2r|\ln\ r|^{1/2}}.$$

Superstrong collapse can exist also for power nonlinearity $|\psi|^{2n}\psi$ where $nD > 4$. Thus, at $D = 4$ the equation (1.81) has an exact singular solution

$$|\psi| = B/r,$$

whose amplitude is defined by the flux P from the equation

$$B^4(B^2-1)=P^2.$$

Finally, when $D>4$, equation (1.81) has quasi-classical stationary solutions with an asymptotic expansion at zero

$$\begin{aligned} |\psi| &= \frac{c}{r^\gamma}(1+c_1 r^\nu+...), \quad c=P^2, \quad \gamma=\alpha(\alpha-1), \\ \nu &= \alpha(2D-8)>0, \quad \alpha=\frac{1}{3}, \quad c_1(D)>0. \end{aligned}$$

It is important that the quasi-classical criterion for this solution improves while approaching the singular point ($r=0$).

The existence of such solutions was also confirmed by the numerical integration of equation (1.78) at $\psi_0=0$ [23, 21].

7 Anisotropic black holes

As we saw in the previous section the black hole regime becomes quasi-classical starting from $D=4$. In this section we want to present an example showing how, due to the medium anisotropy, the "effective dimension D'' can be greater than 4 and, as a consequence, the black-hole regime can be realized.

We consider the upper-hybrid waves Langmuir waves in a plasma with sufficiently small magnetic field ($\omega_{pe}\gg\omega_{ce}$) when all changes in the dispersion law are expressed in the form of the additive term

$$\omega_k=\omega_{pe}\left(1+\frac{3}{2}k^2r_d^2+\frac{1}{2}\frac{\omega_{ce}^2}{\omega_{pe}^2}\frac{k_\perp^2}{k^2}\right), \tag{1.84}$$

where ω_{ce}, ω_{pe} are electron gyrofrequency and electron plasma frequency, respectively, $r_d=v_{Te}/\ \omega_{pe}$ is the Debye radius, $\boldsymbol{k}_\perp$ is the component transverse to the external magnetic field $\boldsymbol{B}_0$, directed along the z axis. In the dispersion law (1.84), the first term describes the potential electron plasma oscillations with a plasma frequency. Other terms are due to slower processes. In isotropic case ($\omega_{ce}=0$) (1.84) transforms into the dispersion law for the Langmuir waves.

The nonlinear effect, in a small-amplitude region ($E^2/8\pi nT\ll m/M$), m and M being the electron and ion masses, respectively, corresponds to the nonlinear frequency shift due to the interaction with slow adiabatic plasma flows induced by high-frequency plasma oscillations. In this limit, the equation for the envelope of high-frequency oscillations in dimensionless variables can be written as follows [38]

$$\Delta(i\psi_t+\Delta\psi)-\sigma\Delta_\perp\psi+\nabla(|\nabla\psi|^2\nabla\psi)=0, \tag{1.85}$$

where ψ is the envelope of high-frequency waves and $\sigma = \omega_{ce}^2/2\omega_{pe}^2$. Respectively, the low-frequency plasma fluctuations follow adiabatically the pondermotive pressure of high-frequency waves,

$$n = -|\nabla\psi|^2.$$

The equation (1.85) at zero magnetic field transforms into the Zakharov equation describing collapse of Langmuir waves [6] in the so-called static approximation.

The equation (1.85) can be further reduced under additional assumptions. It is known that due to weak turbulent processes, such as induced scattering of ions or four-wave interaction, the energy transfer by cascade to the region $\omega_k \to \omega_{pe}$. If one studies these processes in more details, it is possible to find that they lead, in the first stage, for waves with $(kr_d)^2 < \sigma$, to a rapid decrease of $k_\perp$, and only subsequently, to a reduction of k_z up to a zero value. This means that the wave condensate will have characteristic longitudinal scales smaller than the transverse ones. Under this assumption Eq. (1.85) reads as follows

$$\frac{\partial^2}{\partial z^2}\left(i\psi_t + \frac{\partial^2}{\partial z^2}\psi\right) - \Delta_\perp\psi + \frac{\partial}{\partial z}\left(\left|\frac{\partial\psi}{\partial z}\right|^2\frac{\partial\psi}{\partial z}\right) = 0. \tag{1.86}$$

where we put, without any restriction, the constant $\sigma = 1$, that corresponds to a simple rescaling.

The equation (1.86) can also be written in the Hamiltonian form

$$i\frac{\partial^2}{\partial z^2}\psi_t = \frac{\delta H}{\delta\psi^*}, \tag{1.87}$$

where the Hamiltonian

$$H = \int(|\psi_{zz}|^2 + |\nabla_\perp\psi|^2 - \frac{1}{2}|\psi_z|^4)d\boldsymbol{r} \equiv I_1 + I_2 - I_3. \tag{1.88}$$

The possible stationary solutions of this equation should correspond to the soliton-like solution

$$\psi = \psi_0\exp(i\lambda^2 t),$$

where ψ_0 satisfies the equation

$$\frac{\partial^2}{\partial z^2}\left(-\lambda^2\psi_0 + \psi_{0zz}\right) - \Delta_\perp\psi_0 + \left(|\psi_{0z}|^2\psi_{0z}\right) = 0. \tag{1.89}$$

Localized solutions of this equation simultaneously represent stationary points of the Hamiltonian for a fixed number of particles $N = \int|\psi_z|^2 d\boldsymbol{r}$ (coinciding up to a constant multiplier with the energy of high-frequency waves),

$$\delta(H + \lambda^2 N) = 0. \tag{1.90}$$

Performing now the scaling transformation retaining N,

$$\psi \to \frac{a^{1/2}}{b}\psi\left(\frac{z}{a}, \frac{r_\perp}{b}\right),$$

instead of (1.47) for the NLSE, H (1.88) becomes a function of two scaling parameters,

$$H(a,b) = \frac{I_1}{a^2} + \frac{I_2}{b^2}a^2 - \frac{I_3}{ab^2}. \tag{1.91}$$

The variational problem (1.90) now yields two relations between integrals I_l $(l = 1, 2, 3)$ on the solution ψ_0

$$-2I_1 + 2I_2 + I_3 = 0, \ -I_2 + I_3 = 0. \tag{1.92}$$

Another relation follows after multiplication of (1.89) by ψ_0 and integration

$$\lambda^2 N + I_1 + I_2 - 2I_3 = 0. \tag{1.93}$$

After a simple algebra based on (1.92) and (1.93), one can show that

$$I_1 = -2\lambda^2 N < 0,$$

contradicting the sign of I_1 which is positive definite. This contradiction implies that for (1.86) stationary soliton solutions do not exist [39]. This is possible to understand considering the NLSE (1.40) as an example. According to (1.49), soliton solutions in the NLSE exist for $D \leq 4$ and are absent for $D > 4$ that corresponds to the well-known general fact, i.e., to increase of the role of nonlinear effects with growth of dimension D. Consider now the parabolic family $b = \gamma a^2$, where γ is a constant. For this kind of curve, the first two terms in $H(a,b)$ (1.91) have the same (self-similar) behavior (dependence)

$$H(a,\gamma) = \frac{1}{a^2}\left(I_1 + \frac{I_2}{\gamma^2}\right) - \frac{1}{a^5}\left(\frac{I_3}{\gamma^2}\right). \tag{1.94}$$

Hence, firstly, one can see that the Hamiltonian is unbounded from below as $a \to 0$ that is one of the necessary conditions for the existence of collapse. Secondly, the comparison of (1.94) with (1.48) shows that the equation (1.86) is equivalent to the NLSE with $D = 5$. According to our classification presented in the previous sections, the dimension $D = 5$ corresponds to superstrong collapse providing existence of quasi-stationary black-hole regime. Moreover, this regime can be described in terms of semi-classical approach. The latter assumes solutions of the equation (1.86) in the form $\psi = Ae^{i\Phi}$ where we impose the following (semi-classical) restrictions on the phase Φ and the amplitude A

$$|\Phi_t|T \gg 1, \quad |\Phi_z|L_z \gg 1, \quad |\nabla_\perp \Phi|L_\perp \gg 1. \tag{1.95}$$

Here T is the characteristic time of the amplitude variation, L_z and $L_\perp$ are characteristic longitudinal and transverse scales of the amplitude, respectively.

Under these assumptions, in the leading order we have the Hamilton-Jacobi equation for the eikonal Φ

$$\Phi_t + \Omega(\nabla\Phi) - n = 0 \tag{1.96}$$

where $\Omega(\boldsymbol{k}) = k_z^2 + k_\perp^2/k_z^2$ is the dispersion relation for small-amplitude waves describing by the linearized equation (1.86), $\boldsymbol{k} = \nabla\Phi$ is the wave vector and $n = |\psi_z|^2 \simeq A^2\Phi_z^2$ is the wave intensity. At the next order we arrive at the continuity equation for n

$$n_t + \mathrm{div}(n\boldsymbol{V}) = 0. \tag{1.97}$$

Here $\boldsymbol{V} = \partial\Omega/\partial\boldsymbol{k}$ is the group velocity. Eqs. (1.96), (1.97) retain the Hamiltonian structure

$$\begin{aligned} n_t &= \frac{\delta H}{\delta\Phi}, \quad \Phi_t = -\frac{\delta H}{\delta n}, \\ H &= \int\left[\Omega(\nabla\Phi)n - \frac{n^2}{2}\right] d\boldsymbol{r}. \end{aligned}$$

It is possible to show that Eqs. (1.96) and (1.97) have the whole family of collapsing solutions starting from semi-classical strong collapse up to the weakest collapse corresponding to self-similar solution of the equation (1.86) (for details, see [40, 39]). All semi-classical collapsing regimes occur to be unstable. Therefore at the initial stage of a collapse we have the formation of a weak singularity which later on serves as the origin for the appearance of a black hole. To find a structure of a black hole, it is enough to take semi-classical equations (1.96), (1.97) and to seek for solutions in the form of an anisotropic funnel

$$\Phi = \frac{1}{z}\phi(\eta), \quad n = \frac{1}{z^4}g(\eta). \tag{1.98}$$

Here $\eta = r_\perp/z^3$ is a new self-similar variable and the function $g(\eta)$, as it is easy to show, obeys the ordinary differential equation

$$gg' + 3\eta(g + 3\eta g')^4 = 0. \tag{1.99}$$

Solutions of this equation only depend on the constant

$$P = \int_0^\infty (1 - 3\eta g' g^{-1})g' d\eta,$$

which is the energy flux into the singularity. Numerical calculations of (1.99) showed the existence of monotonically vanishing solutions with the asymptotics $g \sim \eta^{-1/3}$ at $\eta \to \infty$. It should be added that the semi-classical

criterion (1.95) for the solution (1.98) improves as $\boldsymbol{r}$ approaches the singular point.

The similar situation arises for lower-hybrid waves near the lower-hybrid resonance ω_{LH}. In the case, when $\omega_{ce} \gg \omega_{pe}$, the dispersion law of waves is

$$\omega_k = \omega_{LH}\left(1 + k_\perp^2 R^2 + \frac{1}{2}\frac{m}{M}\frac{k_z^2}{k_\perp^2}\right),$$

where $R = [3/4 + (3T_i/T_e)]r_{ce}$, r_{ce} is the electron gyro-radius, while m and M are the electron and ion mass, respectively. For low wave intensity, as for UH waves, the low-frequency plasma-density variation is related to the high-frequency pondermotive force through [41, 42]

$$n = i[\nabla\psi \times \nabla\psi^*]_z \ . \tag{1.100}$$

Here, as in a previous case, we write (1.100) in dimensionless variables, ψ stands for the envelope of the high-frequency electric potential of LH waves.

The evolution equation for the envelope is obtained by usual average over the high-frequency ω_{LH} [41, 42]

$$\Delta_\perp(i\psi_t + \Delta_\perp\psi) - \alpha\partial_z^2\psi - \nabla_\perp([\nabla_\perp\psi \times \nabla_\perp\psi^*]_z[\boldsymbol{n} \times \nabla_\perp\psi]) = 0. \tag{1.101}$$

Here $\boldsymbol{n} = \boldsymbol{B}_0|\boldsymbol{B}_0|$ and α is a constant. This equation can be written in the Hamiltonian form

$$\Delta_\perp i\psi_t = \frac{\delta H}{\delta\psi^*}, \tag{1.102}$$

$$H = \int\left(|\Delta_\perp\psi|^2 + |\psi_z|^2 + \frac{1}{2}[\nabla_\perp\psi \times \nabla_\perp\psi^*]^2\right)d\boldsymbol{r} \equiv I_1 + I_2 - I_3.$$

The same analysis as it was done for UH waves demonstrates that solitons are absent for the model (1.101). The Hamiltonian under scaling transformations, $\psi(z, r_\perp) \to a^{-1/2}\psi(z/a, r_\perp/b)$, regaining the wave energy $N = \int|\nabla_\perp\psi|^2 d\boldsymbol{r}$, behaves as follows [39]

$$H(b,\gamma) = \frac{1}{b^2}\left(I_1 + \gamma^2 I_2\right) - \frac{1}{b^4}\left(\frac{I_3}{\gamma}\right),$$

where $H(b,\gamma)$ is taken along the parabolas $a = \gamma b^2$. Thus, for the effective dimension $D = 4$ weak collapse forms initially a singularity, which eventually transforms into a black hole. It is interesting to note that the effective dimension D for the black-hole regime corresponds to a lower boundary of the semiclassical black holes [45].

At the end of this Section we would like to note the recent experimental observations of quasi-stationary localized structures in the auroral ionosphere (at altitudes near 800 km) [43]. The wavelet analysis of these measurements by the plasma wave interferometer aboard the AMICIST rocket

demonstrated in the region of lower-hybrid frequency the existence of long-life-time solitary structures possessing rotating eigenmodes [44]. These observations are consistent with the results of three-dimensional numerical experiments which showed the presence of a cavity density. In our opinion, these objects are the first candidates for black holes.

8 Structure in media with weak dispersion

Let us consider another situation where phase stochastization plays a less important role than coherent structures. This is propagation of waves in a media with weak dispersion. Suppose first that dispersion is absent entirely. In an isotropic medium

$$\omega(k) = c|k| \tag{1.103}$$

where c has a meaning of sound speed.

Now resonant conditions (1.9) can be satisfied only if all three vectors $\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3$ are parallel. In particular, they are satisfied, if $\boldsymbol{k}_2 = \boldsymbol{k}_3 = \boldsymbol{k}, \boldsymbol{k}_1 = 2\boldsymbol{k}$. It means that the monochromatic wave cannot exist for a long time; it produces second harmonic, then zero and higher harmonics. Phases of all harmonics are correlated. This creates favorable conditions for the formation of coherent structures. This family is especially rich, if the dispersion relation is not exactly linear

$$\omega(|\boldsymbol{k}|) = c(|\boldsymbol{k}| + L(\boldsymbol{k})), \quad |L(\boldsymbol{k})| \ll \boldsymbol{k}, \quad L(0) = 0. \tag{1.104}$$

If $L''(\boldsymbol{k}) > 0$, the resonant conditions (1.9) are satisfied when all three vectors $\boldsymbol{k}_i$ are almost parallel. Then it is possible to consider the situation when the support of the function $a(\boldsymbol{k})$ is concentrated on an almost one dimensional set. In other words, one can present the wave vector in the form

$$\boldsymbol{k} = (p, \boldsymbol{q})$$

and consider the complex amplitude $a(p, \boldsymbol{q}) \neq 0$ only if $|\boldsymbol{q}| \ll p,\ p > 0$. Here $p, \boldsymbol{q}$ are components along and across the direction of the wave propagation. Now

$$|\boldsymbol{k}| = \sqrt{p^2 + \boldsymbol{q}^2} \simeq p + \frac{1}{2}\frac{\boldsymbol{q}^2}{p}, \tag{1.105}$$

and one can put approximately

$$\omega(p, \boldsymbol{q}) \simeq c\left(p + \frac{1}{2}\frac{\boldsymbol{q}^2}{p} + L(p)\right), \qquad L(-p) = -L(p). \tag{1.106}$$

The most natural model of acoustic waves is compressible ideal hydrodynamics with dependence of internal energy of both density ρ and its gradient. In particular, ion-acoustic waves in isotropic plasma relate to this kind of waves and can be described in terms of ideal hydrodynamics with

dispersion. We can use this model for calculation of the coupling coefficient for three-wave interaction. Skipping the details (see [3], e. g.), we present the result of these calculations

$$V(\boldsymbol{k}, \boldsymbol{k}_1, \boldsymbol{k}_2) \simeq V(p, p_1, p_2) = \mu (p p_1 p_2)^{1/2}; \qquad p_i > 0 \tag{1.107}$$

where μ is a constant expressing through the characteristics of the media: mean density, sound speed and internal energy.

Let D be the space dimension. One can introduce a new unknown function

$$u(x, r, t) = -\frac{1}{(2\pi)^D} \int\limits_{p>0} \sqrt{p} (a_{p,q} + a^*_{-p,-q}) e^{ip(x-t)+i\boldsymbol{qr}}\, dp d\boldsymbol{q}. \tag{1.108}$$

After simple transformation one can find that in this case (1.8) takes the form

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \hat{L}\, u - \nu \frac{\partial^2 u}{\partial x^2}\right) = -\Delta_\perp u. \tag{1.109}$$

Here

$$\hat{L}\left(\frac{\partial}{\partial x}\right) = i\, \hat{L}\left(-i\frac{\partial}{\partial x}\right) = -\,\hat{L}\left(-\frac{\partial}{\partial x}\right) \tag{1.110}$$

is the operator responsible for dispersion and $\Delta_\perp$ is the Laplacian with respect to $\boldsymbol{r}$. In the two dimensional space $\Delta_\perp u = u_{yy}$, in the three-dimensional case $\boldsymbol{r} = (y, z)$ and $\Delta_\perp u = u_{yy} + u_{zz}$. In equation (1.109) we introduced the dissipative (viscous) term νu_{xx}.

If $\omega^2 = \omega^2(k^2)$ is an analytic function of k^2 $L(p)$ is an odd function: $L(p) = -L(-p)$. In the simplest case

$$L(p) = \pm p^3, \quad \hat{L}\left(\frac{\partial}{\partial x}\right) = \mp \frac{\partial^3}{\partial x^3}. \tag{1.111}$$

In the 2D case for $\nu = 0$ we obtain now the Kadomtsev-Petviashvili equations [46, 47]:

the KPI equation -

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - \frac{\partial^3 u}{\partial x^3}\right) = -\frac{\partial^2 u}{\partial y^2} \tag{1.112}$$

and the KPII equation -

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3}\right) = -\frac{\partial^2 u}{\partial y^2}. \tag{1.113}$$

The KPI equation describes acoustic-type waves with positive dispersion. These are magneto-acoustic waves in strongly magnetized plasma

with propagation angles not close to transverse and parallel directions of an external magnetic field, phonons under some conditions [48], gravity-capillary waves for the shallow water limit.

The KPII equation corresponds to the negative dispersion. Gravity waves for shallow water, ion-acoustic waves in isotropic plasma, magneto-acoustic waves for perpendicular to magnetic field propagation belong to such type of waves.

The general equation (1.109) can be called the generalized KP equation. We note that this equation is written in a frame moving with the sound speed c, which coincides with the group velocity of the waves at $k = 0$. All other terms describe slow process with respect to this propagation: the second term represents the nonlinear renormalization of the sound speed, the third one is responsible for weak dispersion and, finally, the term in the r.h.s. of (1.109) refers to transverse diffraction of acoustic waves.

In many interesting cases, $L(p)$ is not an analytic function in the vicinity of $p = 0$. In this case $L\,(\partial/\partial x)$ is a pseudo-differential operator. For instance, among the operators

$$L(p) = \pm p|p|^{2s}, \quad s > 0, \quad \hat{L} = \mp \frac{\partial}{\partial x} \left| \frac{\partial}{\partial x} \right|^{2s} \tag{1.114}$$

the choice $s = 1/2$, $\Delta_\perp u = 0$, $\nu = 0$ corresponds to the well-known Benjamin-Ono equation applicable for description of internal waves. The generalized KP equation (GKP) describes the wide spectrum of coherent structures including solitons, collapses and black holes. Let us consider the simplest examples of such structures.

Neglecting by dispersion, dissipation and diffraction, one arrives at the Hopf equation

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0. \tag{1.115}$$

The general solution of this equation is given in the implicit form

$$x = ut + F(u) \tag{1.116}$$

where $F(u)$ is arbitrary function. Let $F(u) = -ut_0 + u^3$. The corresponding solution has a self-similar form

$$u = (t_0 - t)^{1/2} g\left(\frac{x}{(t_0 - t)^{3/2}} \right). \tag{1.117}$$

Here $g(\xi)$ is the solution of the cubic equation

$$g^3 = g + \xi. \tag{1.118}$$

The solution (1.117) describes self-similar wave collapse. As a result, the first derivative of u becomes infinite in a finite time. Indeed, according to (1.117)

$$\frac{\partial u}{\partial x}|_{x=0} \simeq \frac{1}{t_0 - t}.$$

At the moment of collapse ($t = t_0$) $u = x^{1/3}$. This example of the wave collapse is known as the wave breaking.

Let us consider the influence of the neglected factors to the process of collapse. Taking into account the dependence of the diffraction term we have the dispersionless KP equation

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right) = -\frac{\partial^2 u}{\partial y^2}. \tag{1.119}$$

In 3D media one has to replace $\partial^2 u/\partial y^2$ by $\Delta_\perp u$. Weak dependence on the perpendicular coordinate in the solution (1.116) might be taken into account by replacing in the solution (1.116) $t_0 \to t_0 + \epsilon y^2$. In this case, comparing competing terms in (1.119) ,

$$\frac{\partial}{\partial x}\left(u\frac{\partial u}{\partial x}\right) \simeq \frac{1}{(t_0 - t)^2}; \quad \frac{\partial^2 u}{\partial y^2} \simeq \frac{\epsilon}{(t_0 - t)^{3/2}}, \tag{1.120}$$

one can see that weak dependence on the perpendicular coordinates does not arrest the wave breaking.

On the contrary, both dissipation and dispersion arrest the collapse. In the presence of dissipation the equation (1.109) transforms (in the one-dimensional case) into the Burgers equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}. \tag{1.121}$$

As soon as for the wave breaking $uu_x \simeq 1/(t_0 - t)^{1/2}$ and the dissipation term can be estimated as $u_{xx} \simeq 1/(t_0 - t)^{5/2}$, collapse is seen to be arrested even by an infinitesimally small viscosity ν. To estimate the dissipation efficiency one can exploit the identity

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} u^2 dx = -\nu\int_{-\infty}^{\infty} u_x^2 dx. \tag{1.122}$$

As soon as $u_x^2 \simeq 1/x^{4/3}$, the integral in the right hand side of (1.122) converges as t approaches t_0

$$\nu\int_{-\infty}^{\infty} u_x^2 dx \simeq \frac{\nu}{(t_0 - t)^{1/2}}.$$

Similarly, the total amount of absorbed energy is

$$\Delta E = \int_0^{t_0} dt\frac{\partial}{\partial t}\int_{-\infty}^{\infty} u^2 dx \simeq 2\nu {t_0}^{1/2}.$$

Note that $\Delta E \to 0$ as $\nu \to 0$. By definition this is *weak collapse.*

What is the remote results of the collapse in this case? The Burgers equation (1.121) has a solution in the form of a stationary propagating shock wave

$$u = \frac{2s}{1 + \exp[s(x + st)/\nu]}. \tag{1.123}$$

Calculating now the rate of dissipation for this solution, one can get

$$\nu \left(\frac{\partial u}{\partial x}\right)^2 = \frac{s^4}{4\nu \cosh^4[s(x + st)/\nu]}.$$

Thus, dissipation is concentrated in a very small domain near $x = -st$. The shock wave is a moving sink of energy, i.e. a black hole of codimension 1. It can be compared with the black hole of dimension zero (codimension three) which can arise after the formation of weak singularity in the 3D NLSE (see Section 6).

Another fundamental effect, arresting the collapse of gradients, is the wave dispersion. Suppose that

$$L\left(\frac{\partial}{\partial x}\right) u = -\frac{\partial}{\partial x}\left|\frac{\partial}{\partial x}\right|^{2s} u, \tag{1.124}$$

where $|\partial/\partial x|^{2s}$ is the operator with symbol $|k|^{2s}$. Then the equation (1.109) in 1D case takes the form of the generalized Korteweg-de-Vries equation (GKDV).

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - \frac{\partial}{\partial x}\left|\frac{\partial}{\partial x}\right|^{2s} u = 0 \tag{1.125}$$

For $s = 1$ this equation transforms into the classical KDV equation. At $s = 1/2$ (1.125) coincides with the Benjamin-Ono equation.

Comparison of linear and nonlinear terms in (1.125) shows that the collapse of gradients (wave breaking) is impossible for any $s > 0$. The equation (1.125) can be presented as follows

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\frac{\delta H}{\delta u} \tag{1.126}$$

where

$$H = T - U, \quad T = \frac{1}{2}\int_{-\infty}^{\infty}\left(\left|\frac{\partial}{\partial x}\right|^{s} u\right)^2 dx, \quad U = \frac{1}{6}\int_{-\infty}^{\infty} u^3 dx. \tag{1.127}$$

The equation (1.125) preserves the Hamiltonian H and the momentum $P = \frac{1}{2}\int_{-\infty}^{\infty} u^2 dx$.

Let us look for soliton solutions of (1.125) in the form of a stationary propagating wave $u = u(x - Vt)$. After one integration we obtain

$$-Vu + \frac{1}{2}u^2 - \left|\frac{\partial}{\partial x}\right|^{2s} u = 0. \tag{1.128}$$

This equation follows from the variational problem

$$\delta(H + VP) = 0, \tag{1.129}$$

demonstrating that soliton solutions are stationary points of the Hamiltonian for fixed momentum P. As far as the operator $|\partial/\partial x|^{2s}$, it is positive definite. Therefore solitons (as localized objects) can exist only for $V > 0$. Only in this case the linear operator $\hat{R}= V + |\partial/\partial x|^{2s}$ in the stationary equation (1.128) is positive definite and reversible. Otherwise the solution of (1.128) has oscillating asymptotics at infinity.

By multiplying (1.128) by $u/2$ and integrating with respect to x one can get the relation

$$\frac{3}{2}U - T = VP. \tag{1.130}$$

Let $u(x)$ be a solution of (1.128). Consider the functions $u(x, a) = a^{-1/2}u\,(x/a)$ depending on the scaling parameter a. Then

$$H = \frac{T}{a^{2s}} - \frac{U}{a^{1/2}} \tag{1.131}$$

(here U, T are calculated on the solution $u(x)$). For this kind of deformation the condition (1.127) now reads as $\partial H/\partial a|_{a=1} = 0$ or

$$\frac{1}{2}U - 2sT = 0, U = 4sT,$$

which in combination with (1.130) gives

$$T = \frac{VP}{6s-1}, \quad U = \frac{4sVP}{6s-1}, \quad H = \frac{1-4s}{6s-1}VP. \tag{1.132}$$

Here both functionals U and T are positive definite for $s > 1/6$. For the model (1.125) this defines the region of the soliton existence. For $s > 1/4$ the Hamiltonian on the soliton solutions is negative, $H_s < 0$. In this case it is possible to show that the stationary point $u(x)$ is not only the local but the global minimum of the functional H that, in accordance with the Lyapunov theorem, provides the soliton stability.

For $s < 1/4$ $H_s > 0$. Letting $a \to 0$ in (1.112) one can see that in this case H can be made arbitrary negatively large. Solitons in this region of the parameter V represent saddle points and one may expect that they are unstable.

For any s the equation (1.124) allows the self-similar substitution

$$u = (t_0 - t)^{1+1/2s} F\,(\xi)\,, \quad \xi = \frac{x}{(t_0 - t)^{1/(1+2s)}}. \tag{1.133}$$

For this family of self-similar solutions we have

$$P \simeq (t_0 - t)^{\frac{1-4s}{1+2s}}. \tag{1.134}$$

Hence one can see that a localized solution can exist only for $s > 1/4$, coinciding with the interval for soliton stability. The solution (1.133) at $0 < s < 1/4$ describes weak collapse leading to the formation of an integrable singularity: $u \approx 1/|x|^{2s}$. At $s = 1/4$ we have the regime of strong collapse corresponding to the critical NLSE considered in Section 6.
The results of this Section can be easily extended to a more general equation

$$\frac{\partial u}{\partial t} + u^p \frac{\partial u}{\partial x} - \frac{\partial}{\partial x} \left| \frac{\partial}{\partial x} \right|^{2s} u = 0. \tag{1.135}$$

We now consider reference [50] where for the critical KDV equation at $s = 1$ and $p = 2$ the corresponding theory for strong collapse was developed. In particular, it was shown that the asymptotic form of the collapsing distribution approaches the soliton form at the collapse time, and the absorbed energy into singularity corresponds to the soliton energy. For $p > 2$ ($s = 1$) collapse becomes weak. Stable solitons appear for $p < 2$ [49].

Very interesting coherent structures are described by the dissipative generalized KDV equation

$$\frac{\partial u}{\partial t} + u^p \frac{\partial u}{\partial x} + \frac{\partial}{\partial x} \left| \frac{\partial}{\partial x} \right|^{2s} u = \nu \frac{\partial^2 u}{\partial x^2}. \tag{1.136}$$

For $s > 1/4$ this equation describes the so-called collisionless shock waves discovered by R.Z. Sagdeev [51]. For the unstable case $s \leq 1/4$ the coherent structures have not been studied yet.

9 Singularities on a fluid surface

For sea surface waves, the wave breaking leads to an infinite second derivative of the surface profile (so that angles or cones appear on the surface). In this field, the first important results date back to the middle of the last century and belong to the famous Stokes [52]. Using the apparatus of complex analysis, Stokes discovered that the critical angle for the surface slope of stationary gravity waves for the deep water case was equal to 120°. For larger angles stationary gravity waves were assumed to be absent. Checking analyticity violation is the most sensitive tool for studying that set of collapses. Loss of analyticity of vortex sheets at the nonlinear stage of the Kelvin-Helmholz instability [53] is such an example. Various aspects of the singularity formation for vortex sheet motion have so far been studied in a number of papers, both numerically and analytically [53]-[58]. The paper [56] should be mentioned in particular, which provides a considerable numerical evidence of arising of the infinite surface curvature in a finite time. The root (in space) character of the arising singularity has also been checked in [56].

Below we present some recent results [59, 60] devoted to the free surface hydrodynamics of an ideal fluid. Adopting only the small slope approximation, in absence of both gravity and capillarity, this system was effectively examined. In particular, it was shown that for two-dimensional flows the velocity component v, tangent to the free surface, obeys the equation, formally coinciding with the Hopf equation (1.115),

$$\frac{\partial v^{\pm}}{\partial t} + v^{\pm}\frac{\partial v^{\pm}}{\partial x} = 0. \tag{1.137}$$

Here $v^{\pm}$ is analytical continuation of $v(x,t)$ to the upper $^{(+)}$ and lower $^{(-)}$ half-planes of the variable x. On the real axis $v = \frac{1}{2}(v^{+} + v^{-})$ and functions $v^{(\pm)}$ are complex conjugate. The free surface elevation $\eta(x,t)$ in this approximation ($|\nabla\eta| \ll 1$) is defined from integration of the equation

$$\frac{\partial \eta}{\partial t} = -\hat{H}\, v. \tag{1.138}$$

Here

$$(\hat{H}f)(x) = \frac{1}{\pi} V.P. \int_{-\infty}^{+\infty} \frac{f(x')}{x' - x}\, dx'.$$

is the Hilbert transform. Both equations for v and η are integrable. The integrability of these equations originates from the solution of the Laplace equation in the fluid bulk.

Autonomy of the equation for the tangent velocity component from elevation η is one of the main features of this system[2]. It admits, as for (1.115), the standard method of characteristics, but the analyticity requirement for functions $v^{\pm}$ leads in comparison with solution (1.116) to some changes in the form of general solution. Omitting all details of the general solution analysis (see [60, 59]), we present here only the main results.

The formation of singularities on the free surface for small angle approximation can be considered as the process of the wave breaking in the complex plane to which the solution can be extended. This results in the motion of both branch points of the analytical continuation of the velocity potential and singular points of the analytical extension of the surface elevation. When for the first time the most "rapid" singular point reaches the

[2]Equation (1.137), after separation of imaginary and real parts, transforms into a system of the gas dynamic type with negative pressure,

$$u_t + (uv)_x = 0,$$

$$v_t + vv_x = \frac{1}{2}(u^2)_x$$

where u is normal component of the velocity. It is interesting that this system also follows from the quasi-classical limit of the fifth NLSE.

real axis it just indicates the appearance of the singularity. Respectively three kinds of singularities are possible. For the first kind at the touching moment, the tangent velocity on the surface has an infinite first derivative and simultaneously the second space derivative of the surface coordinate $z = \eta(x,t)$, i.e. η_{xx}, also tends to infinity. These are weak singularities of the root character ($\eta_{xx} \sim \mid x \mid^{-1/2}$). This kind of singularities turns out to be consistent with an assumption about small surface angles. It is shown that the interaction of two movable branch points of the tangent velocity can lead under some definite conditions to the formation of the second type of singularities - wedges on the surface shape. Close to the collapse time the self-similar solution for such singularities is compatible with the complete system of equations describing arbitrary angle values. The third type is caused by the initial analytical properties of $\eta_0(x)$, resulting in the formation of strong singular surface profile.

As was shown in [61], the equation of motion for free surface hydrodynamics with finite depth in the absence of capillarity can also be integrated effectively in the small angle approximation. Of course, the root singularities, as well as all others have the same asymptotic behavior as for the deep water case. Another interesting effect is connected with the possibility to integrate the free surface hydrodynamics in the limit of large surface gradients. In this case, as it was shown in [62], the equation can be reduced to the so-called Laplacian growth equation (LGE)[3] which allows application of the pole decomposition. The latter means that a system of equations has an exact solution in the form of finite sum of poles, residues of which are constants and pole positions (in complex plane) obey a closed dynamical system of ordinary differential equations. In the case of the LGE this dynamical system allows complete integration. Similarly, the solution can be written in an implicit form (for more details see [65, 66, 62, 61]).

10 Solitons and collapses in the generalized KP equation

Let us take into account the diffraction term, that corresponds to consideration of the dependence on perpendicular coordinates in the GKDV equation. Assuming maximum symmetry in the perpendicular plane we

[3] At first this equation was derived in 1945 by Polubarinova-Kochina and Galin [63, 64] for boundary flows in porous media. Later it became clear that this equation is applicable for description of the boundary motion for phase transition of the first kind.

will study the following version of the equation (1.109)

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}-\frac{\partial}{\partial x}\left|\frac{\partial}{\partial x}\right|^{2s}u\right)=\frac{\alpha}{r^{d-1}}\frac{\partial}{\partial r}r^{d-1}\frac{\partial}{\partial r}u. \tag{1.139}$$

Here $\alpha=\pm1$, d is the dimension of the perpendicular plane, r is the radius in this plane.

The equation (1.139) can be written in the Hamiltonian form (1.126)

$$H=T-U+W \tag{1.140}$$

where

$$T = \frac{1}{2}\int\left(\left|\frac{\partial}{\partial x}\right|^{s}u\right)^{2}dxdr,\quad U=\frac{1}{6}\int u^{3}dxdr, \tag{1.141}$$

$$2W = \frac{\alpha}{2}\int(\nabla_{\perp}w)^{2}dxdr\quad w=\int_{-\infty}^{x}udx. \tag{1.142}$$

This equation conserves the Hamiltonian and the momentum

$$P=1/2\int u^{2}dxdr. \tag{1.143}$$

Soliton solutions of the equation (1.139) have the form

$$u=u_{s}(x-Vt,r) \tag{1.144}$$

with the boundary condition $u_s\to0$ in all directions at infinity, $\sqrt{x^2+r^2}\to\infty$, and requiring finite momentum $P<\infty$.

Solitons are solutions of the stationary KP equation

$$\hat{R}w=\left[\left(V+\left|\frac{\partial}{\partial x}\right|^{2s}\right)\frac{\partial^{2}}{\partial x^{2}}+\frac{\alpha}{r^{d-1}}\frac{\partial}{\partial r}r^{d-1}\frac{\partial}{\partial r}\right]w=\frac{\partial}{\partial x}\left(\frac{u^{2}}{2}\right), \tag{1.145}$$

which can be presented as the variational problem (1.129) where H and P are given by (1.141) and (1.143), respectively. Soliton solutions exist only for $\alpha=+1$ and positive V. For all other cases the operator $\hat{R}$ is not sign definite and it cannot provide a vanishing soliton solution at $\sqrt{x^2+r^2}\to\infty$ (for more details see [67, 72, 68, 13]).

Multiplying now (1.145) by $w/2$ and integrating with respect to x and r, one obtain

$$-VP-T-W+\frac{3}{2}U=0. \tag{1.146}$$

Let us take a trial function for the variational problem (1.129) with H and P given by (1.141) and (1.143) in the form, retaining the total momentum P,

$$u(x,r,a,b)=a^{-1/2}b^{-d/2}u_{s}\left(x/a,r/b\right). \tag{1.147}$$

As a result, the Hamiltonian becomes the function of two scaling parameters a and b

$$H = \frac{T}{a^{2s}} + \frac{a^2}{b^2}W - \frac{U}{a^{1/2}b^{d/2}}. \tag{1.148}$$

By inserting (1.147) into (1.129) we get

$$\left.\frac{\partial H}{\partial a}\right|_{a=b=1} = 0, \quad \left.\frac{\partial H}{\partial b}\right|_{a=b=1} = 0. \tag{1.149}$$

This yields

$$-2sT + 2W + \frac{1}{2}U = 0, \quad -2W + \frac{d}{2}U = 0. \tag{1.150}$$

Solving now the linear system (1.146) and (1.150) one can get

$$\begin{aligned} T &= \frac{d+1}{6s-1-d(1+s)}VP_s,\ U = \frac{4s}{6s-1-d(1+s)}VP_s \\ W &= \frac{sd}{6s-1-d(1+s)}VP_s,\ H_s = \frac{d(1+s)+1-4s}{6s-1-d(1+s)}VP_s. \end{aligned} \tag{1.151}$$

These formulas become identical to (1.132) at $d = 0$. From (1.151) one can see that on the soliton solutions T and W must have the same (positive) sign. Hence it follows that multidimensional solitons exist only if $\alpha > 0$. In other words, multidimensional solitons exist only for the KPI equation and its generalization. This conclusion ($\alpha > 0, V > 0$) corresponds completely to the requirement of sign-definiteness of the operator R in the equation (1.145). In the following we shall assume $\alpha = 1$.

From relations (1.151), we get the necessary conditions for existence of solitons

$$s > \frac{1+d}{6-d}. \tag{1.152}$$

The sufficient condition for soliton stability is again $H_s < 0$, implying

$$s > \frac{d+1}{4-d}. \tag{1.153}$$

In the interval

$$\frac{1+d}{6-d} < s \leq \frac{d+1}{4-d}, \tag{1.154}$$

solitons are unstable.

Let us consider now the most important physical examples of equation (1.139). For $d = 1$ and $s = 1$ (1.139) is nothing more than the classical (2D) KPI equation (1.112) (where it is necessary to change $u \to -u$ and $t \to -t$). We see that the condition (1.153) is satisfied now, and the soliton is stable [69]. The soliton in this case has the form of a lump and can be found analytically [70]. Another example arises if $s = 1$ and $d = 2$. In this case (1.139) is the KPI equation for a three-dimensional media. Now the

criterion (1.154) is satisfied and the soliton exists but is unstable. This fact was established in the paper [69]. Development of the soliton instability results into collapse of acoustic waves confirmed by numerical experiments [71, 72].

Consider the GKPI equation in the three-dimensional case with power nonlinearity

$$\frac{\partial}{\partial x}[u_t + \gamma(\gamma-1)u^{\gamma-2}u_x] = \Delta_\perp u \tag{1.155}$$

for which the Hamiltonian is of the form

$$H = \frac{1}{2}\int (u_x)^2 d\boldsymbol{r} + \frac{1}{2}\int (\nabla_\perp w)^2 d\boldsymbol{r} - \int u^\gamma d\boldsymbol{r}. \tag{1.156}$$

This equation generalizes the KDV equation (1.135) with power nonlinearity to many dimensions. In particular, the classical KPI equation corresponds to $\gamma = 3$. The case $\gamma = 4$ is possible if for some physical reason the three-wave matrix element vanishes. For instance, such a situation takes place for special angles of propagation of acoustic-type waves in a ferromagnet [75]. In this case, as it was shown at first in this paper it is possible to write down the analog of the virial theorem.

Consider the quantity $I = \int r_\perp^2 u^2 d\boldsymbol{r}$ which, because of conservation of the x component of the momentum, $P_x = \frac{1}{2}\int u^2 d\boldsymbol{r}$, has the meaning of mean transverse size of the wave distribution. Let us find the first derivative of I with respect to time. By means of (1.155) we have

$$I_t = -4\int u(\boldsymbol{r}_\perp \nabla_\perp) w d\boldsymbol{r}.$$

Calculating now the second derivative of I one can get

$$I_{tt} = 4\left[2\int (\nabla_\perp w)^2 d\boldsymbol{r} - d(\gamma-2)\int u^\gamma d\boldsymbol{r}\right].$$

By use of (1.156) the r. h. s. of this equation can be rewritten

$$I_{tt} = 4\left[4H - 2\int (u_x)^2 d\boldsymbol{r} + \beta\int u^\gamma d\boldsymbol{r}\right], \tag{1.157}$$

where $\beta = 4 - d(\gamma-2)$. At $d = 2$ (the 3D case) and $\gamma = 4$ the coefficient $\beta = 0$. In this case from the equation (1.157) one can get the following inequality [75]

$$I_{tt} < 16H. \tag{1.158}$$

Hence we have the same sufficient condition $H < 0$, as for the NLSE. For the classical 3D KPI equation ($d = 2$, $\gamma = 3$) the coefficient $\beta = 4 - 2(\gamma-2) > 0$ and in the virial identity (1.157) the two last terms have different signs and therefore, even for $H < 0$, it is difficult to get a certain

answer about the sign of the r. h. s. of (1.157) and that is so, despite the unboundedness of the Hamiltonian from below. But if the Hamiltonian of some region Ω is negative, then, following to the arguments analogous to section 3, it is possible to show that radiation of waves from this area promotes collapse. Radiation reduces the Hamiltonian of the cavity Ω so that H_Ω becomes more negative. Simultaneously, due to the unboundedness of the Hamiltonian, the maximal value of the wave amplitude into the cavity will increase, and this process continues up to the singularity formation [72]. At the moment there are no analytical arguments whether the collapse time (for $d = 2$ and $\gamma = 3$) is finite or infinite. Meanwhile, the numerical experiments performed in [71, 72] indicate that this time is finite.

One more physical example is $s = 1/2$, $d = 1$. We have now the following equation

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - \frac{\partial^2}{\partial x^2}\,\hat{H}\,u\right) = \frac{\partial^2 u}{\partial y^2}. \tag{1.159}$$

This equation describes two-dimensional Tolman-Schlichting waves in the laminar boundary layer. In this case again the condition (1.154) is fulfilled and 2D solitons are unstable.

In the case of soliton instability

$$s < \frac{d+1}{4-d} \tag{1.160}$$

the equation (1.139) describes weak collapse. The corresponding self-similar solution is

$$u = (t_0 - t)^{-2s/(2s+1)} F\left(\frac{x}{(t_0-t)^{1/(2s+1)}}, \frac{r}{(t_0-t)^{(s+1)/(2s+1)}}\right). \tag{1.161}$$

Collapse leads to the formation of an integrable singularity for $t \to t_0$

$$u = \frac{1}{x^{2s}}\varphi\left(\frac{r}{x^{s+1}}\right). \tag{1.162}$$

More detailed structures of collapses as well as the role of dissipation in arresting collapse and the formation of black holes have not properly been studied so far.

11 Self-focusing in the boundary layer

Now we demonstrate how the tools considered above work for the two-dimensional model,

$$u_t = \frac{\partial}{\partial x}\hat{k}u - 6uu_x = \frac{\partial}{\partial x}\frac{\delta H}{\delta u} \tag{1.163}$$

where the Hamiltonian becomes

$$2H = \int \left(\frac{1}{2} u \hat{k} u - u^3 \right) d\boldsymbol{r} \equiv \frac{1}{2} I_1 - I_2.$$

Here $\hat{k}$ is the integral operator, its Fourier transform is the modulus $|\boldsymbol{k}| = (k_x^2 + k_y^2)^{1/2}$. This equation describes low-frequency oscillations of the boundary layer within the high Reynolds number, $Re \gg 1$, with the mean velocity profile $\boldsymbol{v} = \hat{x}\, U(z)$ $(0 \leq z < \infty)$. The function $U(z)$ is assumed to be a monotonically growing function with a constant value at the infinity. The dimensionless amplitude u is connected with the velocity fluctuations along mean flow by means of the relation

$$\delta v_x \approx -6huU'(z), \tag{1.164}$$

where $h = U(0)/U'(0)$ is a thickness of the boundary layer.

Equation (1.163) was derived first by V.I. Shrira [77]. It represents the two-dimensional generalization of the well-known Benjamin-Ono (BO) equation describing long waves in stratified liquids. One should note that for this problem this equation was also derived in the 1D case first in the papers [81], taking into account the small viscosity.

The simplest soliton in this model are of the form $u = u_s(x - Vt, y)$. Their shape is defined from the equation

$$-Vu_s - \hat{k}u_s + 3u_s^2 = 0. \tag{1.165}$$

For the 1D case the solution of this equation can be found explicitly

$$u_s = \frac{2V}{3(x^2V^2 + 1)} \quad (V > 0). \tag{1.166}$$

In the 2D case the model has a ground state soliton that is a cylindrically symmetric solution without nodes. Such a solution was found in [82] numerically.

It is very important to note that the velocities of the 2D ground state solitons as well as their amplitudes are positive quantities. In physical variables both the 1D and 2D solitons, upon applying the relation (1.164), move in the upstream direction and have negative amplitudes. The latter means that in the real hydrodynamic system solitons look like holes in the mean velocity profile, therefore they move slower than the main flow. When the soliton amplitude grows, the soliton velocity decreases and *vice versa*. This physical reasoning suggests a possibility for the appearance of the wave collapse in this system and the instability of 1D solitons with respect to two-dimensional perturbations as well (for details, see [78, 79]). This instability is analogous to the Kadomtsev-Petviashvili instability [46, 47]. In the two-dimensional case, as was shown in [80], it is possible to develop a quasi-classical nonlinear theory of this instability taking a solution in the

form of the 1D soliton (1.166) with slowly varying parameters depending on y and t.

The soliton (1.163) in this model, as many others, represents a stationary point of H for fixed x-projection of the momentum $P = 1/2 \int u^2 d\boldsymbol{r}$

$$\frac{\delta}{\delta u}(H + VP_x) = 0.$$

A minimum of H (for fixed P) is found in the one-dimensional case. It follows from the estimates analogous to (1.56)

$$\int u^3 d\boldsymbol{r} \leq C \left(\int u\hat{k}u d\boldsymbol{r} \right)^{D/2} \left(\int u^2 d\boldsymbol{r} \right)^{(3-D)/2},$$

with the best constant C attaining its value at the ground state soliton

$$C = I_{2s} I_{1s}^{-D/2} \left(2P_s\right)^{(D-3)/2}.$$

Hence it is easy to get the estimate

$$H \geq H_s + 1/2(I_1^{1/2} - I_{1s}^{1/2})^2, \tag{1.167}$$

which becomes precise on the 1D soliton. Thus, in the one-dimensional case the soliton is proved to be stable with respect to 1D perturbations, but optionally against small ones [83, 78, 79].

In 2D case this system demonstrates the critical behavior like the 2D cubic NLSE. In particular, the Hamiltonian is bounded from below by zero,

$$H \geq \frac{1}{2} \left[1 - \left(\frac{P}{P_s} \right)^{1/2} \right] \int u\hat{k}u d\boldsymbol{r},$$

if the total perturbation power does not exceed the critical value equal to P_s. If initially the Hamiltonian is negative, $H < 0$, then it is unbounded from below. The latter follows from the scaling transformation, retaining P,

$$u_s(\boldsymbol{r}) \to \frac{1}{a^{d/2}} u_s(\boldsymbol{r}/a). \tag{1.168}$$

Under these transformations H becomes a function of the scaling parameter a,

$$H(a) = \frac{I_{1s}}{2a} - \frac{I_{2s}}{a^{d/2}}, \tag{1.169}$$

and is unbounded from below as $a \to 0$, starting from $d \geq 2$. It is enough to state that in this case the formation of a singularity is possible due to small amplitude waves radiated from the region with negative Hamiltonian. In this case the inequality corresponding to (1.63) is of the form

$$\max_{x\in\Omega} u \le \frac{|H_\Omega|}{2P_\Omega}. \tag{1.170}$$

Numerical integration of the equation (1.163) confirmed the main theoretical predictions.

For all initial conditions with $P > P_s$ and $H < 0$, the significant growth of amplitude was observed at the peak moving with increasing acceleration. The temporal behavior of the peak velocity and of the peak amplitude are familiar, indicating that the collapse is of self-similar nature. Upon approaching the singularity the peak anisotropy vanishes, and the peak distribution becomes nearly symmetric.

For the initial conditions with $P < P_s$ ($H > 0$) a slow evolution took place: the distribution of u slowly decayed near the maximum. The spectrum evolution for $P_x < P_{x,cr}$ demonstrated the energy transfer to the long-wave region, which on a qualitative level is in agreement with the estimate (1.170).

In conclusion of this section, we would like to point out several interesting experiments [76], summarizing the results of many years of experimental studies of the onset of the coherent structures in the boundary layer of the blowing plate by the mechanical vibrating system near the edge of the plate (see, also [84], [85]). According to these experimental data, one-dimensional solitons are exited at the initial stage, later (for larger distances from the plate edge) one-dimensional solitons demonstrate their instability which results "in the formation of thorns", i.e., the localized three-dimensional coherent structures. Self-focusing of the above structures is observed at longer distances. A later stage of the development of the thorns-solitons leads to the formation of vortices and to their eventual separation.

The above theory and numerical experiments as well explain all these experimental observations, but not the formation of vortices, for which equation (1.163) becomes inapplicable. The threshold character of the wave collapse in the boundary layer described by Eq. (1.163) also explains why in many other experimental studies in the boundary layer such bright phenomena as self-focusing of solitons and collapse have not been observed or have not been distinguished on the background of the turbulent noise. The collapse is possible to observe starting from the finite energy of the pulse as it was in experiments [76]. If the pulse amplitude is small enough then this phenomena is absent.

Acknowledgments: This work was supported by the Russian Foundation of Basic Research under Grant no. 97-01-00093, by the INTAS under Grant no. 96-0413 and by the NATO Linkage Grant OUTR.LG 970583.

12 References

[1] E.A. Kuznetsov, A.M. Rubenchik and V.E. Zakharov, Phys. Rep. **142**, 103 (1986).

[2] J. Juul Rasmussen and K. Rypdal. Physica Scripta **33**, 481 (1986).

[3] V.E. Zakharov and E.A. Kuznetsov, Uspekhi Fizicheskihk Nauk (Physics Uspekhi) **167**, 1137 (1997).

[4] N. Bloembergen, *Nonlinear Optics*, Ed. W.A. Benjamin, Reading, Mass. (1965).

[5] V.E. Zakharov, V.S. L'vov and G.E. Falkovich, *Kolmogorov Spectra of Turbulence I*, Springer-Verlag, Berlin-Heidelberg-New York (1992).

[6] V.E. Zakharov, ZhETF **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].

[7] V.E. Zakharov, In: *"Handbook of Plasma Physics"*, Vol.2, Basic Plasma Physics II, A. Galeev and R. Sudan, Eds. North-Holland (984) p. 81.

[8] V.E. Zakharov and E.A. Kuznetsov, Zh. Eksp. Teor. Fiz. **91**, 1310 (1986) [Sov. Phys. JETP **64**, 773 (1986)].

[9] O.A. Ladyzhenskaya, *Mathematical Problems of the Dynamics of a Viscous Incompressible Fluids*, Fizmatgiz, Moscow (1961) (in Russian).

[10] L. Nirenberg, Ann. Sci. Norm. Sup. Pisa, **20**, No. 4, 73 (1966).

[11] E.A. Kuznetsov, Chaos **6**, 381 (1996).

[12] M.I. Weinstein, Commun. Math. Phys. **87**, 567 (1983).

[13] V.E. Zakharov and E.A. Kuznetsov, ZhETF (JETP) **113** 1892 (1998).

[14] R.J. Chiao, F. Gardmire and C.H. Townes, Phys. Rev. Lett. **13**, 479 (1964).

[15] N.G. Vakhitov and A.A. Kolokolov, Radiofizika, **16**, 1020 (1973).

[16] V.E. Zakharov and A.B. Shabat, Sov.Phys. JETP **34**, 62 (1972).

[17] S.N. Vlasov, V.A. Petrishchev and V.I. Talanov, Izv.Vys. Uchebn. Zaved. Radiofizika **14**, 1353 (1971) [Radiophys. Quantum Electron. **14**, 1062 (1974)].

[18] S.K. Turitsyn, Phys. Rev. E **47**, R13 (1993).

[19] E.A. Kuznetsov, J.J. Rasmussen, K. Rypdal and S.K. Turitsyn, Physica ***D*87**, 273 (1995).

[20] V.E. Zakharov and L.N. Shchur, Sov. Phys. JETP **54**, 1064 (1981).

[21] N.E. Kosmatov, and V.F. Shwets, and V.E. Zakharov, Physica ***D*52**, 210 (1991).

[22] S. Dyachenko, A.C. Newell, A. Pushkarev and V.E. Zakharov, Physica ***D*57**, 96 (1992).

[23] V.E. Zakharov, N.E. Kosmatov and V.F. Shwets, Pis'ma ZhETF (JETP Letters) **49**, 431 (1989).

[24] V.E. Zakharov and V.S. Synakh, Sov. Phys. JETP **41**, 465 (1975).

[25] V.E. Zakharov, Sov. Phys. Uspekhi **31**, 672 (1988).

[26] V.E. Zakharov and V.F. Shwets, Pis'ma ZhETF (JETP Letters) **47**, 275 (1988).

[27] G.M. Fraiman, Sov. JETP **88**, 390 (1985).

[28] A.I. Smirnov and G.M. Fraiman, Physica ***D*52**, 2 (1991).

[29] L. Bergé and D. Pesme, Phys. Rev E **48**, R684 (1993).

[30] L. Bergé and D. Pesme, Phys. Lett A **166**, 116 (1992).

[31] S.N. Vlasov, V.A. Petrishchev and V.I. Talanov, in: Nonlinear and Turbulent Processes in Physics, Proceedings of the Kiev Workshop, Kiev, vol. 2, 210 (1988).

[32] V.M. Malkin, Pis'ma v ZhETF (JETP Letters) **48**, 603 (1988).

[33] K. Landman, G.C. Papanicolau, C. Sulem and P.L. Sulem, Phys. Rev. A **34**, 1200 (1988).

[34] B.J. Le Mesurier, G.C. Papanicolau, C. Sulem and P.L. Sulem, Physica ***D*31**, 78 (1988).

[35] L. Bergé, E.A. Kuznetsov and J.J. Rasmussen, Phys. Rev E **53**, R1340 (1996).

[36] N.A. Zharova, A.G. Litvak, T.A. Petrova, A.M. Sergeev and A.D. Yunakovsky, JETP Lett **44**, 13 (1986).

[37] L. Bergé, E.A. Kuznetsov, J.J. Rasmussen, E.G. Shapiro, and S.K. Turitsyn, JOSA B **13**, 1879 (1996).

[38] V.V. Krasnosel'skikh and V.I. Sotnikov, Fizika Plazmy (Soviet Plasma Physics) **48**, 603 (1988).

[39] E.A. Kuznetsov and M.M. Scoric, Phys. Rev. A **38**, 1422 (1988).

[40] E.A. Kuznetsov and S.K. Turitsyn, Fizika Plazmy (Sov. Plasma Phys.), **16**, 901 (1990).

[41] B.I. Sturman, ZhETF **71**, 613 (1976) [Sov. Phys. JETP **44**, 322 (1976)]; S.L.Musher, B.I.Sturman and A.M.Runbenchik, Plasma Phys. **20**, 1131 (1978).

[42] S.L. Musher and B.I. Sturman, Pisma ZhETF, **22**, 537 (1975) [JETP Lett. **22**, 265 (1975)].

[43] J. Bonnell, P. Kintner, J.E. Wahlund, K.Lynch and R.Arnoldy, Geophys. Res. Lett. **23**, 3297 (1996).

[44] J.L. Pincon, P.M. Kintner, P.W. Schuck, and C.E. Seyler, *Observations and analysis of lower hybrid solitary structures as rotating eigenmodes*, Preprint, LPCE/CNRS, Orleans, France (1996).

[45] E.A. Kuznetsov, unpublished (1997).

[46] B.B. Kadomtsev and V.I. Petviashvili, Sov.Phys.Dokl. **15** 539 (1970).

[47] B.B. Kadomtsev. Collective Phenomena in Plasma. Moscow, Nauka, (1976).

[48] V.L. Gurevich, *Kinetics of Phonon Systems,* Nauka, Moscow (1980) (in Russian).

[49] E.A. Kuznetsov, Physics Letters ***A*101**, 314 (1984).

[50] D.E. Pelinovsky and R.H.J. Grimshaw, *An asymptotic approach to solitary wave instability and critical collapse in long-wave KDV-type evolution equations*, Preprint of Monash University, No. 95/35.

[51] R.Z. Sagdeev, *Voprosy teorii plazmy (Problems of plasma theory)*, Atomizdat, Moscow, vol.4 (1964).

[52] G.G. Stokes, Mathematical Physical Papers, vol.**1** p. 225, Cambridge University Press (1880).

[53] D.W. Moore, Proc. Roy. Soc. ***A*365**, 105 (1979).

[54] G.R. Baker, D.I. Meiron and S.A. Orzag, J. Fluid Mech. **123**, 477 (1982).

[55] M.S. Longuet-Higgins, In: Nonlinear Waves, ed. L.Denath (Cambridge Univ. Press 1983) p.1.

[56] M. Shelley, J. Fluid Mech. **244**, 493 (1992).

[57] R.E. Caflish, O.F. Orellana and M. Siegel, SIAM J. Appl.Math. **50**, 1517 (1990).

[58] G. Baker, R. Caflish and M. Siegel, J.Fluid Mech. **252**, 51 (1993).

[59] E.A. Kuznetsov, M.D. Spector and V.E. Zakharov, Phys. Lett. **182*A*** 387 (1993).

[60] E.A. Kuznetsov, M.D. Spector and V.E. Zakharov, Phys. Rev. E **49** 1283 (1994).

[61] A.I. Dyachenko, E.A. Kuznetsov, and V.E. Zakharov, Fizika Plazmy (Russian Plasma Physics) **22**, 916 (1996).

[62] A.I. Dyachenko and V.E. Zakharov, Phys. Lett. A **221**, 80 (1996).

[63] P.Ya. Polubarinova-Kochina, Dokl. Akad. Nauk SSSR **47**, 254 (1945); Prikl. Matem. Mech. **164**, 383 (1945).

[64] L.A. Galin. Dokl. Akad. Nauk SSSR **47**, 246 (1945).

[65] M.B. Mineev and S.P. Dawson, Phys. Rev. E **50**, 24 (1994).

[66] S.P. Dawson and M.B. Mineev, Physica ***D*73**, 373 (1994).

[67] V.I. Petviashvili, in: *Nelineinye volny (Nonlinear waves)*, ed. A.V.Gaponov-Grekhov, Moscow, Nauka, p. 5 (1979); Fizika plazmy (Sov. Plasma Physics) **2**, 469 (1976).

[68] J. Nycander, Chaos **4**, 253 (1994).

[69] E.A. Kuznetsov and S.K. Turitsyn, ZhETF **82**, 1457 (1982) [Sov.Phys. JETP **55**, 844 (1982)].

[70] L.A. Bordag, A.R. Its, A.V. Matveev, S.V.Manakov and V.E.Zakharov, Phys. Lett. **63*A***, 205 (1979).

[71] E.A. Kuznetsov, S.L. Musher and A.V. Shafarenko, Pis'ma ZhETF **37**, 204 (1983) [JETP Lett. **37**, 241 (1983)].

[72] E.A. Kuznetsov and S.L. Musher, ZhETF **91**, 1605 (1986) [Sov.Phys. JETP **64**, 947 (1986)].

[73] V.E. Zakharov and E.A. Kuznetsov, Zh. Eksp. Teor. Fiz. **66**, 594 (1974) [Sov. Phys. JETP **39**, 285 (1974)].

[74] K. Rypdal, J. Juul Rasmussen and K. Thomsen, Physica ***D*16**, 339 (1984).

[75] S.K. Turitsyn and G.E. Falkovich, Zh. Eksp. Teor. Fiz. **89**, 258 (1985) [Sov.Phys. JETP **62**, 146 (1985)].

[76] Yu.S. Kachanov and O.S. Ryzhov, Sib. Fiz.-Tekh.Journ. N1, 34, (1992) (in Russian); Y.S.Kachanov, O.S.Ryzhov and F.T.Smith, J. Fluid Mech. **251**, 273 (1993).

[77] V.I. Shrira, Doklady AN SSSR **308**, 732 (1989) (in Russian).

[78] A.I. Dyachenko and E.A. Kuznetsov, Pis'ma ZHETF, **59** 103 (1994) [JETP Lett. **39**, 108 (1994)].

[79] A.I. Dyachenko and E.A. Kuznetsov, Physica ***D*87**, 301 (1995).

[80] D.E. Pelinovsky and V.I. Shrira, Phys. Lett. A **206**, 195 (1995).

[81] V.I. Zhuk and O.S. Ryzhov, Doklady AN SSSR **263**, 56 (1982) (see also: O.S.Ryzhov, Zh.Vych.Mat.i Mat.Fiz. **29**, 1804 (1990)) (both in Russian).

[82] A.A. Abramyan, Ya.A. Stepanyants and V.I. Shrira, Doklady AN SSSR **327**, 460 (1992).

[83] M. Weinstein, Comm. Partial Diff. Eq. **12** (10), 1133 (1978).

[84] V.N. Borodulin and Yu.S. Kachanov, Izv. Sib. Branch USSR Ac.Sci. N18, 65 (1988) (in Russian).

[85] Yu.S. Kachanov, V.V. Kozlov, V.Ya. Levchenko and P.M. Ramazanov, Sib.Branch USSR Ac.Sci. N2, 124 (1989) (in Russian).

2

Perturbation Theories for Nonlinear Waves

Lev Ostrovsky
Konstantin Gorshkov

ABSTRACT Some ideas and theories developed since 1960s to describe nonlinear waves with slowly varying parameters (modulated waves) are outlined. These theories are associated with different versions of the asymptotic perturbation method. In this framework, both quasi-periodic and solitary waves (solitons) can be treated. A scheme for reduction of a quasi-hyperbolic system to one or more evolution equations is also presented. Some challenges for the theory are briefly discussed.

1 Introduction

Among the most remarkable achievements of nonlinear wave theory by the eve of the 21st century, many of us will cite the development of exact methods such as the inverse scattering method in the theory of integrable nonlinear wave equations. At the same time, among the most effective tools for solving practical problems (apart from direct numerical simulations), the outstanding role of approximate analytical methods will probably be appreciated as well.

Perturbation methods in mechanics had been developed in the 19th century for astronomical applications. Later, they occupied an outstanding place in quantum physics, to mention only two very important applications. These methods were first elaborated for the cases when small perturbations in the equations or initial conditions would result in small changes in the solutions. However, in many cases, small perturbations may change the solutions rather strongly, and the smallness of the former is reflected only in the slowness of the latter's variation compared to the characteristic time scale of the process. An adequate tool for solving such problems is the asymptotic perturbation theory, which constructs a series in a small parameter that may not converge, but its lower-order terms turn out to be the closer to the exact solution, the smaller is the expansion parameter. The main term of such an expansion differs from the unperturbed solution in that it contains slowly varying parameters; their variation can be

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 47−65, 2000.

found from "compatibility," or "orthogonality," conditions which secure the finiteness of the higher-order perturbations.

For ordinary differential equations, these methods have been most thoroughly elaborated by Bogolubov and Mitropolsky [1]. The idea of the method can be demonstrated at an example of a simple anharmonic oscillator with damping

$$\frac{d^2x}{dt^2} + \omega_0^2 x = -\varepsilon\nu\omega_0^2 x^3 - \varepsilon a\frac{dx}{dt}, \tag{2.1}$$

where ε is a small parameter. At $\varepsilon = 0$ the immediate solution is $x = X = A\sin(\omega_0 t+\psi)$. For $\varepsilon \neq 0$, seeking a solution in the form $x = X+\varepsilon x_1+\varepsilon^2 x_2...$, and substituting into (2.1), we readily obtain $x_1 = (3\nu\omega_0 A^3/8)t\cos(\omega_0 t + \psi) + (aA/2)t\sin(\omega_0 t + \psi)$, i.e., a secularly increasing solution. To get rid of this growth, it is sufficient to represent the zero-order solution in the form $X = A(\varepsilon t)\sin(\omega t + \psi)$ with $\omega = \omega_0 + \varepsilon\omega_1 + ...$ After that, it is easy to see that the basic equation will be satisfied to the first approximation if $A = A_0 e^{-at/2}$, and $\omega_1 = 3\nu A^2/8$. The first-order perturbation remains finite: $x_1 = -(\nu A^3/32)\sin 3(\omega t + \psi)$. (The dissipative term yields only a second-order variation of oscillation frequency.)

In this elementary example the main idea of the asymptotic theory is readily seen: a small perturbation in governing equations results in finite deviations in the solution of the unperturbed one due to *resonance* action of the perturbation in the basic equation. Fortunately, however, due to its resonance character, this deviation has the same structure as the basic solution, and it can be incorporated into the latter by letting its parameters (amplitude, phase) vary slowly in time, thus keeping the rest of the perturbations small (nonresonant).

Another important class of systems is associated with fast-relaxing systems when the small parameter appears as a factor in highest-derivative terms. For example, Eq. (2.1) can have a small factor ε in front of the first term on its left side instead of those on the right. In this case, the motion can be divided into the "slow" stage when the first term can be neglected, and the "fast" one when the zero-derivative terms are negligible. The corresponding oscillations can then be described by using matched asymptotic expansions; they acquire a characteristic "sawtooth" shape [1, 40, 41].

There exists a huge variety of nonlinear ODEs that are important for physical and technical applications and have been effectively solved with the help of the asymptotic perturbation theory. Among the historic achievements, we only mention the theory of the van der Pol oscillator, for which the perturbation theory gives an analytical expression for an attractor: limit cycle in the phase plane of the system, and which has played an outstanding role as a realistic model of electronic generators (see, e.g., [2]). Also, many early models of a laser were based on ODEs and their solutions with a slowly-varying amplitudes.

2 Modulated waves

The word "modulated" came from radio engineering and signal theory. In application to waves, it was first used by Ryutov in 1940 [3] for quasiharmonic waves of the form

$$u = A(\mathbf{r},t)e^{i\theta(\mathbf{r},t)} + c.c., \tag{2.2}$$

where the amplitude, A, instantaneous frequency, $\omega = \theta_t$, and wave vector, $\mathbf{k} = -\nabla\theta$, are slowly varying in time and space. In other words, they depend on "slow" variables $\tau = \varepsilon t$ and $\rho = \varepsilon r$ (again, ε is a small parameter). In that work, Ryutov considered mostly classical geometrical optics of linear inhomogeneous media.

For a nonlinear case, a systematic consideration of modulated wave propagation started in 1960s with, as it seems, the works by Ostrovsky and especially Whitham. The former paper [4] deals with a plane electromagnetic wave of the form (2.2). Substitution into Maxwell equations with nonlinear equations of state yields the solution for the wave amplitude in the form of a simple wave well known in fluid mechanics

$$|A| = F\left[t - x/c_g\left(|A|\right)\right], \tag{2.3}$$

where c_g is the nonlinear group velocity, depending on the wave amplitude; e.g., in a nonlinear dielectric, $c_g^2 = dD_0/dE_0$, D_0 and E_0 being the amplitudes of electric induction and electric field, respectively. The wave phase and frequency become modulated due to amplitude modulation, also. It was also stated that, as a result of breaking the "envelope wave" (2.3) , an *envelope shock* can be formed as a "step" in the amplitude, provided the medium has an inertia (such as relaxation of the Kerr-effect in liquids), and a boundary condition at the shock was written. The results obtained in [4] are valid when the wave dispersion in the vicinity of carrier frequency is weak in comparison with nonlinearity (as in nonlinear optics of very short and intensive laser pulses).

A very general analysis of the problem has been performed by Whitham [5, 6]. He considered an arbitrary system described by a Lagrangian $L(q, q_t, q_x)$, where q is a canonical variable or a vector of variables. Representing again these variables in the form (2.2) , substituting into the expression for L, and averaging over the wave period, one obtains an averaged Lagrangian, $L(A, \omega, k)$, where again, $\omega = \theta_t$ and $k = -\theta_x$. Considering A and θ as new "canonical" variables, Whitham introduced the "*averaged variational principle*" to obtain variational equations

$$\frac{\partial L}{\partial A} = 0, \tag{2.4}$$

and

$$\frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \omega}\right) - \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial k}\right) = 0. \tag{2.5}$$

The first of these equations represents a dispersion relation $\omega = \omega(k, |A|)$; in a linear case, when $L = g(\omega, k)\,|A|^2$, it gives just $g = 0$, and consequently, $L = 0$. This means that in a linear progressive wave, the virial theorem is valid: average kinetic and potential energies are equal. The second equation describes space-time variation of slowly varying quantities; for closure, one should add a kinematic equation

$$\frac{\partial k}{\partial t} + \frac{\partial \omega}{\partial x} = 0. \tag{2.6}$$

Whitham applied this approach to a wider class of nonlinear quasiperiodic (not necessarily harmonic) waves in the form $u = F(\theta, A)$, where A is a parameter (or parameters) of a wave, such as its amplitude. This can be applied, for instance, to the solutions of the Korteweg-de Vries equation in the form of "cnoidal" waves. Later, a similar approach was applied to describe the transformation of shallow-water sea waves from weakly nonlinear to soliton-like while approaching the shore [7] and to construct a self-similar solution describing the disintegration of an initial stepwise wave to solitons [8].

Whitham's paper [5] was followed by Lighthill's [9] who considered an important case of weakly nonlinear waves for which the dispersion relation takes the form $\omega = f(k) + \alpha|A|^2$. He found that the system (2.5), (2.6) can be of either a hyperbolic or elliptic type, and in the latter case, a periodic wave train is unstable with respect to modulating perturbations. He also gave a criterion for wave stability (hyperbolicity), which includes both nonlinearity and dispersion parameters and is known as *Lighthill's criterion*.

Shortly afterwards, higher-order dispersive effects were considered to yield the nonlinear Schrödinger equation (NSE), and many of its important consequences, including the existence of envelope solitons, "dark solitons," etc. This development was particularly stimulated by the progress in nonlinear optics (see the pioneering works by Bloembergen and Khokhlov ([10] and [11]), devoted to the resonance wave interactions, and the papers of the 1960s [12]-[19] that have introduced the NSE and such notions as self-modulation, envelope solitons, and envelope shocks. An important contribution was made by Benjamin and Feir [20], who have shown that the surface gravity (Stokes) waves in water are modulationally unstable. However, all these physical results are beyond the scope of this paper.

Finally, a concept of modulated solitary waves was also introduced. Namely, it was supposed that the unperturbed governing equations have a family of localized, solitary solutions, and then a perturbation method is used to generate both slow variations of the "body" of the solution in the vicinity of a solitary wave and a nonstationary "radiation tail" that propagates behind (and sometimes ahead of) the primarily localized wave. These problems are also considered below.

3 Direct perturbation method

In what follows, we briefly outline the perturbation schemes, starting with that for quasiperiodic (not necessarily quasiharmonic) waves [21].

Consider a system of N nonlinear equations

$$M(u, u_t, u_x, \tau, \rho) = \varepsilon R, \tag{2.7}$$

where M is a column of nonlinear functions, R describes perturbations, $\tau = \varepsilon t, \rho = \varepsilon x$, and u is a vector of unknown variables. Suppose that for $\varepsilon = 0$, this system has a family of 2π-periodic solutions

$$u = U(\theta, A), \tag{2.8}$$

where again $\theta = \omega t - kx$, and $A = A_1, ...A_m$ is a set of integration constants. For $\varepsilon \neq 0$, the solution is represented by an asymptotic series

$$u = U(\theta, A, \tau, \rho) + \sum_n \varepsilon^n u^{(n)}(\theta, \tau, \rho), \tag{2.9}$$

with the same definition of variables $\omega(\tau, \rho) = \theta_t$, $k(\tau, \rho) = -\theta_x$, and $A(\tau, \rho)$ is a slowly-varying amplitude. Substituting into system (2.7) we obtain, in each order of ε, a linear system

$$\hat{P}\, u^{(n)} = H^{(n)}, \qquad \hat{P} = \frac{\partial M^{(0)}}{\partial U_\theta}\frac{d}{d\theta} - \frac{\partial M^{(0)}}{\partial U}. \tag{2.10}$$

Here $M^{(0)}$ is taken from (2.7) for $\varepsilon = 0$. The "forcing" $H^{(n)}$ depends on previous-order functions; for example,

$$H^{(1)} = R^{(0)} - \frac{\partial M^{(0)}}{\partial U_t} U_\tau - \frac{\partial M^{(0)}}{\partial U_x} U_x. \tag{2.11}$$

A general solution for (2.10) in each order can be represented in the form

$$u^{(n)} = Y\left[C^{(n)} + \int_0^\theta Y^* H^{(n)} d\theta'\right], \tag{2.12}$$

with $C^{(n)}(\tau, \rho)$ as integration constants. Here, Y and Y^* are the eigenmatrices for the operator $\hat{P}$ and its conjugate counterpart, respectively.

It is important that $m + 1$ columns of matrix Y are known here just because $\hat{P}$ follows from the variation of the vector M; namely,

$$Y_1 = U_\theta, \quad Y_i = U_{A_i}, \quad i = 2, 3, ...m + 1. \tag{2.13}$$

These solutions remain bounded only if the orthogonality conditions are fulfilled

$$\int_0^{2\pi} Y_i^* H^{(n)} d\theta' = 0, \qquad i = 1, 2, ...m + 1 \tag{2.14}$$

together with a proper selection of parameters $C^{(n)}$. The relations (2.14) are, in fact, differential equations for slowly varying functions in each approximation.

A more detailed consideration is available for Lagrangian systems [22]

$$\frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial u_t} + \frac{\partial}{\partial x}\frac{\partial \mathcal{L}}{\partial u_x} - \frac{\partial \mathcal{L}}{\partial u} = \varepsilon H^{(n)}. \tag{2.15}$$

In this case, the variational operator $\hat{P}$ is of second order and self-conjugate

$$\hat{P} = \frac{d}{d\theta}\left(\frac{\partial^2 \mathcal{L}}{\partial U_\theta^2}\frac{d}{d\theta} + \frac{\partial^2 \mathcal{L}}{\partial U_\theta \partial U}\right) - \frac{\partial^2 \mathcal{L}}{\partial U^2} - \frac{\partial^2 \mathcal{L}}{\partial U_\theta \partial U}\frac{d}{d\theta}. \tag{2.16}$$

For a Lagrangian system when $R = 0$ and the wave is modulated due to initial of boundary conditions, the system (2.15) can be written, in the first-order approximation, in the form (2.5). This means that Whitham's averaged variational principle follows from the perturbation theory as a first-order result.

If x is a vector, these solutions can be expanded to describe three-dimensional waves, thus yielding the equations of nonlinear space-time geometrical optics.

4 Perturbation theories for solitary waves

A very actively developed branch of "perturbation wave science" is devoted to solitons (and, more generally, solitary waves) [1]. Several schemes have been suggested for solitons, which can be divided into two classes: "direct" methods similar to that discussed above for periodic waves, and "inverse" methods based on perturbations in the scheme of inverse scattering for equations close to fully integrable ones. We shall briefly outline each of them.

4.1 Direct perturbation method for solitons: quasistationary approach

This method has been discussed by the authors and Pelinovsky in 1974 [23] in an application to nonlinear electromagnetic waves, and then in several other papers (e. g., by Grimshaw for the KdV equation [32] and by the

[1] We do not discuss the definition of the term "soliton" here; note that we are among those who believe that this term should be applied to all conservative localized nonlinear steady waves which are able to preserve their integrity upon interactions and other perturbations, rather than only those escaping unchanged from interactions.

authors for a general system [24]). Up to some point, this scheme is similar to that described above for periodic waves. Considering system (2.7), we suppose that, at $\varepsilon = 0$, it has a family of solitary solutions $u = U(\zeta, V, A)$, where $\zeta = x - Vt$. These solutions tend to constants at infinities: $U(\zeta \to \pm\infty) = U_{\pm}(A)$. (In general, $U_+ \neq U_-$; moreover, this consideration can also be applied for nonconservative solitary formations such as shock waves and autowaves.) Then at $\varepsilon \neq 0$, we represent the solution in the form (2.9) with $\zeta = x - \int V(\tau, \rho)dt$ instead of θ. As a result, we obtain an equation similar to (2.12). To prevent the growth of perturbations at infinities, the orthogonality conditions must be met for all eigenmatrices Y_i (suppose $i = 2, 3, ...l$) tending to zero at infinity

$$\int_{-\infty}^{\infty} Y_i^* H^{(n)} d\zeta' = 0, \qquad i = 2, 3, ...l. \tag{2.17}$$

Note that in cases of solitons with exponential asymptotics, they prevent exponential growth at infinity rather than power-law (secular) one as in the case of periodic waves. However, if the limits for some matrices Y_i are nonzero but finite, a secular growth at infinity is also possible, and to eliminate it, the following algebraic conditions must be added

$$\lim_{\zeta \to \pm\infty} Y_i^* H^{(n)} = 0, \qquad i = l+1, l+2, ...m+1. \tag{2.18}$$

Together with a proper selection of the integration parameters C, the above conditions (2.17) and (2.18) are necessary for the finiteness of perturbations. The first set of them gives equations for the soliton amplitude. The second set of equations matches the solitary wave to the nonlocalized field component, a radiation. This matching is somewhat analogous to that mentioned above for the relaxing systems described by ODE and also to the problems typical of hydrodynamic boundary layers. (Indeed, in the case of a gas-dynamic shock wave, condition (2.18) gives the boundary conditions on a shock which connect it with isenthropic flow outside the narrow shock area.) The nonzero asymptotic perturbations $U_{\pm}^{(n)}$ arise in higher approximations ($n > 1$). By using (2.12), they can be matched in the vicinity of a solitary wave by a boundary condition

$$U_+^{(n)} - U_-^{(n)} = Y_i \int_{-\infty}^{\infty} Y_i^* H^{(n)} d\zeta', \qquad i = l+1, l+2, ...m+1. \tag{2.19}$$

As a result, we have a complete description of the behavior of a solitary wave (the small radiation field is described by the linearized basic system (2.7)) with perturbations uniformly bounded for all x.

Analogously, the perturbation scheme can be constructed for solitary waves in Lagrangian systems [24], as discussed below for soliton interaction.

4.2 Nonstationary approach

To describe the nonstationary part of the solution, a radiation, a more general scheme, first developed by Keener and McLaughlin in 1977 [25, 26], is more adequate. The solution is represented as

$$u = U(\theta, A, \tau, \rho) + \sum_n \varepsilon^n u^{(n)}(x,t)\,. \tag{2.20}$$

It yields again a linear system for $u^{(n)}$ in each order

$$\hat{P}\, u^{(n)} = H^{(n)}, \qquad \hat{P} = \frac{\partial M^{(0)}}{\partial U_t}\frac{\partial}{\partial t} + \frac{\partial M^{(0)}}{\partial U_x}\frac{\partial}{\partial x} - \frac{\partial M^{(0)}}{\partial U}, \tag{2.21}$$

and, for example,

$$H^{(1)} = R^{(0)} - \frac{\partial M^{(0)}}{\partial U_t}U_\tau - \frac{\partial M^{(0)}}{\partial U_x}U_x. \tag{2.22}$$

A general solution to (2.21) in each order can be represented in the form

$$u^{(n)}(x,t) = \sum_n C_d(t)\, Y_d(x,t) + \int_{-\infty}^{\infty} dk C_c(t)\, Y_c(x,t)\,, \tag{2.23}$$

where C are coefficients to be found. Subscripts "d" and "c" denote the discrete and continuous parts of the operator $\hat{P}$ spectrum, respectively. If $\hat{P}$ is written in a normal form with respect to $\partial_t u^{(n)}$, functions $C_{d,c}$ satisfy a system

$$\frac{dC_{d,c}}{dt} = \int_{-\infty}^{\infty} Y^*_{d,c} H^{(n)} dx. \tag{2.24}$$

A characteristic feature of such a scheme is secular divergence of coefficients C in time. For the continuous-spectrum coefficients, C_c, such a divergence is typically associated with nonlocalized functions Y_k (radiation), so that secular growth of C is due to spreading of the region occupied by radiated waves that remain finite in each point; hence, the corresponding parts of perturbations $u^{(n)}$ do not build up. On the other hand, the coefficients C_d correspond to spatially-localized, discrete-spectrum functions Y_d, so that the discrete-spectrum parts of $u^{(n)}$ grow in time rather than spread in space. Orthogonality conditions (2.17) suppress the C_d growth and define the variation of soliton parameters (such as its amplitude) in time. Here again, the variations of the basic solution with respect to soliton parameters (cf.(2.13)) are among the localized eigenfunctions. Keener and McLaughlin [26] have shown that, if the basic (at $\varepsilon = 0$) system is fully integrable, the entire discrete spectrum is covered by these "variational" solutions. Moreover, in this case all continuous-spectrum functions can be represented in a similar way. As a result, the orthogonality conditions (2.17)

are sufficient to eliminate all divergences; at least in the first approximation, the amount of resonances is equal to that of the parameters $A_i(t)$ to be determined. As a result, the description is reduced to the solution of a finite-dimensional ODE system

$$\frac{dA_i}{dt}\int_{-\infty}^{\infty} Y_d^* U_{A_i} dx = \int_{-\infty}^{\infty} Y_d^* H^{(n)} dx; \qquad i,d = 1,2,...m. \tag{2.25}$$

It is worth noting that when the orthogonality conditions similar to (2.17) are satisfied, the discrete-spectrum perturbation disappears, and $u^{(n)}$ consists of the continuous-spectrum part of (2.23) only.

4.3 Inverse-scattering perturbation scheme

The schemes outlined above do not depend on the integrability or non-integrability of the basic equations. For systems close to integrable ones, an alternative scheme is also applicable that is based on the inverse scattering method, first suggested in 1977 by Karpman and Maslov [28] (see also [29]-[31] and [46]). According to this method, the basic equation is represented in an evolutional form

$$\partial_t u = \hat{S}[u] + \varepsilon \hat{R}[u], \tag{2.26}$$

where $\hat{S}[u]$ and $\varepsilon \hat{R}[u]$ are nonlinear differential operators, and (2.26) is fully integrable at $\varepsilon = 0$ and can be solved by the inverse scattering method (see, e.g., [18]). As known, integrable evolution equations (including KdV, MKdV, SG, and NLS [2]) can be represented in terms of a compatibility condition for a pair of linear equations for a function ψ

$$\partial_x \psi = X(\lambda, u)\psi; \qquad \partial_t \psi = T(\lambda, u)\psi. \tag{2.27}$$

Here, the matrices X and T have the elements depending on the spectral parameter λ and having the solutions of the basic evolution equation as coefficients. As a first stage, the direct problem is solved for the first of these equations, in which the variable t is treated as a parameter. In cases when both X and T are 2×2 matrices, the solution procedure is equivalent to finding the scattering coefficient $r(\lambda)$ ($-\infty < \lambda < \infty$), together with eigenvalues λ_n and amplitude coefficients b_n for discrete-spectrum eigenfunctions. These scattering data can be expressed in terms of Yost coefficients that relate two equivalent sets of eigenfunctions of continuous spectrum: $r(\lambda,t) = b(\lambda,t)a^{-1}(\lambda.t)$. The values of λ_n are given by zeroes of the function $a(\lambda.t)$ in the upper half-plane of the complex parameter λ, where this function is analytical; then $b_n(t) = b(\lambda = \lambda_n, t)$.

---

[2] We use generally accepted abbreviations for the equations: Korteweg-de Vries, Modified Korteweg-de Vries, Sine-Gordon, and Nonlinear Schrödinger.

The equations for a and b follow from the analysis of asymptotic behaviour of the continuous-spectrum functions at $x \to \infty$; they read as

$$\dot{a}\,(\lambda,t)=0, \qquad \dot{b}\,(\lambda,t)=i\Omega\,(\lambda)b(\lambda,t). \tag{2.28}$$

(The function $\Omega(\lambda)$ coincides with the left-hand part of the dispersion relation $\omega(k)$ for the linearized equation (2.26) at $\varepsilon=0$.) From the analytical properties of functions a and b in the upper half-plane of λ, one readily finds equations for λ_n and b_n

$$\dot{\lambda}_n=0, \qquad \dot{b}_n=i\Omega(\lambda_n)b_n(t). \tag{2.29}$$

From the scattering data known for all t, one can unambiguously restore $u(x,t)$ by using the integral equation of the inverse scattering problem. Note that all known N-soliton solutions are associated with nonreflecting potentials, for which $r(\lambda,t)\equiv 0$. In these cases, the constants λ_n define the amplitudes and velocities of interacting solitons, and $b_n(\lambda_n,t)$ describe their phases. From (2.29) it directly follows that the amount and parameters of solitons are the same before and after interactions (elastic collision).

For $\varepsilon\neq 0$, the perturbed equation (2.26) is associated with the same linear operator X, so that the direct scattering problem remains unchanged. However, temporal evolution of scattering data does not obey Eqs. (2.28) and (2.29) anymore. One of the possible ways to describe this evolution is based on the representation of scattering data as functionals of the "scattering potential" $u(x,t)$. By definition, the time derivatives of these functionals are represented as [18]

$$\frac{dF}{dt}=\int_{-\infty}^{\infty}\frac{\delta F}{\delta u}\frac{\partial u}{\partial t}dx, \qquad F=\{a,b,\lambda_n,b_n\}, \tag{2.30}$$

where the variational derivatives $\delta F/\delta u$ can be expressed in terms of eigenfunctions of the scattering problem. For Eq. (2.26) , this relationship takes the form

$$\frac{dF}{dt}=\int_{-\infty}^{\infty}\frac{\delta F}{\delta u}\hat{S}\,(u)dx+\int_{-\infty}^{\infty}\frac{\delta F}{\delta u}\hat{R}\,(u)dx, \tag{2.31}$$

where the first term in the right-hand part is the same as for $\varepsilon=0$. As a result, evolution equations for the scattering data acquire the following structure [26]

$$\begin{aligned}
\dot{a}\,(\lambda,t) &= \varepsilon A(u,\lambda), \\
\dot{b}\,(\lambda,t) &= i\Omega(\lambda)b(\lambda,t)+\varepsilon B(u,\lambda), \\
\dot{\lambda}_n &= \varepsilon\Lambda_n(u), \\
\dot{b}_n &= i\Omega(\lambda_n)b_n(t)+\varepsilon B_n(u).
\end{aligned} \tag{2.32}$$

Here, A, B, Λ_n, and B_n can be defined from the second term in (2.31) , but only as functionals of the unknown solution $u(x,t)$. However, by using

the expansion of (2.9) type for u and the corresponding expansions for the parameters in question, $A(u) = A(U) + \varepsilon A^{(1)} + ...$ etc., where $U(x,t)$ is again a known basic solution for $\varepsilon = 0$, such as an N-soliton solution. As a result, in the first approximation, the functional derivatives are transformable to explicit functions of U depending on parameters λ_n, b_n, and λ that closes the equations (2.32). Their solution provides the time dependences of scattering data and soliton parameters. Then, by restoring the scattering potential, one finds the perturbation $u^{(1)}(x,t)$. After that, the next-order terms can be added to the system (2.32), and so on.

4.4 "Equivalence principle"

As expected, the results obtained in a framework of different perturbation schemes are essentially the same. In particular, the "direct" and "inverse" schemes considered above yield identical equations for soliton parameters and for (at least) first-order perturbations, although some formal differences can exist. For relatively simple but very typical examples of perturbed KdV, MKdV, and SG equations, when a basic soliton $U(x,t)$ depends on one (phase) variable $\zeta = x - V(\lambda_0)t - X_0$ and contains two parameters, λ_0 and X_0, the equations for these parameters, obtained from the inverse scheme in the first approximation, have the form

$$\frac{d\lambda_0}{dt} = \varepsilon E(X_0, \lambda_0, t); \qquad \frac{dX_0}{dt} = V(\lambda_0) + \varepsilon D(X_0, \lambda_0, t). \tag{2.33}$$

When solving the same problem with the direct method, one obtains the same equations, but without the last term (εD) in the second equation. This difference is due to the different amounts of localized eigenfunctions used in the two schemes. Whereas the solution $U_{X_0} \sim U_\zeta$ is present in both schemes, the solution $U_{\lambda_0} \sim U_V$ appears only in the inverse (partial-difference) scheme (in the quasistationary approach, U_V is not an eigenfunction). This circumstance was mentioned, e.g., by Ablowitz and Kodama [47]. However, the term with D is, in fact, a second-order correction to the soliton velocity. In the direct scheme, a similar correction ($\sim U_V$) appears in the expression for $u^{(1)}$. Moreover, a consistent use of the term εD to find a correction to the phase variable ζ is possible only together with terms of order ε^2 in the first equation (2.33); this fact is well known in oscillation theory [1]. In general, the two schemes satisfy the "equivalence principle," necessary for the correct use of either of them. Personally, we prefer the direct method because of its relative simplicity and universality: the scheme is constructed similarly both for integrable and non-integrable equations and, moreover, for solitons and quasiperiodic waves. [3]

[3] Eqs. (2.33) also describe small perturbations in the form of small-scale oscillations [26, 28, 29], which are absent in the quasistationary equations; as mentioned, however, these equations are not completely consistent.

4.5 Example: soliton interaction in Lagrangian systems

In [24], the case was considered when two or more (N) solitons interact on the $x-$axis in such a way that at any moment, they are separated by finite but large (compared to their characteristic lengths) distances. In this case, they interact by their "tails," and the solution in the vicinity of an i-th soliton can be sought in the form

$$u_i = U(\zeta - s_i) + \sum_{j \neq i} U(\zeta - s_j) + \sum_n \varepsilon^n u^{(n)}(\zeta - s_i, \tau, \rho), \tag{2.34}$$

where s is the soliton center coordinate, so that $ds_i/dt = \varepsilon V$, and V is taken the same for all solitons. It is supposed that $V_i - V_j = \Delta V_{ij}$ is of order ε, and the values of "tails" of neighbouring solitons at the site of the i -th one are of order ε^2.

After some transformations, from the orthogonality condition (2.17) , it follows that in the second approximation (which is the first nontrivial in this case), the coordinates of solitons satisfy the equations of Newtonian mechanics for classical particles

$$M\ \ddot{s_i} = \sum_{i \neq j} f(s_i - s_j), \tag{2.35}$$

where f is defined by the soliton asymptotics (exponential, algebraic, etc.). This equation corresponds to the collision of particles with a Lagrangian

$$L_{eff} = \frac{M}{2} \sum \dot{s}_i^2 + \sum \Phi(s_i - s_j), \tag{2.36}$$

with the pair potential Φ expressed in terms of soliton asymptotics. It is worth noting that

$$M = \frac{dP}{dV}, \qquad P = \int_{-\infty}^{\infty} d\zeta \frac{\partial L}{\partial U_t} U_\zeta, \tag{2.37}$$

so that P is the total wave momentum of the soliton, and M is its field mass.

From here, it is easy to understand the main types of two-soliton interaction [24, 57]. Indeed, its character depends on the sign and shape of the function f in (2.35). Positive f corresponds to repulsion of particles, as in the majority of integrable equations (like KdV); negative f yields attraction producing a breather, as in the sine-Gordon equation. Finally, for oscillatory behaviour of f, when solitary solutions have tails of the type $e^{-\lambda|x|}cos(\lambda x)$, the interactions are of a more complicated character: they include equilibrium points, oscillating bound states, and repulsing interactions, depending on initial conditions. Note that the equilibrium is, in fact, a stationary bound state that corresponds to a new, "two-hump" soliton; a countable amount of "multisolitons" can be constructed in this

way [42]. The latter case is realized, for example, for the KdV equation modified by an additional dispersive term with the 5th derivative of the unknown function [38]. Many types of soliton interactions were observed in electromagnetic lines [43, 44]. Note that in these experiments, the birth of new solitons upon interaction was observed. An approximate approach to description of such cases was also discussed [35, 51].

Of interest also are "soliton lattices", i.e., quasiperiodic sets of solitons. In particular, for repulsing interactions these lattices are stable, for attracting interactions they are unstable, and for oscillating potentials both stable and unstable lattices are possible. In the latter case, the stochastization of the soliton set can occur [44].

4.6 Radiation from solitons

An important problem is that of soliton radiation due to perturbations, such as interaction with other solitons. Different problems of soliton radiation were considered, for example, in [39, 33, 55]. In the Lagrangian scheme, the radiation appears in the second approximation; in general, $u^{(2)}$ can be represented as $u_l^{(2)} + u_r^{(2)} + \hat{A}\, U_2$, where the first term is a localized perturbation deforming a soliton, the second is nonlocalized radiation, and the third is the "tail field" of the neighbouring soliton(s) on the site of the considered one. Interestingly, in such integrable equations as MKdV, SG, and NSE, the radiative part $u_l^{(2)}$ for soliton interactions is not present, whereas for KdV it does occur, but only in the region between interacting solitons and it asymptotically disappears at $t \to \infty$, when solitons diverge in space. It can be said, consecutively, that along with the exchange of energy and momentum, the interacting KdV solitons "exchange radiation:" one of the solitons (the frontal one) radiates a field that is eventually completely absorbed by the other.

In other cases (small dissipation, solitary wave interaction in non-integrable equations, etc.), the radiation spreads beyond solitons. (The condition of its decay at infinity for any finite t is still fulfilled.) The corresponding nonstationary term should be matched with the quasistationary asymptotic of a solitary wave, which, in turn, obeys the boundary condition (2.19) . In fact, the method of matched asymptotic expansions, that is well known in oscillation theory [40] and the theory of boundary layer [45], is employed here. Naturally, the energy loss for radiation results in the decay of the localized part.[4]

Finally, we shall make one practical, important remark regarding soli-

[4] An interesting recent example [54] is "terminal damping" of solitary waves in rotating systems. It was shown before that rotation prevents the existence of solitary waves. As a result, an initial pulse close to a KdV soliton radiates a dispersive wave and disappears in a finite time.

ton interaction. The applicability of the above method is often actually wider than for well-separated, close-velocity solitons. In another extreme case, when one soliton is much larger (and narrower) than the other, the smaller one can be considered as a slowly varying pedestal for the stronger one, and the perturbation method is applicable again [39]. Moreover, the direct application of the expansion (2.34) is justified even for an arbitrary ratio of soliton amplitudes. Indeed, a real small parameters of the problem, are, e.g., $\varepsilon = (V_1 - V_2)/(V_1 + V_2)$ for KdV and $\varepsilon = V$ for SG, both are always less than 1. This makes the quasistationary perturbation approach reasonably adequate for all ratios of soliton parameters. A confirmation can be readily found by comparison with exact solutions for such equations as KdV, MKdV, SG, and NSE. For example, an exact two-soliton solution of MKdV can be represented as

$$\begin{aligned} u_{ss}(x,t) &= \sum_{1,2} (\theta_{1,2})_x \, (\cosh \theta_{1,2})^{-1} \qquad (2.38) \\ \theta_{1,2} &= \varsigma - 2\varepsilon^2 t \pm \cosh^{-1}\left[\frac{1}{\varepsilon}\cosh \varepsilon(\varsigma - 2t)\right], \end{aligned}$$

where $\varsigma = V_0\left(x - 4V_0^2 t\right)$, V_0 is the average of partial velocities of solitons, $V_{1,2}$, and $\varepsilon = (V_1 - V_2)/(V_1 + V_2) < 1$. The signs + and - correspond to the first and second soliton, respectively. Expanding (2.38) in ε, it is easy to see that the first terms in this expansion coincide with the solution obtained by the perturbation method, both for the perturbation $u^{(1)}$ and for the parameters of solitons.

5 Asymptotic reduction of nonlinear wave equations

When dealing with evolution equations like KdV in mathematics, one has to remember that in physics, such equations are usually an approximation for a situation when both nonlinear and dispersive terms in a basic system are small. In other words, this system must be close to a linear hyperbolic one. There exist numerous ways to reduce the basic equations to the equations like Burgers, KdV, and others. We can refer, for example, to the early paper by Khokhlov [52], who reduced the equations for nondispersive electromagnetic waves to the form of Burgers equation just by introducing the "travelling" coordinate $\zeta = x - ct$, c being the linear wave speed. The series of papers by Taniuti and others [48, 49] was devoted to such a reduction for a general system of 1-D equations. Another, iterative scheme, which directly exploits the quasi-hyperbolicity of basic equations (which is, in fact, equivalent to weak dispersion), was suggested in [50] (see also [57]).

We shall outline this scheme briefly. Consider a system

$$u_t + \hat{B}\, u_x = \varepsilon\, \hat{R}\,(u, x, t)\,, \tag{2.39}$$

where u is a column vector of N variables, $\hat{B}$ is a square matrix of constant coefficients, and $\hat{R}$ is a vector operator that is responsible, in general, for nonlinearity, dispersion, and dissipation effects. At $\varepsilon = 0$, this system is supposed to be hyperbolic. It is convenient to diagonalize this system by a change of variables

$$u(x,t) = \sum_{n=1}^{N} F_n(x,t)\mathbf{r}_n, \tag{2.40}$$

where F_n are scalar functions, and $\mathbf{r}_n$ are eigenvectors of the problem. At $\varepsilon = 0$, we have $F_n = F_n(x - c_n t)$, where c_n are characteristic velocities of normal waves defined as eigenvalues of matrix $\hat{B}$; i.e.,

$$\det |\, \hat{B} - c_n\, \hat{I}\, | = 0, \qquad \hat{B}\, \mathbf{r}_n = c_n \mathbf{r}_n. \tag{2.41}$$

($\hat{I}$ is a unit matrix.)

For $\varepsilon \neq 0$, after substituting (2.40) into (2.39) and multiplying by the left eigenvector $\mathbf{l}_m$ of the matrix $\hat{B}$ (i.e. $\mathbf{l}_m \cdot \hat{B} = c_m\, \hat{I}$), the latter system can be readily transformed into a set of equations for coupled normal waves F_m

$$F_{mt} + c_m F_{mx} = \varepsilon\, (\mathbf{l}_m \cdot \mathbf{r}_m)\, \mathbf{R}\,(F_n, x, t)\,. \tag{2.42}$$

As a result, we have N weakly coupled normal modes, each propagating along its own characteristic, $dx/dt = c_n$. In practice, only few of these modes are usually excited, so that in the basic approximation, the order of the system is much less than N. The rest of the modes can be found from linear equations of (2.42) with known right-hand parts. Then, the iteration procedure is used: substitute the solution of the linear group of equations into the "main," coupled equations, and solve them again, then find again the perturbing modes from linear equations, etc. A majority of evolution equations, such as KdV, follow from this scheme as a first one-wave approximation (after introduction of a travelling coordinate, $\zeta = x - ct$), and the two-dimensional equations, like Kadomtsev-Petviashvili and Khokhlov-Zabolotskaya-Kuzhetsov equations, in the second approximation.

6 Conclusions

In this book, a question like "What would one expect of nonlinear wave theory (let us call it NWT) on the eve of the 21th century?" seems quite natural, but there is hardly anyone able to give a general answer. It can be

stated that in the second half of our century, NWT has been consolidated as an independent branch of science and as such, it includes a solid mathematical basis and numerous applications. Like those in oscillation theory, approximate methods form one of the most important parts of NWT. In the above consideration, we did not present all known perturbation schemes for waves, while space restrictions prevented us from discussing a variety of applications, some of them quite remarkable. Still, we hope that the paper has demonstrated the universality of asymptotic approaches in NWT and their generic connections with other areas of mathematical physics and oscillation theory. Practical applications of asymptotic perturbation methods are extended from nonlinear optics to plasma physics to electronics to astrophysics.

A look at the recent development in this area shows a number of relevant problems that can be expected to remain a challenge through the beginning of the next century. (Some of them have been already mentioned above). Evidently, the analysis of higher approximations (including soliton radiation) and of 2-D and 3-D schemes is still far from being completed. The same is true for complex cases of soliton interaction in non-integrable systems that may result in birth and disappearance of solitons in the process of interaction. Multisoliton dynamics and soliton lattices are worth studying, in particular, the stochastic soliton ensembles and "soliton turbulence." A perturbation approach to solitons in discrete systems is quite possible, too. Also, an extension of perturbation schemes used for solitons, to localized hydrodynamic structures such as vortices, can give new results in this rather classical area [53]. However, the perspectives of the theory might go far beyond the soliton-related problems. In fact, "nonlinear geometrical acoustics" employed the approximate approach for shock waves for a rather long time (see, e.g., the recent books [56]). Finally, slow-varying autowaves of the reaction-diffusion type will probably become a subject for the perturbation theory. Even this rather unorganized list demonstrates that asymptotic methods will almost definitely remain an effective tool helping to meet the new challenges in wave theory.

Acknowledgments: This work was partially supported by the INTAS grant and the grant of Russian Foundation for Basic Research. The authors are grateful to I. A. Soustova for an assistance in preparation of this paper.

One of the authors (L.O.) is glad to note that his fruitful personal contacts with Robert Parmentier have stimulated some of the studies mentioned in this paper.

7 References

[1] N. N. Bogolubov and Y. A. Mitropolsky. *Asymptotic Methods in Oscillation Theory.* N. Y., Gordon & Breach, 1961.

[2] A. A. Andronov, A. A. Vitt, and S. E. Khaikin. *Oscillation Theory.* Moscow, Fizmatgiz, 1959 (in Russian).

[3] S. M. Ryutov. Modulated oscillations and waves. *Proc. Lebedev Inst.*, 2**,** USSR Acad. Sci., 1940 (in Russian).

[4] L. A. Ostrovsky. Electromagnetic waves in nonlinear media with dispersion. *Sov. Phys. Techn. Phys.*, 8, 679-683, 1964 (in Russian, 1963).

[5] G. B. Whitham. Nonlinear dispersive waves. *Proc. Roy. Soc. Lond.*, A283, 238, 1965.

[6] G. B. Whitham. *Linear and Nonlinear Waves.* John Wiley & Sons, 1974.

[7] L. A. Ostrovsky and E. N. Pelinovsky. Transformation of surface waves in a variable-depth liquid. *Izvestija, Oceanic and Atmospheric Physics*, 6, 934, 1970.

[8] A. V. Gurevich and L. P. Pitaevsky. Decay of initial discontinuity in the Korteweg-de Vries equation. *JETP Lett.*, 17, 193-195, 1973.

[9] M. J. Lighthill. Contributions to the theory of waves in nonlinear dispersive systems. *J. Inst. Math. Appl.*, 1, 269-306, 1965.

[10] N. Bloembergen. *Nonlinear Optics (a Lecture Note).* NY-Amsterdam, W. A. Benjamin Inc., 1965.

[11] S. A. Akhmanov and R. V. Khokhlov. *Problems of Nonlinear Optics: Electromagnetic Waves in Nonlinear Dispersive Media.* N. Y., Gordon & Breach Sci. Publ., 1972.

[12] L. A. Ostrovsky. Propagation of wave packets and space-time self-focusing in a nonlinear medium. *Sov. Phys. JETP*, 24, 797-800, 1967.

[13] D. T. Benney and A. C. Newell. The propagation of nonlinear wave envelopes. *J. Math. Phys. (Stud. Appl. Math.)*, 46, 133-139, 1967.

[14] L. A. Ostrovsky. Envelope shock waves. *Sov. Phys. JETP*, 27, 660-664, 1968.

[15] V. I. Karpman. Self-modulation of nonlinear plane waves in dispersive media. *JETP Lett.,* 6, 277-279, 1967.

[16] V. I. Karpman and E. M. Krushkal. Modulated waves in nonlinear dispersive media. *Sov. Phys. JETP*, 28, 277-281, 1968.

[17] A. G. Litvak and V. I. Talanov. A parabolic equation for calculating the fields in dispersive nonlinear media. *Radiophys. Quant. Electron.*, 10, 296-302, 1967.

[18] V. E. Zakharov. Stability of periodic waves of finite amplitude on the surface of a deep fluid. J. Appl. Mech. Tech. Phys., 2, 86-96, 1968.

[19] T. B. Benjamin. Instability of periodic wavetrains in nonlinear dispersive systems. *Proc. Roy. Soc. Lond.*, Ser. A, 299, 59-75, 1967.

[20] T. B. Benjamin and J. F. Feir. The disintegration of wavetrains on deep water. Part 1: Theory. *J. Fluid Mech.* 27, 417-430, 1967.

[21] L. A. Ostrovsky and E. N. Pelinovsky. Averaging method for nonsinusoidal waves. *Doklady*, 195, 804, 1970.

[22] L. A. Ostrovsky and E. N. Pelinovsky. Averaging method and generalized variational principle for nonsinusoidal waves. *Prikladnaya Matematika i Mekhanika*, 36, 71, 1972.

[23] K. A. Gorshkov, L. A. Ostrovsky, and E. N. Pelinovsky. Some problems of asymptotic theory for nonlinear waves. *Proc. IEEE*, 62, 1511, 1974.

[24] K. A. Gorshkov and L. A. Ostrovsky. Interaction of solitons in nonintegrable systems: direct perturbation method and applications. *Physica D*, 3, 428, 1981.

[25] J. P. Keener and D. W. McLaughlin. Green's function for a linear equation associated with solitons. *J. Math. Phys.,* 18, 2008, 1977.

[26] J. P. Keener and D. W. McLaughlin. Soliton under perturbation. *Phys. Rev.* A16, 777, 1977.

[27] V. E. Zakharov. *Soliton Theory: Inverse Scattering Method* (ed. S. P. Novikov). Moscow, Nauka, 1980.

[28] V. I. Karpman and E. M. Maslov. Perturbation theory for solitons. *JETP*, 73, 537, 1977.

[29] D. J. Kaup and A. C. Newell. Soliton as particle, oscillator in slowly changing media: A singular perturbation theory. Proc. Roy. Soc., A301, 413, 1978.

[30] V. I. Karpman. Soliton evolution in presence of perturbations. *Physica Scripta,* 20, 462, 1979.

[31] E. M. Maslov. To the perturbation theory in the second approximation. *Teor. i Mat. Fiz.*, 42, 362, 1980 [in Russian].

[32] R. Grimshaw. Slowly varying solitary waves. 1. Korteveg-de Vries equation. *Proc. Roy. Soc.*, A368, 359, 1979.

[33] R. Grimshaw and H. Mitsudera. Slowly-varying solitary wave solution of the perturbed Korteweg-de Vries equation revisited. *Stud. Appl. Math.*, 90, 75, 1993.

[34] A. Bondeson, M. Kisak, and D. Anderson. Soliton perturbation. A variational principle for soliton parameters. *Physica Scripta*, 20, 479, 1979.

[35] J. C. Fernandez, C. Froeschle, and G. Reinisch. Adiabatic perturbation of solitons and shock waves. *Physica Scripta,* 20, 54, 1979.

[36] Ko K. and H. H. Kuehl. Korteweg-de Vries soliton in slowly varying medium. *Phys. Rev. Lett.* 1978, 40, 233.

[37] M. Ablowitz and H. Segur. *Solitons and the inverse scattering transform.* SIAM, Philadelphia, 1981.

[38] T. Kawahara. Oscillatory solitary wave in dispersive media. *J. Phys. Soc. Japan*, 33, 60, 1972.

[39] K. A. Gorshkov and L. A. Ostrovsky. Interaction of solitons with their radiation fields. *III Int. Symp. on Select. Topics in Stat. Mech.,* Part 1, Dubna, 222, 1984.

[40] J. D. Cole. *Perturbation Methods in Applied Mathematics.* Blaisdell, Waltham, MA, 1968.

[41] A. H. Nayfeh. Perturbation Methods. N. Y., Wiley-Interscience, 1973.

[42] K. A. Gorshkov, L. A. Ostrovsky, V. V. Papko, and A. S. Pickovsky. On the existence of stationary multisolitons. *Phys. Lett. A,* 74, 177-179.

[43] K. A. Gorshkov, L. A. Ostrovsky, and V. V. Papko. Interactions and bound states of solitons as classical particles. *Sov. Phys. JETP*, 44, 306-311, 1976.

[44] K. A. Gorshkov and V. V. Papko. Dynamic and stochastic oscillations of soliton lattices. *Sov. Phys. JETP*, 46, 92-97, 1977.

[45] M. Van Dyke. Perturbation methods in fluid dynamics. N. Y., *Acad. Press,* 1964.

[46] Y. Kodama and M. J. Ablowitz. Perturbations of solitons and solitary waves. *Stud. Appl. Math.*, 64, 225-245, 1981.

[47] M. J. Ablowitz and Y. Kodama. Note on asymptotic solutions of the Korteweg-de Vries equation. *Stud. Appl. Math.*, 66, 159-170, 1982.

[48] T. Taniuti and C. C. Wei. Reductive perturbation method in nonlinear wave propagation. *J. Phys. Soc. Japan*, 24, 941, 1968.

[49] N. Asano and T. Taniuti. Reductive perturbation method for nonlinear wave propagation i inhomogeneous media. *J. Phys. Soc. Japan,* 27, 1059, 1969.

[50] L. A. Ostrovsky and E. N. Pelinovsky. On approximate equations for waves in media with small nonlinearity and dispersion. *Prikl. Math. Mech.*, 38, 121, 1974.

[51] B. A. Malomed and Yu. S. Kivshar. Dynamics of solitons in nearly integrable systems. *Rev. Math. Phys.*, 61, 763, 1989.

[52] R. V. Khokhlov. To the theory of shock radiowaves in nonlinear lines. *Radiotechn. i Elektron.*, 6, 917-925, 1961 (in Russian).

[53] K. A. Gorshkov, I. A. Soustova, and L. A. Ostrovsky. Perturbation theory in vortex dynamics. *J. Fluid Mech.* (in print).

[54] R. H. J. Grimshaw, J-M. He, and L. A. Ostrovsky. Terminal damping of a solitary wave due to radiation in rotational systems. *Stud. Appl. Math.,* 101, 197-210, 1998.

[55] R. Grimshaw, E. Pelinovsky, and X. Tian. Interaction of a solitary wave with an external force. *Physica D*, 77, 405-433, 1994.

[56] K. A. Naugolnykh and L. A. Ostrovsky. *Nonlinear Wave Processes in Acoustics.* N. Y.-Cambridge-Melbourne, Cambridge Univ. Press, 1998.

[57] L. A. Ostrovsky and A. I. Potapov. *Modulated Waves: Theory and Applications.* Baltimore-London, Johns Hopkins Univ. Press, 1999.

3
Josephson Devices

Antonio Barone
Sergio Pagano

ABSTRACT In this chapter we briefly review the main applications of Josephson effect together with the most successful devices realized. We will give an overview of the various devices, providing also some basic concepts of the underlying physical mechanisms involved, and the associated limit performances. Some considerations on the concrete possibilities of successful "market ready" implementation will also be given.

1 Introduction

Some stimulating aspects of Josephson structures have been considered in detail in other contributions to the present volume. In this review we shall confine our attention to the main features of superconductive electronics which, as far as the active devices is concerned, can be considered as entirely based on the Josephson junctions.

We consider first the physical principles underlying the operation of such devices in the various applications both potential or already realized.

It is quite usual to distinguish electronic components in the two classes of "analog" and "digital" devices. In both cases one can resort to general properties of superconductivity and specific features of the Josephson effect.

Within these classes of devices the most stimulating aspects which make the superconductive technology quite competitive are the possibility of enhancing processing speed of digital circuits and realizing magnetic field sensors of extreme sensitivity. The former is based on the intrinsic speed that Josephson junctions, in different configurations (latching mode, RSFQ mode, etc.), can offer and, even more, the extremely low power dissipation involved. This last issue is further improved by the possibility of superconductive interconnection of microstrips which allows a sort of ballistic transfer of pulses due to the very low attenuation and dispersion.

As for the magnetic field sensors the unparalleled competitiveness lies in the combination of fundamental properties of superconductivity, such as the flux quantization, with the intrinsic possibilities of Josephson circuits.

Of course, when discussing of the actual potential of superconductive devices, it is mandatory to take into account the fantastic growth of the

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 69–86, 2000.

performances of semiconducting devices in the last decades.

In this chapter we will briefly recall the main features of the Josephson effect, we will discuss the main applications of SQUID systems, also in consideration of the new possibilities offered by the high-T_c superconductors, and finally we will discuss in some detail the digital and high frequency devices. In the concluding remarks we will address some issues related to market opportunities and limitations.

2 Elements of the Josephson effect

To make this chapter self-contained, we shall first recall the basic relations which govern the Josephson effect [14]

$$I = I_c \sin\phi \tag{3.1a}$$

$$V = \frac{\hbar}{2e}\frac{\partial\phi}{\partial t}. \tag{3.1b}$$

As we know these express the circumstance that $V = 0$ implies constant phase difference between the two superconductors and therefore a supercurrent of maximum value Ic flowing through the barrier with zero voltage drop. This is the essence of dc-Josephson effect (see Figure 1).

If we apply a voltage, V, across the junction, equation (3.1b) gives, by integration, $\phi = \phi_0 + 2eV/\hbar$ and thus, through (3.1a), an alternating current $I = I_c \sin(\phi_0 + 2\pi\nu t)$ with a frequency to voltage ratio $\nu/V = 483.6\ MHz/\mu V$ which is the a.c. Josephson effect. Equation (3.1b) gives a time modulation of the phase by the voltage. It does occur however also a space modulation by an applied magnetic field as described by the following equation

$$\nabla\phi = \frac{2e}{\hbar} d\mu_0 \quad \underline{H} \times \underline{n} \tag{3.1c}$$

with $d = t + \lambda_{L_1} + \lambda_{L_2}$ where μ_0 is the vacuum permeability, t is the junction barrier thickness ($\cong 1nm$) and λ_{L_1} and λ_{L_2} are the London penetration depths of the two superconducting layers. Equations (3.1) can be combined [3] giving the well known Sine-Gordon equation

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{1}{v^2}\frac{\partial^2\phi}{\partial t^2} = \frac{1}{\lambda_j^2}\sin\phi \tag{3.2}$$

with $v = c\sqrt{t/\varepsilon_r\varepsilon_0\mu_0 d}$ and $\lambda_j = \sqrt{\hbar/2e\mu_0 J_c d}$.

λ_j gives a measure of the distance, from the junction edge, within which the tunneling supercurrent is confined by the self magnetic field, namely

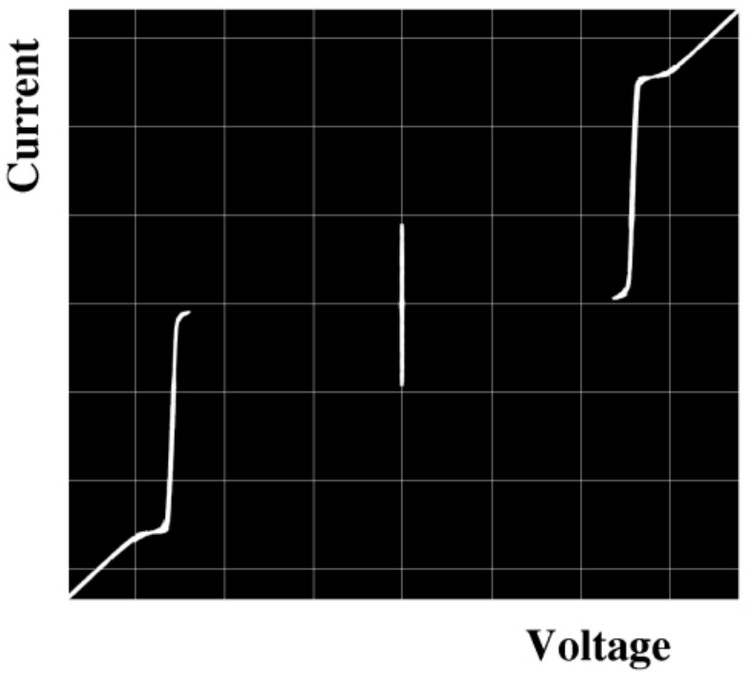

FIGURE 1. Experimental current-voltage characteristics of a high quality Nb-based Josephson junction. The scale is $X = 1mV/div$ and $Y = 0.1mA/div$. Courtesy of Istituto di Cibernetica del CNR, Italy.

the field generated by the currents flowing into the junction. Junctions of linear dimensions smaller, or larger, than λ_j are classified as "small" and "large" junctions respectively.

In the case of small junctions it is easy to show that

$$I_c = \left| \int_{-\infty}^{+\infty} I(x) e^{\imath k x} dx \right| \tag{3.3}$$

where $I(x)$ is the current density profile in the junction barrier and $k = 2\pi d/\Phi_0 H_y$ (Φ_0 being the flux quantum). That is, the dependence of the maximum Josephson current on the applied magnetic field is given by the modulus of the Fourier transform of the current density profile $I(x)$. For

a rectangular junction, in terms of magnetic flux threading the junction, $\Phi = H_y L d$, we have a Fraunhofer-like pattern

$$I_c = \quad I_c(0) \left| \tfrac{1}{\pi\Phi/\Phi_0} \sin(\pi\Phi/\Phi_0) \right| . \tag{3.4}$$

In addition to this diffraction behavior, interference phenomena can also occur when a system of two Josephson junctions closed by a superconducting loop is considered

$$I_c = \quad I_c(0) \ \cos(\Phi_e/\Phi_0) \tag{3.5}$$

where Φ_e represents the flux enclosed in the loop and the periodicity is given by a flux quantum. As we shall see in the following section, these concepts lead to the most important family of devices: the SQUID's.

3 SQUIDs

The SQUID (Superconducting QUantum Interference Device) is probably the most simple and successful device realized using Josephson junctions. It cleverly combines two important physical principles of superconductivity: the correlation between the magnetic field and the order parameter phase, which leads to the flux quantization, and the correlation between the order parameter phase and the current through a tunnel junction (equations 1), which leads to the Josephson effect. The former is made explicit by

$$\underline{J} = \quad \rho e/m \, (\hbar \nabla \phi - 2e\underline{A}) \tag{3.6}$$

where $\underline{J}$ is the current density, ρ the order parameter amplitude, e and m the electron charge and mass respectively, ϕ the order parameter phase, and $\underline{A}$ the vector potential. It is straightforward to show that, by integrating (3.6) along a path inside ($\underline{J} = 0$) a superconducting loop, the flux of magnetic field threading the loop itself must be an integer multiple of Φ_0. A SQUID consists generally of a superconducting loop interrupted by one (rf-SQUID) or two (dc-SQUID) Josephson junctions. In this way a magnetic field coupled to the loop will modify the phase along it, and, consequently, the current crossing the junctions (3.1a). As a result we have a device capable of converting small changes of the magnetic flux, in the scale of fractions of Φ_0, into electrically measurable quantities.

Although the first realizations of SQUIDs, which date back three decades, involved bulk superconductors with junctions formed by point contacts, nowadays the planar thin film technology is widely used which allows fine control of the device geometry and characteristics, as shown in Figure 2.

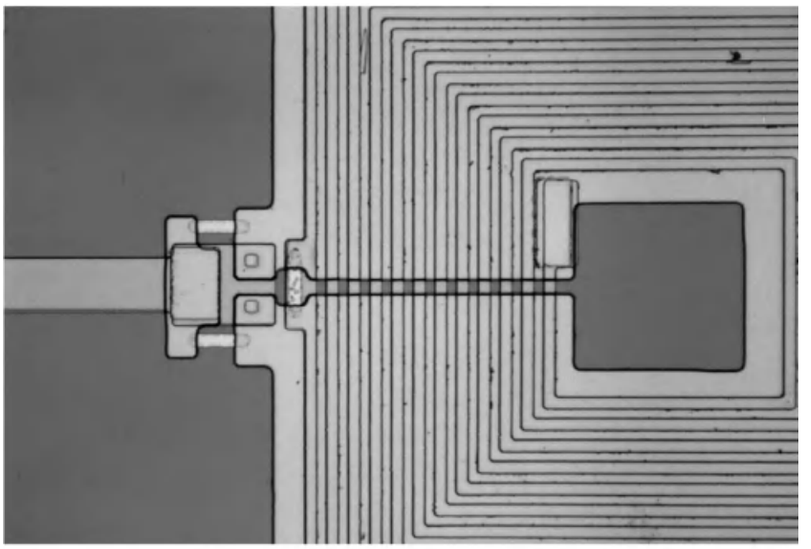

FIGURE 2. Micro-photograph of a planar Nb-based dc-SQUID magnetometer for biomagnetic measurements. The superconductive loop is made by the large square washer to the right, on top of which a spiral coil, belonging to the input transformer, is integrated. The two junctions forming the SQUID are the small squares to the left of the figure, while the two small horizontal stripes are shunting resistors used to make the junctions I-V characteristic single-valued. Courtesy of the Istituto di Cibernetica del CNR, Italy.

The complete theory of operation of a SQUID is rather complex, as it is necessary to take into account the loop inductance, possible junction differences and thermally generated noise, as well as the effect of externally coupled circuits. A detailed treatment can be found in [3, 37].

An important SQUID parameter is: $\beta = 2\pi LI_c/\Phi_0$, which is the ratio between the maximum flux that can be generated by a persistent current in the loop and the flux quantum. For practical SQUID operation this parameter has to be about unity, in order to obtain an optimal critical current modulation, as described in (3.5). This requirement implies the need of small values of L and consequently small area loop (few tens of μm^2). This would greatly reduce the magnetic field sensitivity of the SQUID. In order to measure very weak magnetic fields normally a superconducting flux transformer is used so that the effective sensing area and field sensitivity result greatly enhanced. Another important component of a SQUID magnetometer is the readout electronics. Indeed, as it is evident from (3.5), the SQUID response is periodic in magnetic field with a periodicity of Φ_0. The purpose of the readout electronics is to amplify and linearize the output

signal. The first goal is accomplished by an rf flux modulation and phase detection technique, and the second one by a negative flux feedback that exactly compensates the external field. A modern SQUID magnetometer has a noise level of few $\mu\Phi_0$ and a magnetic field sensitivity of few fT (10^{-10} the Earth magnetic field). This performance, together with a dc limited low frequency response, make the SQUID an ideal sensor in all situations where very weak magnetic fields have to be measured.

A major field of application of SQUID magnetometers in undoubtedly biomagnetism, where the very tiny magnetic fields generated by currents flowing in neuronal and muscular tissues have to be recorded. The main interests are in magneto-encephalography, for the study of normal, evoked and pathological currents generated in the brain by neuronal activity, in magneto-cardiography, for the study of pattern of the currents flowing in the heart tissues, in fetal magneto-cardiography, for the investigation of the fetal heart in situations when standard techniques (electro-cardiography) are not usable, and in the study of magnetic contaminants in lungs and liver.

The last years have witnessed an impressive improvement in biomagnetic SQUID systems. From single sensor systems, which were difficult to use, required extreme shielding against external noise, and needed a long time to perform a complete scanning of the surface of interest (head or chest), to modern multi-channel magneto-encephalographic systems, which contain many SQUID sensors (up to more than 100) in various gradiometric configurations, combined analog and digital readout electronics to optimize SQUID operation and performances, active shielding, and concurrent bio-electrical signal and head position recording [35].

Another application where the extreme sensitivity of the SQUID can be very useful is the non-destructive testing of materials. This field has a very broad interest, ranging from aircraft aging to nuclear power plants pipeline testing. In all these cases the sensitivity at low frequencies of the SQUIDs allows the investigation of material properties and defects to a depth not easily accessible by traditional systems. As an example, with a SQUID eddy current system it is possible to locate defects in multi-layered aluminum foils down to a depth in excess of 1 cm. However, differently from the case of biomagnetism, this application requires the possibility of moving the SQUID system into an uncontrolled environment. This portability issue raises two main problems: refrigeration and noise cancellation. While for the second problem various solutions have been and still are being, developed like active shielding and gradiometric sensors, the first one is still a major task. In this context SQUIDs based on high-T_c materials may have a better chance. Indeed in this case the advantages of simplified cryogenic requirements, $T < 90K$ instead of $T < 10K$, largely compensate for the limitations in sensitivity and reliability due to the use of unfriendly ceramic materials [35].

Finally it is worth mentioning some small but significant SQUID applica-

tions where the unusual performances of these sensors are essential. These are the displacement sensor for the gravitational wave antenna, capable of detecting displacements of the order of $10^{-18} m/\sqrt{Hz}$, low noise amplifiers for dark matter detection experiments, magnetic microscopy, and low field nuclear magnetic resonance imaging [35].

4 Digital devices

Besides early attempts (40 years ago!) based on the "switching" between the superconducting and normal state used by the archetype "cryotron", a large amount of research activity was developed all over the world in the following years to realize superconductive computer devices. Among the largest efforts we must recall the IBM and the following MITI projects which were essentially closed in the 1983 and 1990 respectively, without reaching the objective of commercialization although after a collection of excellent specific results [8, 3]. Original experimental results obtained by Matisoo [19] demonstrated the possibility of transition between the $V = 0$ state to the $V = 2\Delta/e$ state with a switching time less than 1 ns. The back-switching to the V=0 requires a longer time. Flip-flop mode operation was also achieved with two junctions systems. In the following years shorter switching time were reached. It is the case of the observation by Jutzi et al. [13] of a switching time of about 38 ps. In all these experiments the switching mode was of the "latching" type requiring thereby a specific resetting operation (decrease of the gate current). Actually, a voltage V_m exists such that, when a resistive loading is considered, if the load line intersects the quasiparticle branch of the I-V characteristics at $V < V_m$, a "self-resetting" process is allowed in which the junction switches back to the zero voltage state independently of the control current. For intermediate values of the loading a "non-latching" behavior can be realized in which the reset occurs as soon as the control current is removed.

Addressing the reader to the specific bibliography [8, 3] we can mention that the "latching" mode operation circuitry was adopted making thereby a superconductive version of the semiconductor computer device architecture. In the 1985 Likharev and coworkers [18] proposed a new family of digital devices based on intrinsic properties of superconductors which should guarantee ultrafast digital electronics. This rapid Single-Flux-Quantum (RSFQ) circuitry is indicated as the most challenging research issue toward a competitive ultrafast superconductive digital electronics. The RSFQ circuit allows the storage of digital bits as single magnetic flux quanta trapped in superconducting loops. In the simplest configuration, one can consider a superconducting loop interrupted by just one Josephson over-damped junction. This represents a DC/SFQ converter in that it generates a SFQ pulse V(t) as soon as an input ramp signal reaches a certain threshold

value. We are dealing, therefore, with a quantum interferometer by which we could appreciate a quantized voltage drop across the junction produced by a change in the magnetic flux of one quantum Φ_0. The switching between the two states occurs within a time of the order of a picosecond (it is actually limited through fundamental considerations by the energy gap value of the superconductor employed). In practical circuit configuration, an important problem lies in the transfer of such a flux quantum bit (SFQ) to another cell. This is accomplished by passive superconductive microstrip lines which allow a sort of ballistic transfer. Picoseconds pulses can indeed travel at a speed close to the velocity of light with negligible dissipation and dispersion. Of course the interconnections can be conceived also as active transmission lines including extra Josephson devices when pulse amplification is required.

The simplest RSFQ logic element can be represented by a flip-flop device configuration consisting of a superconducting loop containing two junctions. In this system a SFQ pulse produces the switching of the device between two flux states 0 and 1. The device is current biased to apply a magnetic flux $\Phi_0/2$ to the loop creating thereby a condition of two possible flux states $n = 0$ and $n = 1$. The flipping from state 0 to state 1 is produced by a SFQ pulse which transfers a quantum Φ_0 across the first junction. A reset SFQ pulse can reproduce the initial state. Such a resetting process expels a quantum Φ_0 from the second junction leading thereby to an output SFQ pulse which, in turn, can drive another device. A variety of a large RSFQ family can be thus generated with quite complex configurations. Design and implementation of such circuitry employing low-T_c superconductors has been widely demonstrated. The niobium trylayer technology, developed also in different contexts, guarantees a quite reliable circuit implementation. As already mentioned, the combination of high speed and low dissipation presents an attractive perspective. Indeed simple circuits with junctions having a size of $1 - 2\mu m$ allowed operation at frequency up to about 400 GHz while submicron technology would promise speed of LSI of 100-200 GHz. The potential of RSFQ circuits technology spans over a large variety of applications. These include such as communication, radio astronomy, NDE (Non Destructive Evaluation) etc. where speed up of the processing of both digital signal and imaging is of paramount importance. Accordingly, different circuit configurations are investigated such as analog to digital converters for radars, communications systems or digital magnetometers for NDE, Biomagnetism, etc.

As for the extremely challenging perspective of a competitive superconducting superfast general-purpose computer there is an ambitious projection made by Likharev [17] in the last year. It consists of a peta flops-scale (10^{15} floating point operations per second) computer based on RSFQ technology. The intrinsic potential of this technology and preliminary estimates of the systems design [7] show great advantages over a possible semiconductor version of a computer with petaflops scale performance. As

for other superconducting applications, the required cryogenic environment remains a drawback although the final evaluation is always related to the cost/benefit ratio. From this perspective, of high-T_c superconductor implementation would be attractive. However the degree of maturity of such a technology, specially for complex systems of loop arrays of junction devices, appears not yet sufficient, although it is continuously improving (e.g. see [12]).

5 Detectors

Another important field of application on which the Josephson junction technology has a significant impact is radiation detection. In this area the history dates back several decades [1]). Originally the main reasons of interest lie in both the possibility of reaching very high energy resolution, due to the smallness of the energy gap (say 3 orders of magnitude smaller than for semiconductors) and in the radiation hardness offered by various superconductors materials. Confining the attention to superconducting detectors in junction configurations, the role of quasiparticle tunneling for energy spectroscopy has been widely demonstrated. In this case the junctions are of Josephson type because of the large tunneling barrier transparency required, though the actual mode of operation is not based on the Josephson supercurrent. Rather the junction operates with a bias loading into the subgap region. The energy released in the absorber, namely the junction electrodes, create quasiparticles and phonons. The resulting non-equilibrium conditions determine an overall scenario that is quite complex. The quasiparticles produced cross the junction barrier leading to an increase of the subgap current which in turn produces the signal. The tunneling of such quasiparticles is in competition with the quasiparticle recombination process and quasiparticle losses. The interplay of these various rates determines the actual response of the detector. Indeed the production of quasiparticles results from the breaking of Cooper pairs by phonons. The whole evolution of such a three-fluid system (pairs, quasiparticles and phonons) is described by fundamental equations that govern non-equilibrium superconductivity. As previously discussed, junctions employed for energy spectroscopy operate in the quasiparticle regime and the Josephson current is suppressed by a suitable applied magnetic field. There is however also the possibility of using the intrinsic properties of the Josephson effect itself in the fast particle detection [4]. Indeed, in this context, junction detector configurations can be considered that offer faster response, digital output and simple interface with Josephson junction circuitry. Fast discrimination in a time down to 1 ps can be obtained. RSFQ latching operation, as discussed previously, can be also adopted. The main drawbacks with this technology lie in the energy threshold to produce Josephson switching and in the extreme sensitivity

to magnetic field. The former requires very small junction dimensions (say $5 \times 5\mu m^2$) while, the latter demands efficient local shielding. Of course the problem of 2D geometrical efficiency can be solved by producing large arrays of junctions. This, in turn, requires an advanced technology in order to reach a close control of the junctions barrier thickness and therefore the necessary Josephson current density uniformity. Spatial resolution is set by the specific lithographic processes and favored by intrinsic superconducting properties. Micron and submicron technologies have been proven. High packing density is allowed by the very low power dissipation. Moreover dispersionless propagation via superconducting microstrips, is also obtained. It is of interest to consider that, so far, we have identified the absorber of the detector with the junction itself (its superconducting electrodes). Indeed it is of interest to investigate detector configurations in which the absorber is decoupled from the sensor. This is the case of ref. [6] where a relatively large crystal has been considered as absorber and the signal detection is performed by a junction sensor (junctions arrays). Finally, let us remark the fundamental importance of combining the various superconducting detector configurations with suitable associate superconductive electronics: Josephson junctions and SQUIDs, SFQ logic etc. Passive and active superconductive devices for processing of the detector signals could play a significant role also in hybrid structures [28].

6 Voltage standard

As we showed in section 1, a constant voltage difference V across a junction determines a linear increase in time of the phase, and consequently the supercurrent in the junction oscillates at a relatively high frequency $f \cong 0.5 \quad GHz/\mu V)$. This behavior, called the ac Josephson effect, implies that a Josephson junction can be essentially considered as a voltage controlled microwave oscillator. A natural question arises whether such oscillator can be phase-locked to an external one. The early experiments of Shapiro [32] indeed showed this possibility with the appearance of constant voltage steps in the dc I-V characteristics of the junction (inverse ac Josephson effect) as a signature. A simple analysis [3] shows that such steps occur at voltages $V_n = nh/2eF_{rf}$, where F_{rf} is the frequency of the applied microwave field, and n is an integer. The corresponding current amplitude can be easily calculated, under the hypothesis of a voltage biased junction, as $I_n = I_0 J_n(2eV_{rf}/hF_{rf})$, where J_n is the nth order Bessel function of first kind and V_{rf} the voltage amplitude of the microwave field. It was immediately clear that the relation between the applied frequency F_{rf} and the step voltage V_n could be used to determine the value of the constant e/h [29], or, once its value was fixed as an universal constant, to define a voltage standard from a frequency standard. In the latter case

the advantage is that the frequency standard is known with a very high precision (10^{-12}), while the traditional voltage standard, based on Weston cells, is accurate only to few parts in 10^7.

The practical implementation of a Josephson based voltage standard, however, has not been not simple, for a number of reasons. One is that, due to the smallness of the constant $h/2e$, the voltage spacing of the rf-induced current steps (called Shapiro steps) is rather small, about $2\mu V/GHz$. Using single junctions connected to standard X-band ($8 - 12Ghz$) systems, step voltages up to few mV were obtained [11]. However the necessity of comparing the Josephson standard with secondary standards (typically Zener diodes at 1 or 10 V) required the use of resistive voltage dividers that, even in their best implementations, spoiled most of the intrinsic precision of the Josephson standard. One additional problem encountered in the development of a suitable Josephson voltage standard was the possible triggering of intrinsic chaotic dynamics of the junction [15], especially when the driving frequency is close to the characteristic Josephson frequency (the so-called plasma frequency) or when the junction size is larger than the characteristic Josephson length (λ_j).

To overcome these difficulties a number of designs were explored. The most successful was developed by NIST-USA and later by PTB-Germany jointly with ETL-Japan, employing a large array of series connected Nb-based small Josephson junctions biased by an high frequency microwave field ($F = 70 - 90GHz$) [10, 27, 21]. In these devices a large number (more than 10000) of junctions are built on specially designed integrated transmission lines. The design is such that each junction receives about the same amount of microwave energy from the external source, while all the junctions are connected in series from the dc point of view, so that the overall voltage is the sum of the voltage of the single junctions. In Figure 3 is shown an example of such arrays capable of generating 10 V fabricated at PTB [20].

In this type of device the main fabrication issues are chip uniformity of the junction characteristics and process yield, considering the large number of elements involved. For these same reasons an implementation using high-T_c superconductors and junctions is impossible with the present level of technology.

Recently a new interesting design for the implementation of an integrated Josephson voltage standard has been proposed [5]. It consists of a double array of Josephson junctions: one is acting as microwave source and the other as detector and generator of constant voltage steps. The peculiarity lies in the fact that the source junctions are operating in a phase-locked resonant flux mode, in which one magnetic flux quantum (fluxon) oscillates inside each junction, and all the oscillators are phase-locked because of mutual interaction. If a phase-locking of the Josephson oscillators to an external low frequency reference signal could be also achieved, then a simple low-cost voltage standard system could be fabricated and made available

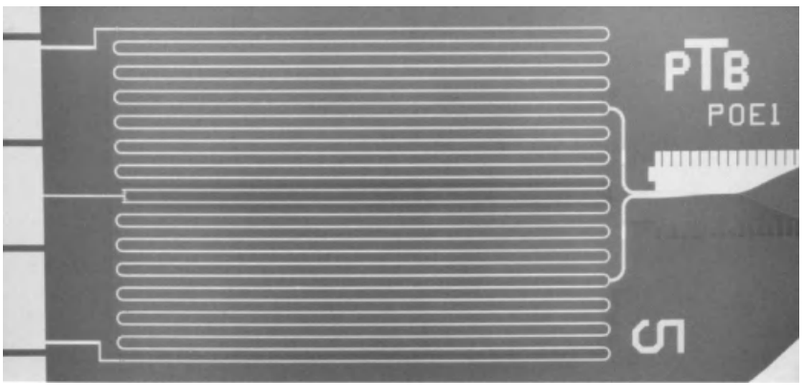

FIGURE 3. Photograph of a 10 V voltage-standard array with 13924 junctions arranged in four microwave paths. The chip is inserted into a microwave transmission line and the microwave signal is coupled through the fin-line antenna to the right. The dc-current bias and voltage pick up contacts are shown on the left. The ground plane and the load extend over the circuit. (Courtesy of PTB-Germany).

to calibration laboratories.

Although the current implementation of the dc-Josephson voltage standard can be considered a mature technology, there is the need for the development of Josephson based ac-voltage standards. This is because modern high resolution ADC and DAC requires high precision ac-waveforms to be calibrated (a 20 bit DAC require a voltage precision better than 1 ppm). In this case a device is needed capable of generating programmable voltage waveforms with frequencies up to few kHz and with very high time and amplitude precision. A number of different solutions have been recently proposed to address this problem [31, 33], some by changing in a programmed way the Shapiro step on which the junctions are biased, others by employing rather complex digital Josephson circuitry. Although very promising, these approaches require relatively complex designs and have to be considered still at a development stage.

7 Microwave oscillators

As we showed in the previous section, a Josephson junction can be considered as a voltage controlled oscillator. It is therefore natural to try to use this property to fabricate cryogenic microwave oscillators to power various types of devices, such as SIS mixers, voltage standard arrays, digital circuits, etc. The main advantages of Josephson oscillators are a wide tunability, compared to other microwave sources, and a small radiation linewidth.

The former derives directly from the linear proportionality between dc-bias voltage and emitted frequency, the latter mainly from the low operating temperature.

A drawback, however, is the very low power available from a single junctions, being generally less than $1nW$ and strongly dependent on the junction parameters and operating frequency. In order to overcome this difficulty two different configurations have been considered: large junction arrays and long junctions in the flux-flow regime. The first approach consists in the realization of arrays of junctions in various configurations: 1D, 2D, stacked, with the aim of summing coherently the microwave signal generated by each junction, thus achieving more emitted power. This technique relies on the possibility of phase-locking the signal generated by a junction by means of an external periodic excitation. In this case the excitation is given by the other junctions of the array by means of an appropriate coupling mechanism. The specific way this mechanism is realized determines the capacity of the array of acting as a coherent source. Against such behavior are the unavoidable differences among the junctions in the array (a direct consequence of the fabrication processes), the effect of external electromagnetic disturbances, which couple to the array as bias current and magnetic field, and the position dependence of the applied microwave field.

An example of microwave oscillator realized with a Josephson array is the device developed at Stony Brook [Lukens], where 40 junctions were connected in series and to a common load. In this case the coupling occurs because of the common load and of the standing wave generated along the array. This type of coupling, named "global coupling", does not provide a very strong coupling among the individual junctions and therefore the mutual phase-locking is easily destroyed. Nevertheless with proper design and fabrication quality it was possible to achieve a microwave generation at $350GHz$ with a power of about $1\mu W$ [36].

Other ways of coupling the junctions of an array are the use of a common microwave resonator which accumulates the signal emitted by the junctions, properly tuned to the resonator fundamental frequency, and feeds it back to the other junctions in the array. In this way a much more efficient coupling can be achieved, if a high-Q resonator is used, although only at frequencies corresponding to the normal modes of the resonator.

A different approach involves junctions connected in parallel arrays. Each junction is coupled to the others through a direct interaction. Indeed, due to the fluxoid quantization rule, the currents circulating in the superconducting loops forming the "elementary cells" of the array are not independent. In this way the current in each branch influences the current in all the others, with a spatially decaying strength, thus achieving a mechanism for phase-locking the junction of the array. However this mechanism tends to be local, being due to mutual magnetic coupling between the various array branches, and long range fluctuations are possible. This type of array realizes the discrete version of a long Josephson junction and, with the proper

parametrization, can exhibit most of the dynamical features of the long junction. In particular a dynamical state in which a train of magnetic flux quanta moves along the array can be realized [9]. This discrete flux flow generates a microwave radiation if the array is properly terminated, and could be employed for the realization of a simple microwave source.

In order to enhance the mutual coupling a much better configuration involves a 2D array of Josephson junctions. Now each junction is close to a relatively large number of other junctions, thus making the mutual phase-locking stronger. However the large number of junctions involved in a 2D array poses severe restrictions on the fabrication technology. An interesting recent experimental investigation [2] reports the coherent emission from 2D arrays of junctions coupled to a common groundplane. A 150 GHz emission is observed with a detected power up $0.12\mu W$ from a 3x230 junctions array. The progressive mutual coupling of the junctions in the array could be obtained by varying the number of active (voltage biased) rows and the expected square law was observed. However an unexplained threshold was also observed for the starting of the phase locking, confirming that in such systems the overall dynamics is quite complex and the number of degrees of freedom quite large.

A completely different type of microwave oscillator can be realized using long Josephson junctions. As we recalled at the beginning of the chapter, a long junction has at least one of its dimensions larger then the Josephson penetration length (λ_j). In this case its electrodynamics is described by a perturbed form of the sine-Gordon equation. It is well known that the sine-Gordon equation admits solitonic solutions. These correspond to magnetic field flux quanta (fluxons) that are trapped in the junction barrier and can move along the junction, pushed by dc the bias current [26]. With the use of an appropriate dc magnetic field, fluxons can be continuously injected in the junction at one edge, accelerated by the bias current toward the opposite edge and here absorbed by a suitable load. As a result a train of fast (few ps wide) voltage pulses can be observed at the load and possibly used signal source. The pulse width and spacing can be, to some extend, modulated by the dc-bias current and magnetic field, thus providing a simple means to tune the generated frequency. This type of oscillator, called Flux Flow Oscillator (FFO), has proven to be a reliable wideband and easy tunable oscillator suitable for integration with a SIS mixer in a single-chip sub-mm wave receiver up to $500GHz$ [22, 23, 24, 25, 34, 30]. The remaining problems for this application are related to the control of the signal linewidth, which should be better than 1 ppm of the oscillator frequency, for which has been recently proposed a solution based on the use of an external precision source, a PLL and an additional mixer stage [16].

8 Conclusions

In concluding this short review of some of the applications of superconductivity we wish to make few considerations on the cryogenic and market aspects.

Since the beginning of the development of Josephson junction devices in the early sixties, it has been shown many times that superconductive circuits can achieve performances beyond the reach of competing technologies. Even compared with the exponential growth of semiconductor circuits speed and complexity, the comparatively simpler Josephson circuits keep concrete, although nowadays substantially reduced, performance margins.

However, with the possible only exception of SQUID biomagnetic systems, where there is no competing technology, the superconductive electronics has never really become a marketable technology. The main problem lies not in the performances, but in the practical, and sometimes psychological, problems related to the cryogenic requirements. It is simply not possible to envisage a large scale product which requires liquid helium cooling, with the accompanying problem of periodic refilling the storage dewar. Indeed the only example of such technology is given by the MRI systems, where again there is no substitute for superconductive magnets, which however are large and expensive systems and have reached a relatively good level of user-friendliness.

The recent improvements in the technology of cryocoolers could change this picture by providing simple electrically powered cryogenic cooling systems that could eliminate the need for liquid cooling. If, or when, compact cooling systems is developed, there will be a chance of seeing portable instruments based on superconductive circuits, which will make superconductive electronics a reality.

Acknowledgments: The authors wish to acknowledge Dr. J. Niemeyer and Dr. M. Russo for providing photographic material of their devices.

9 References

[1] A. Barone. Why superconductive detectors? *Nucl. Phys. B*, 44:645–666, 1995.

[2] P. Barbara, A.B. Cawthorne, S.V. Shitov, and C.J. Lobb. Stimulated emission and amplification in Josephson junction arrays. *Phys. Rev. Lett.* to be published 1999.

[3] A. Barone and G. Paternò. *Physics and Applications of the Josephson Effect.* J. Wiley & Sons, New York, 1982.

[4] A. Barone and S. De Stefano. More on the possibility of nuclear radiation detectors by superconductors. *Nucl. Instrum & Methods*, 202:513–514, 1982.

[5] M. Cirillo, F. Mueller, J. Niemeyer, and R. Poepel. Coupling of Josephson radiation to voltage standard arrays. *J. of Supercond.*, 1999. to be published.

[6] S. Cooper, editor. *Proceedings of Seventh International Workshop on Low Temperature Detectors*. Max Planck Institute of Physics, Munich, 1997.

[7] M. Dorojets, P. Bunyk, D. Zinoviev, and K.K. Likharev. Petaflops RSFQ System Design. In *IEEE Trans Appl. Superc.*, Proc. Appl. Supercond. Conference, Palm Desert, Ca, 1998. to be published 1999.

[8] For a more comprehensive review and an extensive bibliography the reader is addressed to the special issue of the IBM Journal of Research and Development (v.24, n2, March 1980).

[9] G. Filatrella, S. Matarazzo, and S. Pagano. Fluxon dynamics in discrete sine-Gordon systems. In P.L.Christiansen J.C. Eilbeck and R.D. Parmentier, editors, *Future Directions on Nonlinear Dynamics in Physical and Biological Systems*, Plenum Press, New York (1993) pp. 347-350.

[10] C.A. Hamilton and C.J. Burroughs. Performance and reliability of NIST 10-V Josephson arrays. *IEEE Trans. Instrum. Measur.*, 44:238–240, 1995.

[11] L.B. Holdeman, B.F. Field, J. Toots, and C.C. Chang. Prototype for a commercial Josephson-effect voltage standard. *FTSE*.

[12] *IEICE Trans. Electron.*, E81(10), 1998. Special Issue on Low and High-Temperature Superconductive Electron Devices and Their Applications.

[13] W. Julzi, T.O. Mohr, M. Gasser, and H.P.Gschwind. Josephson junctions with $1\mu m$ dimensions and with picosecond switching times. *Electron Lett.*, 8:589–591, 1972.

[14] B.D. Josephson. Possible new effects in superconductive tunneling. *Phys. Lett.*, 1:257–253, 1962.

[15] R.L. Kautz and R. Monaco. survey of chaos in the rf-biased Josephson junction. *J. Appl. Phys.*, 57:875, 1985.

[16] V.P. Koshelets, S.V. Shitov, A.V. Shchukin, L.V. Filippenko P.V. Dmitriev, V.L. Vaks, J. Mygind, A.B. Baryshev, W. Luinge, and H. Golstein. Flux-flow oscillators for sub-mm wave integrated receivers. In *IEEE Trans Appl. Superc.*, Proc. Appl. Supercond. Conference, Palm Desert, Ca, 1998. to be published 1999.

[17] K.K. Likharev. Superconductors speed up computation. *Physics World*, pages 39–43, May 1997.

[18] K.K. Likharev, D.A. Mukhanov, and V.K. Semenov. In W. de Gruiter, editor, *SQUID'85*, page 1103, Berlin, 1985.

[19] J. Matisoo. Sub-nanosecond pair tunneling to single-particle tunneling transition in Josephson junctions. *Appl. Phys. Lett.*, 9:167–168, 1966.

[20] F. Mueller, R. Poepel, J. Kohlmann, J. Niemeyer, W. Meier, T. Weilmann, L. Grimm, F.W. Duenschede, and P. Gutmann. Optimized 1 V and 10 V Josephson series arrays. *IEEE Trans. Instrum. and Measur.*, 46:229–232, 1997.

[21] Y. Murayama, Y. Sakamoto, A. Iwasa, M. Nakanishi, H. Yoshida, U. Klein, and T. Endo. Ten-Volt Josephson junction array. *IEEE Trans. Instrum. Measur.*, 44:219–222, 1995.

[22] T. Nagatsuma, K. Enpuku, F. Irie, and K. Yoshida. Flux-flow type Josephson oscillator for millimeter and sub-millimeter wave region. *J. Appl. Phys*, 54:3302–3309, 1983.

[23] T. Nagatsuma, K. Enpuku, F. Irie, and K. Yoshida. Flux-flow type Josephson oscillator for millimeter and sub-millimeter wave region: part II. *J. Appl. Phys*, 56:3284, 1984.

[24] T. Nagatsuma, K. Enpuku, F. Irie, and K. Yoshida. Flux-flow type Josephson oscillator for millimeter and sub-millimeter wave region: part III. *J. Appl. Phys*, 58:441, 1985.

[25] T. Nagatsuma, K. Enpuku, F. Irie, and K. Yoshida. Flux-flow type Josephson oscillator for millimeter and sub-millimeter wave region: part IV. *J. Appl. Phys*, 63:1130, 1988.

[26] R. D. Parmentier. Fluxons in long Josephson junctions. In H. Weinstock and R.W. Ralston, editors, *The New Superconducting Electronics*, Dordrect, 1993. Kluwer. and references reported therein.

[27] R. Poepel, J. Niemeyer, R. Fromknecht, W. Meier, and L. Grimm. 1- and 10-V series array Josephson voltage standard in $Nb/Al_2O_3/Nb$ technology. *J. Appl. Phys.*, 68:4294–4303, 1990.

[28] S. Pagano, V.G. Palmieri, A. Esposito, O. Mukhanov, and S. Rylov. First Realization of a Tracking Detector for High Energy Physics Experiments Based on Josephson Digital Readout Circuitry. In *IEEE Trans Appl. Superc.*, Proc. Appl. Supercond. Conference, Palm Desert, Ca, 1998. to be published 1999.

[29] W. H. Parker, B.N. Taylor, and D.N. Langerberg. Measurement of 2e/h using the ac Josephson effect and its implications for quantum electrodynamics. *Phys. Rev. Lett.*, 18:287–291, 1967.

[30] S.V. Shitov, A.B. Ermakov, L.V. Filippenko, V.P. Koshelets, A.B. Baryshev, W. Luinge, and J.R. Gao. Superconductive chip receiver for imaging applications. In *IEEE Trans Appl. Superc.*, Proc. Appl. Supercond. Conference, Palm Desert, Ca, 1998. to be published 1999.

[31] S.P.Benz and C.A. Hamilton. Pulse driven programmable Josephson voltage standard. *Appl. Phys. Lett.*, 68:3171–3173, 1996.

[32] S. Shapiro, A.R. Janus, and S. Holly. Effect of Microwaves on Josephson current in superconducting tunneling. *Rev. Mod. Phys.*, 36:223–225, 1964.

[33] H. Sasaki, S. Kiryu, F. Hirayama, T. Kikuchi, M. Maezawa, and A. Shoji. RFSQ-based D/A converter for AC voltage standard. In *IEEE Trans Appl. Superc.*, Proc. Appl. Supercond. Conference, Palm Desert, Ca, 1998. to be published 1999.

[34] V.P.Koshelets, S.V. Shitov, L.V. Filippenko, A.M. Baryshev, W.Luinge, H. Golstein, H. van de Stadt, J.R. Gao, and T. de Graauw. An integrated 500 GHz receiver with superconducting local oscillator. *IEEE Trans Appl. Supercond.*, 7:3589–3592, 1997.

[35] H. Weinstock, editor. *SQUID sensors: Fundamentals, Fabrication and Applications*. Kluwer Academic Publishers, 1996.

[36] K.L. Wan, A.K. Jain, and J.E. Lukens. Sub-millimeter wave generation using Josephson junction arrays. *IEEE Trans. Mag.*, pages 1076–1079, 1989.

[37] H. Weinstock and M. Nisenoff, editors. *Superconductive Electronics*. Springer - Verlag, Berlin, 1989.

4

Josephson Flux-Flow Oscillators in Microwave Fields

Mario Salerno
Mogens Samuelsen

ABSTRACT In this chapter we investigate, both analytically and numerically, the flux-flow resonance of long Josephson junctions in the presence of microwave fields. We consider the microwaves to be coupled to the junction both uniformly in space and through boundary conditions. Using a perturbation analysis around the rotating background we derive analytical expressions for the current-voltage characteristics. The dependence of the flux-flow step on the amplitude of the rf-field and the appearance of satellite steps are explained. As a result we find that for uniform microwave fields the satellite steps are spaced around the main flux-flow resonance by both odd and even harmonics of the rf-frequency, while for non-uniform fields only the even harmonic satellites are present. These results are in good agreement with direct numerical integration of the system.

1 Introduction

In the past years a great deal of interest have been devoted to the study of the Josephson flux-flow oscillator, i.e. a long Josephson junction operating in the flux-flow regime [1]-[10]. The interest in such a device resides in its relatively high output power (of the order of μW at .4 - 1 THz) [1, 7], wide bandwidth, and easy tunability, these being attractive features for applications in superconducting millimeter-wave electronics [3, 5, 8, 9]. The flux-flow regime is achieved when a sufficiently large external magnetic field is applied to the junction so that an unidirectional motion of fluxons, created at one edge and destroyed at the other, is generated. This dynamics gives rise in the current-voltage (IV) characteristics to an high voltage step (flux-flow step) split into a series of equally spaced Fiske [11] substeps. Recently, an analytical description of these resonances in the absence of external microwave fields was provided [10]. On the other hand, it is known that the application of a microwave field to the junction creates new resonances (satellites) around the main flux-flow resonance [4, 6]. Although this phenomenon was observed both numerically and experimen-

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 87−101, 2000.

tally, a theoretical explanation of it is still lacking.

The aim of this chapter is to provide a simple theory that accounts for the appearance of satellite steps in the IV- characteristics of Josephson flux-flow oscillators. To this end we use a perturbative expansion [12] around the uniform rotating background to derive analytical expressions for the IV-characteristics in the presence of both uniform and non-uniform microwave fields. As a result we show that for microwave fields uniformly applied to the junction, flux-flow satellite steps appear at both even and odd harmonics of the rf-frequency, while for non-uniform microwave fields applied at the edges of the junction, only even satellites steps apparently appear. Moreover, the flux-flow and the satellite steps can be reduced or enhanced by changing the amplitude of the microwave field. We check these results by comparing our analytical expressions for the IV-characteristics with direct numerical integrations of the system, finding an excellent agreement between theory and experiment.

The content of the chapter is summarized as follows. In section 2 we consider the case of flux-flow oscillators in the presence of uniform microwave fields and use a first order perturbation theory to derive the analytical expression for the IV-curve. We show that satellite steps around the main flux-flow resonance are spaced by both odd and even harmonics of the rf-frequency. In section 3 we consider the case of microwave fields applied to the junction through boundary conditions. We show that in this case, only even satellite steps appear in the IV-characteristics. In section 4 we compare our analytical results with direct numerical simulations of the perturbed sine-Gordon system. Finally, in the conclusions we summarize the main results of the chapter.

2 Flux-flow oscillators in uniform microwave fields

The electrodynamics of a Josephson junction in the presence of both microwaves and external magnetic fields is described by the perturbed sine-Gordon equation [11, 13]

$$\Phi_{xx} - \Phi_{tt} = \sin(\Phi) + \alpha\Phi_t - \eta + \eta_{rf}\sin(\Omega t - \theta_0), \tag{4.1}$$

subject to the boundary conditions

$$\begin{aligned} \Phi_x(0,t) &= \Gamma + \Gamma_1\sin(\Omega t), \\ \Phi_x(L,t) &= \Gamma + \Gamma_2\sin(\Omega t). \end{aligned} \tag{4.2}$$

In Eq. (4.1) space and time have been normalized to the Josephson penetration length λ_J and to the inverse plasma frequency ω_0^{-1}, respectively. α denotes the loss parameter associated to the quasiparticle tunneling, η represent the dc-bias current and Γ is the normalized external magnetic

field. Note that the rf-field can be applied to the junction either uniformly i.e. $\eta_{rf} \neq 0$, $\Gamma_1 = \Gamma_2 = 0$, or non-uniformly through boundary conditions i.e. $\eta_{rf} = 0$, $\Gamma_1 \neq 0, \Gamma_2 \neq 0$.

In this section we consider the case of uniform coupling, so we fix $\eta_{rf} \neq 0$ and $\Gamma_1 = \Gamma_2 = 0$ in Eqs. (4.1) and (4.2). Since the flux-flow regime is characterized by excitations which travel on top of a fast rotating background, it is natural to assume a solution of Eq. (4.1) of the form

$$\Phi = \Phi_0 + \Psi(x,t), \tag{4.3}$$

where

$$\Phi_0 = \omega t + \Gamma x + \beta \sin(\Omega t) + \theta_1 \tag{4.4}$$

represents the rotating background field with frequency ω, on which is superimposed a uniform oscillation with the same frequency of the rf-field. Here $\Psi(x,t)$ is a small modulation ($\Psi \ll 1$) around Φ_0 and θ_1 is an arbitrary phase. We see that the boundary conditions in Eq. (4.2) (with $\Gamma_1 = \Gamma_2 = 0$) are fulfilled if $\Psi(x,t)$ satisfies

$$\Psi_x(0,t) = \Psi_x(L,t) = 0. \tag{4.5}$$

In the following we look for fields Ψ which satisfy Eq. (4.5) and which have both zero space and time averages $< \Psi >= \overline{\Psi} = 0$ (here and in the following $<>$ and overline denote respectively time and space averages). Inserting the expansion Eq. (4.3) into Eq. (4.1) and using the smallness of Ψ, we obtain the following linearized equation

$$\begin{aligned} \Psi_{xx} - \Psi_{tt} - \alpha\Psi_t &= \alpha\omega - \eta + (\eta_{rf}\cos(\theta_0) - \beta\Omega^2)\sin(\Omega t) - \\ &\quad (\eta_{rf}\sin(\theta_0) - \alpha\beta\Omega)\cos(\Omega t) + \sin(\Phi_0) + \cos(\Phi_0)\Psi. \end{aligned} \tag{4.6}$$

To eliminate from this equation the explicit dependence on the frequency Ω we choose

$$\beta = \frac{\eta_{rf}}{\Omega\sqrt{\Omega^2 + \alpha^2}}, \qquad \tan(\theta_0) = \frac{\alpha}{\Omega}. \tag{4.7}$$

Note that this elimination is exact and puts no restriction on the size of η_{rf}. By expanding the last two terms of Eq. (4.6), using the Bessel relation

$$e^{i\beta\sin(\Omega t)} = \sum_{m=-\infty}^{\infty} J_m(\beta) e^{im\Omega t}, \tag{4.8}$$

and using Eqs. (4.4) and (4.7), we can rewrite Eq. (4.6) in the form

$$\begin{aligned} \Psi_{xx} - \Psi_{tt} - \alpha\Psi_t &= \alpha\omega - \eta + \\ &\quad \sum_{m=-\infty}^{\infty} J_m(\beta)(\sin(\Gamma x + \tilde{\omega}_m t + \theta_1)) + \\ &\quad \sum_{m=-\infty}^{\infty} J_m(\beta)(\cos(\Gamma x + \tilde{\omega}_m t + \theta_1))\Psi, \end{aligned} \tag{4.9}$$

where

$$\tilde{\omega}_m = \omega + m\Omega. \tag{4.10}$$

Note that the dependence on the amplitude of the rf-field is in the argument of the Bessel function $J_m(\beta)$ and that the frequency ω is shifted to $\tilde{\omega}_m = \omega + m\Omega$. To solve Eq. (4.9) we expand the function Ψ as a double Fourier series of the form

$$\Psi(x,t) = \sum_{n=0,m=-\infty}^{\infty} (A_{n,m}\cos(\tilde{\omega}_m t) + B_{n,m}\sin(\tilde{\omega}_m t))\cos(k_n x), \tag{4.11}$$

with $k_n = \frac{\pi}{L}n$, so that the boundary conditions in Eq. (4.5) are automatically satisfied. By substituting this expression in Eq. (4.9) and projecting along the k_n, $\tilde{\omega}_m$ modes we obtain

$$A_{n,m} = \frac{2}{1+\delta_{n,0}} \frac{(\tilde{\omega}_m^2 - k_n^2)\tilde{I}_c(n,m) + \alpha\tilde{\omega}_m\tilde{I}_s(n,m)}{(\tilde{\omega}_m^2 - k_n^2)^2 + \alpha^2\tilde{\omega}_m^2}, \tag{4.12}$$

$$B_{n,m} = \frac{2}{1+\delta_{n,0}} \frac{(\tilde{\omega}_m^2 - k_n^2)\tilde{I}_c(n,m) - \alpha\tilde{\omega}_m\tilde{I}_s(n,m)}{(\tilde{\omega}_m^2 - k_n^2)^2 + \alpha^2\tilde{\omega}_m^2}, \tag{4.13}$$

where

$$\begin{aligned} \tilde{I}_s(n,m) &= J_m(\beta)\Gamma L \frac{(\cos(\theta_1) - \cos(k_n L)\cos(\Gamma L + \theta_1))}{(\Gamma^2 L^2 - k_n^2 L^2)}, \\ \tilde{I}_c(n,m) &= J_m(\beta)\Gamma L \frac{(\sin(\Gamma L + \theta_1)\cos(k_n L) - \sin(\theta_1))}{(\Gamma^2 L^2 - k_n^2 L^2)}. \end{aligned} \tag{4.14}$$

Assuming $\tilde{\omega}_m \neq 0$ the dc- part of Eq. (4.9) is computed as

$$\eta - \alpha\omega = \overline{< \Psi\cos(\Phi_0) >}, \tag{4.15}$$

with the double averaged term (which is different from zero mainly for $\tilde{\omega}_m \neq 0$) given by

$$\overline{< \Psi\cos(\Phi_0) >} = \frac{1}{2L}\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}(A_{n,m}\tilde{I}_c(n,m) - B_{n,m}\tilde{I}_s(n,m)). \tag{4.16}$$

The IV-characteristics is then obtained from the balance between the dc-part of Eq. (4.16) and $\alpha\omega - \eta$, this giving

$$\eta = \alpha\omega + \frac{1}{2}\sum_{n,m=-\infty}^{\infty} J_m^2(\beta)\frac{(\frac{\Gamma L}{2})^2}{(\frac{\Gamma L + k_n L}{2})^2}\frac{\sin^2(\frac{\Gamma L - k_n L}{2})}{(\frac{\Gamma L - k_n L}{2})^2}\frac{\alpha\tilde{\omega}_m}{(\tilde{\omega}_m^2 - k_n^2)^2 + \alpha^2\tilde{\omega}_m^2}. \tag{4.17}$$

In deriving this expression we have substituted Eq. (4.16) into Eq. (4.15) and used Eqs. (4.12), (4.13), and (4.14), together with the assumption $\tilde{\omega}_m \neq 0$. It is interesting to note that Eq. (4.17) is independent on the phase angle θ_1 and it has vertical asymptotes for $\omega = m\Omega$, m=0,1,2... . We also remark that although Eq. (4.17) is derived under the assumption $\omega \neq m\Omega$, the above singularities reflect the presence of phase locking [15]. One can indeed show that the weight of the singularity at $\omega = m\Omega$ in Eq. (4.17) is proportional to the square of the locking range in current of the m-th phase locked step and that the vertical distance between the two hyperbolic-like branches of the singularity is equal to the locking range [16]. Finally, we note that for $\eta_{rf} = 0$ is $J_m(\beta) = \delta_{m,0}$ and the above expression for the IV-curve reduces to the one derived in Refs. [10, 14] in absence of microwaves.

3 Flux-flow oscillators in non-uniform microwave fields

In this section we consider the case of a Josephson flux-flow oscillator coupled to the microwave field through the boundary conditions. The model is now

$$\Phi_{xx} - \Phi_{tt} = \sin(\Phi) + \alpha\Phi_t - \eta, \tag{4.18}$$

with

$$\begin{aligned} \Phi_x(0,t) &= \Gamma + \Gamma_1 \sin(\Omega t), \\ \Phi_x(L,t) &= \Gamma + \Gamma_2 \sin(\Omega t). \end{aligned} \tag{4.19}$$

We assume a solutions for the field equation Eq. (4.18) of the form

$$\Phi = \omega t + \Gamma x + f(x)\sin(\Omega t) + g(x)\cos(\Omega t) + \Psi(x,t) + \theta_1, \tag{4.20}$$

where Ψ is, as before, a small field ($\Psi \ll 1$) with zero averages satisfying the boundary conditions in Eq. (4.5), θ_1 is an arbitrary phase, and $f(x)$, $g(x)$ are unknown functions which we assume to satisfy

$$\begin{aligned} g'(0) = g'(L) = 0, \\ f'(0) = \Gamma_1, \quad f'(L) = \Gamma_2, \end{aligned} \tag{4.21}$$

(here prime denotes the x-derivative). Note that these boundary conditions are consistent with the ones in Eq. (4.19) if the field Ψ satisfies Eq. (4.5). Inserting Eq. (4.20) into Eq. (4.18) and using the smallness of Ψ we obtain the following linearized equation

$$\begin{aligned} \Psi_{xx} - \Psi_{tt} - \alpha\Psi_t = \alpha\omega - \eta \\ -(f''(x) + f(x)\Omega^2 + \alpha g(x)\Omega)\sin\Omega t \\ -(g''(x) + g(x)\Omega^2 - \alpha f(x)\Omega)\cos\Omega t \\ +\sin(\Gamma x + \omega t + f(x)\sin\Omega t + g(x)\cos\Omega t + \theta_1) \\ +\cos(\Gamma x + \omega t + f(x)\sin\Omega t + g(x)\cos\Omega t + \theta_1)\Psi. \end{aligned} \tag{4.22}$$

To simplify this equation we eliminate the explicit Ω dependence by taking $f(x)$ and $g(x)$ to be solutions of the system

$$\begin{array}{c} f''(x) + \Omega^2 f(x) + \alpha\Omega g(x) = 0, \\ g''(x) + \Omega^2 g(x) - \alpha\Omega f(x) = 0. \end{array} \tag{4.23}$$

In the following we assume the existence of the functions $f(x)$ and $g(x)$, leaving the solution of the system Eq. (4.23) with the boundary conditions in Eq. (4.21) to the Appendix. Knowing $f(x)$ and $g(x)$, we can rewrite Eq. (4.22) in the form

$$\begin{array}{c} \Psi_{xx} - \Psi_{tt} - \alpha\Psi_t = \alpha\omega - \eta \\ + \sin(\Gamma x + \omega t + h(x)\sin(\Omega t + \phi(x)) + \theta_1) \\ + \cos(\Gamma x + \omega t + h(x)\sin(\Omega t + \phi(x)) + \theta_1)\Psi, \end{array} \tag{4.24}$$

with $h(x)$, $\phi(x)$, given by

$$h(x) = \sqrt{f(x)^2 + g(x)^2}, \quad \tan\phi(x) = \frac{g(x)}{f(x)}. \tag{4.25}$$

Note that the derivation of Eq. (4.24) is exact and there are no restriction on the size of Γ_1 and Γ_2. Using the Bessel identity of Eq. (4.8) we can rewrite Eq. (4.24) as

$$\begin{array}{c} \Psi_{xx} - \Psi_{tt} - \alpha\Psi_t = \alpha\omega - \eta \\ + \sum_m J_m(h(x))(\sin(\Gamma x + m\phi(x))\cos(\tilde{\omega}_m t + \theta_1) \\ + \cos(\Gamma x + m\phi(x))\sin(\tilde{\omega}_m t + \theta_1)) \\ + \sum_m J_m(h(x))(\cos(\Gamma x + m\phi(x))\cos(\tilde{\omega}_m t + \theta_1) \\ - \sin(\Gamma x + m\phi(x))\sin(\tilde{\omega}_m t + \theta_1))\Psi, \end{array} \tag{4.26}$$

with $\tilde{\omega}_m$ given by Eq. (4.10). To solve this equation we expand the function Ψ as a double Fourier series of the form

$$\Psi = \sum_{nm}(C_{nm}\cos(\tilde{\omega}_m t + \theta_1) + D_{nm}\sin(\tilde{\omega}_m t + \theta_1))\cos k_n x, \tag{4.27}$$

with $k_n = \frac{\pi}{L}n$, so that the boundary conditions in Eq. (4.5) are automatically satisfied. By substituting Eq. (4.27) into Eq. (4.26) and projecting along the k_n, $\tilde{\omega}_m$, modes we get

$$C_{nm} = \frac{(\tilde{\omega}_m^2 - k_n^2)\tilde{S}(n,m) + \alpha\tilde{\omega}_m\tilde{C}(n,m)}{((\tilde{\omega}_m^2 - k_n^2)^2 + \alpha^2\tilde{\omega}_m^2)}, \tag{4.28}$$

$$D_{nm} = \frac{(\tilde{\omega}_m^2 - k_n^2)\tilde{C}(n,m) - \alpha\tilde{\omega}_m\tilde{S}(n,m)}{((\tilde{\omega}_m^2 - k_n^2)^2 + \alpha^2\tilde{\omega}_m^2)}, \tag{4.29}$$

where

$$\tilde{C}(n,m) = \frac{1}{L}\int_0^L \cos(k_n x)J_m(h(x))\cos(\Gamma x + m\phi(x))dx, \tag{4.30}$$

$$\tilde{S}(n,m) = \frac{1}{L}\int_0^L \cos(k_n x)J_m(h(x))\sin(\Gamma x + m\phi(x))dx. \tag{4.31}$$

The IV-characteristics then follows from the dc-part of Eq. (4.26) as

$$\begin{aligned}\eta &= \alpha\omega + \frac{1}{2}\sum_{n,m}(C_{nm}\tilde{C}(n,m) - D_{nm}\tilde{S}(n,m)) \\ &= \alpha\omega + \frac{1}{2}\sum_{n,m}(\tilde{C}(n,m)^2 + \tilde{S}(n,m)^2)\frac{\alpha\tilde{\omega}_m}{(\tilde{\omega}_m^2 - k_n^2)^2 + \alpha^2\tilde{\omega}_m^2}.\end{aligned} \tag{4.32}$$

(Note that in deriving Eq. (4.32) we used $\tilde{\omega}_m \neq 0$). This expression, although exact, is quite complicated to analyze and in general it must be numerically evaluated. To get a physical understanding of Eq. (4.32), however, we can simplify it by assuming

$$\Omega \ll \omega, \Gamma. \tag{4.33}$$

In this case the integrals $\tilde{C}$ and $\tilde{S}$ in Eq. (4.32) can be approximated as

$$\begin{aligned}\tilde{C}(n,m) &\simeq C(n,\Gamma)B(m,C) - S(n,\Gamma)B(m,S), \\ \tilde{S}(n,m) &\simeq S(n,\Gamma)B(m,C) + C(n,\Gamma)B(m,S),\end{aligned} \tag{4.34}$$

with

$$\begin{aligned}B(m,C) &= \tfrac{1}{L}\textstyle\int_0^L J_m(h(x))\cos(m\phi(x))dx, \\ B(m,S) &= \tfrac{1}{L}\textstyle\int_0^L J_m(h(x))\sin(m\phi(x))dx,\end{aligned} \tag{4.35}$$

and

$$\begin{aligned}C(n,\Gamma) &= \tfrac{\Gamma L\cos(k_nL)\sin(\Gamma L)}{(\Gamma L)^2-(k_nL)^2}, \\ S(n,\Gamma) &= \tfrac{\Gamma L(1-\cos(k_nL)\cos(\Gamma L))}{(\Gamma L)^2-(k_nL)^2}.\end{aligned} \tag{4.36}$$

Eq. (4.32) can be then rewritten as

$$\begin{aligned}\eta = \alpha\omega + \tfrac{1}{2}\textstyle\sum_{n,m}(C(n,\Gamma)^2 + S(n,\Gamma)^2)\times \\ (B(m,C)^2 + B(m,S)^2)\times \\ \tfrac{\alpha\tilde{\omega}_m}{((\tilde{\omega}_m^2-k_n^2)^2+\alpha^2\tilde{\omega}_m^2)}.\end{aligned} \tag{4.37}$$

A further simplification is achieved by observing that the integrals in Eq. (4.35) can be put in the form [17]

$$\begin{aligned}B(m,C) &= \tfrac{1}{L}\textstyle\int_0^L\sum_M(-1)^M J_{m+2M}(f(x))J_{2M}(g(x))dx, \\ B(m,S) &= \tfrac{1}{L}\textstyle\int_0^L\sum_M(-1)^M J_{m+2M+1}(f(x))J_{2M+1}(g(x))dx.\end{aligned} \tag{4.38}$$

From the appendix we have that $f(x)$ and $g(x)$ are almost harmonic functions and they oscillate many times in the interval $[0, L]$. This implies that the integrals in Eq. (4.38) for odd values of m are very small, i.e.

$$B(2m+1,C) \simeq 0, \quad B(2m+1,S) \simeq 0. \tag{4.39}$$

Using Eq. (4.39) together with the relations

$$B(-2m,C) = B(2m,C), \quad B(-2m,S) = -B(2m,S), \quad B(0,S) = 0, \tag{4.40}$$

we can finally approximate the IV-curve as

$$\eta \simeq \alpha\omega + \sum_{n,m=-\infty}^{\infty} \frac{1}{2}(B(2m,C)^2 + B(2m,S)^2) \times \frac{(\Gamma L)^2 \sin^2(\Gamma L - k_n L)/2)}{(\Gamma L + k_n L)^2 (\Gamma L - k_n L)/2)^2} \times \frac{\alpha(\omega + 2m\Omega)}{((\omega + 2m\Omega)^2 - k_n^2)^2 + \alpha^2(\omega + 2m\Omega)^2}. \tag{4.41}$$

Note that the sum on n gives rise, as before, to Fiske resonances spaced by $\frac{\pi}{L}$ with the Fraunhofer factor enhancing the ones for which k_n is close to $\Gamma + M\Omega$, $M = 0, 2, 4, \ldots$. It is remarkable that, in spite of the different coupling mechanism, the IV-curve has almost the same structure as in the uniform case (see Eq. (4.17)). In the present case, however, the sum is only over the even $M = 2m$ values, this implying that in the IV-curve satellite steps appear at voltages $\omega = \Gamma + M\Omega$, i.e. they are spaced by even harmonics of the rf-frequency (note that it is the spatial variation of $f(x)$ and $g(x)$ which average out the odd's satellites).

4 Numerical experiment

In order to check the analytical expressions for the IV- characteristics derived above, we have performed a direct numerical integration of Eq. (4.1) subject to the boundary conditions in Eq. (4.2). For fixed values of magnetic field, damping constant, amplitude and frequency of the rf-field and length of the junction, we computed the normalized average voltage across the junction $V = \overline{< \Phi_t >} \equiv \omega$ as a function of the bias current, taking as initial condition $n = \frac{\Gamma L}{2\pi}$ fluxons equally spaced along the junction. The numerical IV- characteristics was obtained by integrating Eq. (4.1) long enough to eliminate all transients and measuring for each value of the bias current the corresponding average voltage. To trace the IV-curves the bias current was increased in small steps from $\eta = 0$ to $\eta = 1$ and then back to zero. At each η-step the final configuration of the field in the junction was used as initial condition for the next η-step. In Fig. 1 we report the IV-characteristics of a long Josephson junction in the presence of an uniform microwave field of amplitude $\eta_{rf} = 3.0$ and without rf-fields at the boundaries ($\Gamma_1 = \Gamma_2 = 0$ in Eq. (4.2)), for parameter values $\alpha = 0.1, L = 15, \Gamma = 6$, $\Omega = 1.4$. From this figure we see the presence of a flux-flow step at $\omega = \Gamma$ and two satellite steps at $\omega = \Gamma \pm \Omega$, both consisting of few Fiske substeps spaced by $\frac{\pi}{L}$ as predicted by our analysis. Small satellite steps at $\omega = \Gamma \pm 2\Omega$ are also visible. Here the thin curve refers to the analytical expression in Eq. (4.17) while the thick one represents the numerical results. The insets of Fig. 1 show the first two satellite steps at $\omega = \Gamma \pm \Omega$ in more details from which we

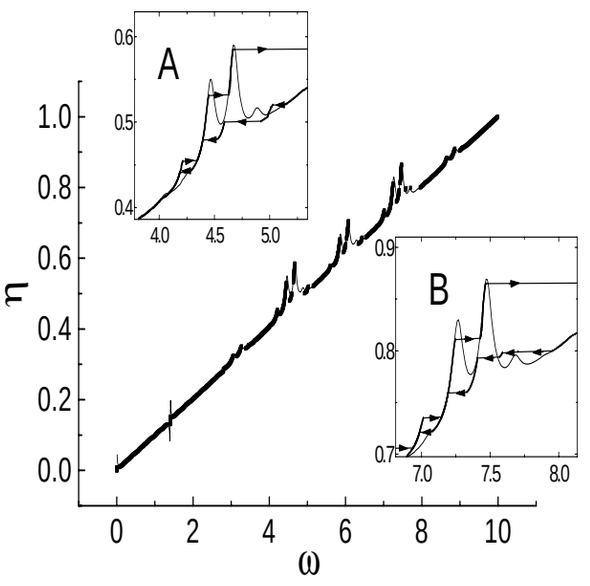

FIGURE 1. IV-characteristics of a long Josephson junction in the presence of an uniform microwave field of amplitude $\eta_{rf} = 3.0$ and frequency $\Omega = 1.4$ and with zero rf-fields at the boundaries ($\Gamma_1 = \Gamma_2 = 0$ in Eq. (4.2)), for parameter values, $\alpha = 0.1, L = 15, \Gamma = 6$. The insets A and B show an enlargement of the satellite steps at $\omega = \Gamma - \Omega$ and $\omega = \Gamma + \Omega$, respectively. The thin curve refers to the analytical expression in Eq. (4.17) while the thick curve is obtained from numerical integrations of Eq. (4.1).

also see the appearance of hysteretic phenomena (arrows show the direction of switching). Note that two of the numerical Fiske substeps appear broken due to our unidirectional procedure of varying the bias current. From this figure it is evident that, except for stability problems not included in our analysis, the two curves overlap in most part of the plot giving an excellent agreement between analytical and numerical results (note that there is no overlapping with the part of the theoretical curve which has negative slope because of its instability). By increasing the amplitude of the microwave field, Eq. (4.17) predicts a $J_m(\beta)^2$ modulation of the flux-flow resonance with a total suppression of the step for appropriate values of η_{rf}. This is actually what we find, as reported in Fig. 2. In this figure the numerical and the analytical IV-curves are shown for the same parameter values of Fig. 1 but for $\eta_{rf} = 5.0$. We see that the satellite steps are still present but the main resonance at $\omega = \Gamma$ has disappeared. Comparing Fig. 1 with Fig. 2 we see that the agreement between numerical and theoretical curves

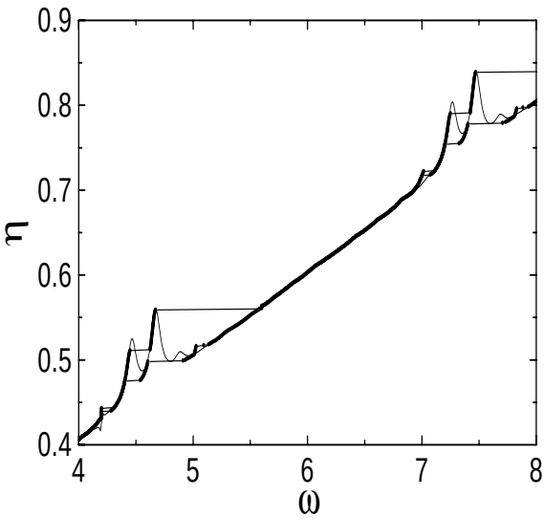

FIGURE 2. Same as in Fig. 2 but for rf-field amplitude $\eta_{rf} = 5.0$.

improves by increasing the amplitude of the microwave field. This can be understood from the fact that by increasing η_{rf} one reduces the size of the flux-flow resonances so that the deviation of the IV-characteristics from the ohmic line (which is a measure of Ψ), becomes small. The number of satellite steps that can appear in the IV-characteristics increases with the amplitude of the rf-field. This suggests the possibility of using the flux-flow oscillator as a mixer device for high frequency electronics (i.e. one can pump a signal at frequency Ω and detect it at frequency $\Gamma \pm m\Omega$). In the case of non-uniform microwave fields applied at the edges of the junction, the behavior is quite different. In the following we shall consider only the symmetric coupling ($\Gamma_1 = \Gamma_2$) since the asymmetric one ($\Gamma_1 = -\Gamma_2$) gives qualitatively similar results. In Fig. 3 we report the IV-characteristics of a long Josephson junction with the same parameter values as in Fig. 1 but in the presence of a symmetric rf-field at the boundaries of amplitude $\Gamma_1 = \Gamma_2 = 3.0$ (with $\eta_{rf} = 0$ in Eq. (4.1)). In this figure the thin line refers to the numerical evaluation of Eq. (4.32) while the thick one represents the numerical integration of Eq. (4.1). From this figure we see a flux-flow step at $\omega = \Gamma$ and two satellite steps at $\omega = \Gamma \pm 2\Omega$. Note that there are no satellite steps at $\omega = \Gamma \pm \Omega$ and that the resonances are split into Fiske substeps spaced by $\frac{\pi}{L}$ as predicted by our analysis. We also note the presence of a strong phase locked step interacting with a Fiske resonance (for uniform

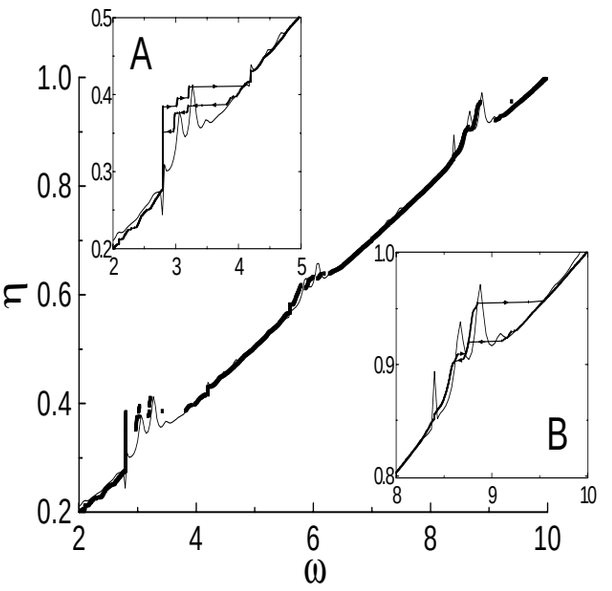

FIGURE 3. The IV-characteristics of a long Josephson junction in the presence of a non-uniform symmetric microwave field at the boundaries of amplitude $\Gamma_1 = \Gamma_2 = 3.0$, frequency $\Omega = 1.4$ and without rf-field along the junction ($\eta_{rf} = 0$). The parameter values of the junction are the same as in Fig. 1. Insets A, B show an enlargement of the satellite steps at $\omega = \Gamma - 2\Omega$ and $\omega = \Gamma + 2\Omega$, respectively. The thin curve refers to the analytical expression in Eq. (4.32) while the thick line is obtained from numerical integrations of Eq. (4.1).

couplings this interaction is absent (see Figs. 1 and 2) due to the weakness of phase lock [18]). This is evident from the inset A of Fig. 3 in which we see a big vertical step at $\omega = 2\Omega$ close to the flux-flow satellite at $\omega = \Gamma - 2\Omega$. Note that the phase locked step is asymmetric around the ohmic line (the upper branch is much bigger that the lower one) and its distance from the Fiske sub-step in the satellite is about $\frac{\pi}{L}$. This suggests that the asymmetric step arises from the rectification and amplification of the Fiske sub-step in the satellite, due to its overlapping with a phase locked resonance (note that in Fig. 3 small symmetric phase locked steps are also visible). This also suggests that, by properly tuning the dc-magnetic field, the length of the junction and the applied rf-frequency, one can create large asymmetric phase lock steps in the IV-characteristics of a long Josephson junction: an interesting feature for practical applications. Finally, in the inset B of Fig. 3 the satellite step at $\omega = \Gamma + 2\Omega$ is reported in more detail. We remark that although the agreement between the analysis and the numerical experiment

is good, it is not as good as in the case of uniform coupling. This is partially due to the lack of a closed exact analytical expression and to the difficulty to compute numerically the sums in Eq. (4.32).

5 Conclusions

We have presented a simple theory for long Josephson flux-flow oscillators in the presence of microwave fields, which accounts for the appearance of satellite steps around the main flux-flow resonance. We derived analytical expressions for the IV-characteristics for both uniform and non-uniform couplings of the microwaves to the junction. As a result we showed that: i) for uniform coupling satellite steps are spaced by both even and odd harmonics of the rf-frequency, ii) for non-uniform couplings (through boundary conditions) only the even satellite steps are present in the IV-characteristics. This different behavior of the system can in turn be useful to distinguish the type of microwave coupling realized in a real experiment. These results were shown to be in good agreement with direct numerical integrations of the system, confirming the validity of our approach.

Acknowledgments: We wish to dedicate this work to the memory of Prof. R. D. Parmentier. We had the pleasure to work with him on several problems related to Josephson junctions. We like to think that, if still alive, he would have enjoyed this paper.

One of us (MS) wishes to acknowledge financial support from the INFM (Istituto Nazionale di Fisica della Materia).

6 Appendix

In this Appendix we show how to solve system Eq. (4.23) together with the boundary conditions in Eq. (4.21). To this end we note that system Eq. (4.23) is equivalent to the fourth order differential equation

$$\left((\partial_x^2 + \Omega^2)^2 + \alpha^2\Omega^2\right) F(x) = 0. \tag{4.42}$$

It is then convenient to shift the x-coordinate $z = x - L/2$ and use the symmetric and the asymmetric (in the coordinate z) solutions of Eq. (4.42) which, in vector notation, can be written as

$$\mathbf{s}(z) = \begin{pmatrix} s_1(z) \\ s_2(z) \end{pmatrix} = \begin{pmatrix} \cos\Omega_1 z \cosh\frac{\alpha_1 z}{2} \\ \sin\Omega_1 z \sinh\frac{\alpha_1 z}{2} \end{pmatrix}, \tag{4.43}$$

$$\mathbf{a}(z) = \begin{pmatrix} a_1(z) \\ a_2(z) \end{pmatrix} = \begin{pmatrix} \cos\Omega_1 z \sinh\frac{\alpha_1 z}{2} \\ \sin\Omega_1 z \cosh\frac{\alpha_1 z}{2} \end{pmatrix}. \tag{4.44}$$

We can express the solution of the general problem in Eqs. (4.23) and (4.21) as a linear combination of the solutions f_s, g_s, f_a, g_a satisfying the boundary conditions

$$g_s'(\pm\frac{L}{2}) = 0, \qquad g_a'(\pm\frac{L}{2}) = 0, \tag{4.45}$$

$$f_s'(\pm\frac{L}{2}) = \Gamma_0, \qquad f_a'(\pm\frac{L}{2}) = \pm\Gamma_0. \tag{4.46}$$

Introducing the matrix

$$\mathbf{M} = \begin{pmatrix} \frac{\alpha_1}{2} & -\Omega_1 \\ \Omega_1 & \frac{\alpha_1}{2} \end{pmatrix}, \tag{4.47}$$

with

$$\Omega_1 = \sqrt{\frac{\sqrt{\Omega^4 + \alpha^2\Omega^2} + \Omega^2}{2}}, \tag{4.48}$$

$$\alpha_1 = \sqrt{2(\sqrt{\Omega^4 + \alpha^2\Omega^2} - \Omega^2)}, \tag{4.49}$$

we can write these solutions in the form

$$\begin{aligned} g_s(z) &= \lambda_s \mathbf{s}^{tr}(\frac{L}{2}) \cdot \mathbf{i} \cdot \mathbf{M}^{-1} \cdot \mathbf{a}(z), \\ g_a(z) &= \lambda_a \mathbf{a}^{tr}(\frac{L}{2}) \cdot \mathbf{i} \cdot \mathbf{M}^{-1} \cdot \mathbf{s}(z), \\ f_s(z) &= -\lambda_s \mathbf{s}^{tr}(\frac{L}{2}) \cdot \mathbf{M}^{-1} \cdot \mathbf{a}(z), \\ f_a(z) &= -\lambda_s \mathbf{a}^{tr}(\frac{L}{2}) \cdot \mathbf{M}^{-1} \cdot \mathbf{s}(z). \end{aligned} \tag{4.50}$$

Here tr means transpose, λ_s, λ_a, are arbitrary constants, and $\mathbf{i}$ is the imaginary unit matrix

$$\mathbf{i} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Note that $\alpha_1\Omega_1 = \alpha\Omega$ and, for small damping, $\Omega_1 \simeq \Omega$ and $\alpha_1 \simeq \alpha$. It is worth to remark that the boundary conditions for the g's in Eq. (4.45) are automatically satisfied independently of the λ's, while the ones for the f's are satisfied if

$$\lambda_s = -\frac{\Gamma_0}{\mathbf{s}^{tr}(\frac{L}{2}) \cdot \mathbf{s}(\frac{L}{2})},$$

$$\lambda_a = -\frac{\Gamma_0}{\mathbf{a}^{tr}(\frac{L}{2}) \cdot \mathbf{a}(\frac{L}{2})}.$$

Finally, the solution of the boundary value problem in Eqs. (4.23) and (4.21) for arbitrary values Γ_1, Γ_2 is constructed as

$$f(x) = f_a(z + \frac{L}{2}) - f_s(z + \frac{L}{2}), \tag{4.51}$$

$$g(x) = g_a(z + \frac{L}{2}) - g_s(z + \frac{L}{2}), \tag{4.52}$$

with λ_s, λ_a in Eq. (4.50) given by

$$\begin{aligned} \lambda_s &= -\frac{\frac{(\Gamma_1+\Gamma_2)}{2}}{\mathbf{s}^{tr}(\frac{L}{2}) \cdot \mathbf{s}(\frac{L}{2})}, \\ \lambda_a &= -\frac{\frac{(\Gamma_1-\Gamma_2)}{2}}{\mathbf{a}^{tr}(\frac{L}{2}) \cdot \mathbf{a}(\frac{L}{2})}. \end{aligned}$$

7 References

[1] T. Nagatsuma, K. Enpuku, F. Irie, and K. Yoshida, J. Appl. Phys. **54**, 3302 (1983); **56**, 3284 (1984).

[2] Y. M. Zhang and P. H. Wu, J. Appl. Physics **68**, 4703 (1990).

[3] A. V. Ustinov, J. Mygind, and V. A. Oboznov, J. Appl. Phys. **72**, 1203 (1992).

[4] A. V. Ustinov, J. Mygind, N.F. Pedersen, and V. A. Oboznov, Phys. Rev. **B46**, 578 (1992).

[5] V. P. Koshelets, A. V. Shchukin, S. V. Shitov, and L. V. Filippenko, IEEE Trans. Appl. Superconductivity **3**, 2524 (1993).

[6] O. H. Olsen, A. V. Ustinov, and N. F. Pedersen, Phys. Rev. **B48**, 13133 (1993).

[7] Y. M. Zhang, D. Winkler, and T. Claeson, Appl. Phys. Lett. **62**, 3195 (1993).

[8] M. Cirillo, F. Santucci, P. Carelli, M. G. Castellano, and R. Leoni, J. Appl. Phys. **73**, 8637 (1993).

[9] Y. M. Zhang and D. Winkler, IEEE Trans. Microwave Theory and Techniques **4**, 726 (1994).

[10] M. Cirillo, N. Grønbech-Jensen, M. R. Samuelsen, M. Salerno, and G. Verona Rinati, Phys. Rev. **B58**, 12377 (1998).

[11] A. Barone and G. Paternò, *Physics and Applications of the Josephson Effect* (J. Wiley, New York, 1982).

[12] D. W. McLaughlin and A. C. Scott, Phys. Rev. **A18**, 1672 (1978).

[13] R. D. Parmentier, in: *The New Superconducting Electronics*, (H. Weinstock and R. W. Ralston, eds.), Kluwer, NATO ASI Series, Vol. 251, Dordrecht, pp. 221-248 (1993).

[14] M. Salerno and M. R. Samuelsen, Phys. Rev. **B59**, 14653 (1999).

[15] N. Grønbech-Jensen and M. Cirillo, Phys. Rev. **B50**, 12851 (1994); N. Grønbech-Jensen, P. S. Lomdahl, and M. Cirillo, Phys. Rev. **B51**, 11690 (1995).

[16] M. Salerno and M. R. Samuelsen (unpublished).

[17] E. Jahnke and F. Emde, *Tables of Functions*, Dover Publications, NY, (1945), 145.

[18] M. Salerno, M. R. Samuelsen, G. Filatrella, S. Pagano, and R. D. Parmentier, Phys. Rev. **B41**, 6641 (1990).

5

Coupled Structures of Long Josephson Junctions

G. Carapella
G. Costabile

ABSTRACT
Long Josephson junctions are very attractive non-linear systems where solitonic dynamics [14] is fully developed and directly related to experimentally observable quantities. In this system the soliton occurs as a solution of the sine-Gordon equation describing the long junction; physically, it represents a fluxon, a current vortex enclosing a flux quantum. Coherent motion of fluxons is an intriguing subject, also because of its possible practical applications. In fact, microwave and far infrared fluxon oscillators [15] greatly enhance their performances if such a motion is established. Here we will consider three different coupled structures of long Josephson junctions where coherent fluxon motion is experimentally demonstrated.

1 Stacks of two long Josephson junctions

In this system, the coupling between junctions originates from the screening currents in the common electrode when its thickness is about equal or smaller than the London penetration depth λ_L. This "magnetic" coupling, which has been formalized in a model [16] for the multilayered structures, accounts for many dynamical phenomena in long stacked junctions, including synchronization of fluxon motion.

1.1 The physical system and its model

If we refer voltage and current polarities to the intermediate electrode in the stack with "double overlap" geometry shown in Fig. 1, the model equations are [16], [9]

$$\begin{aligned} \varphi_{xx} - \varphi_{tt} &= \sin(\varphi) + \alpha\varphi_t + \varepsilon\psi_{xx} - \gamma_A, \\ \psi_{xx} - \psi_{tt} &= \sin(\psi) + \alpha\psi_t + \varepsilon\varphi_{xx} + \gamma_B, \\ \varphi_x(0) &= \varphi_x(l) = \eta(1+\varepsilon), \\ \psi_x(0) &= \psi_x(l) = \eta(1+\varepsilon), \end{aligned} \tag{5.1}$$

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 103−119, 2000.

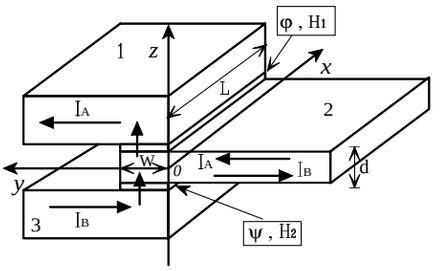

FIGURE 1. Stack with "double overlap" geometry.

where ε, $-1 < \varepsilon < 0$, is the magnetic coupling constant, defined as a function of the thickness d of the intermediate electrode of the stack and the thickness t of the insulating barriers by

$$\varepsilon = -\frac{\lambda_L}{\sinh(d/\lambda_L)} \frac{1}{[t + \lambda_L + \lambda_L \coth(d/\lambda_L)]}. \tag{5.2}$$

The lengths in Eqs. (5.1) are normalized to the Josephson penetration length $\lambda_J = \sqrt{\hbar c^2/8\pi e d'(1-\varepsilon^2)J_0}$, where $d' = t + \lambda_L + \lambda_L \coth(d/\lambda_L)$ and J_0 is the critical current density of the junctions. The time is normalized to the inverse of the plasma frequency $\omega_J = \overline{c}/\lambda_J$, where $\overline{c} = c\sqrt{t/\epsilon_r d'(1-\varepsilon^2)}$ is the Swihart velocity. Moreover, $l = L/\lambda_J$ is the normalized length of the junctions, $\alpha = (1/R)\sqrt{\hbar/(2eCJ_0)}$ is the ohmic dissipation (where R is the transverse resistance) and $\gamma_{A,B} = I_{A,B}/J_0 LW$ are the normalized bias currents. The η term in the boundary conditions accounts for an external magnetic field H (e.g., given by a coil) applied perpendicularly to the long dimension of the stack: $\eta = Hc/(4\pi\lambda_J J_0(1-\varepsilon^2))$.

In the absence of perturbations ($\gamma_A = \gamma_B = \alpha = \eta = 0$), the Hamiltonian of the coupled system is

$$H = \int_{-\infty}^{\infty} \left[\frac{1}{2}\varphi_x^2 + \frac{1}{2}\varphi_t^2 + 2 - \cos\varphi + \frac{1}{2}\psi_x^2 + \frac{1}{2}\psi_t^2 - \cos\psi - \varepsilon\varphi_x\psi_x\right] dx, \tag{5.3}$$

where, for the sake of simplicity, we assumed infinite length junctions. An exact solution of the unperturbed coupled system is [12]

$$\varphi = \sigma\psi = 4\arctan\left[-\exp\left[\gamma\left(\frac{u}{\sqrt{1-\sigma\varepsilon}}\right)\frac{x-ut}{\sqrt{1-\sigma\varepsilon}}\right]\right] \tag{5.4}$$

with $\gamma(y) \equiv 1/\sqrt{1-y^2}$. For $\sigma = 1$ this solution describes a fluxon-fluxon bound state travelling with velocity u upperly bound by the asymptotic velocity $u^+ = \sqrt{1-\varepsilon}$, while for $\sigma = -1$ it describes a fluxon-antifluxon bound state whose asymptotic velocity is $u^- = \sqrt{1+\varepsilon}$. By inserting solution Eq. (5.4) in Eq. (5.3) we get the energy of the two bound states

as

$$H_b = 16\sqrt{1-\sigma\varepsilon}\;\gamma\left(\frac{u}{\sqrt{1-\sigma\varepsilon}}\right). \tag{5.5}$$

From Eq. (5.5) it is seen that the antipolar state is energetically favorable with respect to the homopolar state, but for velocity larger than about u^-. The stability of the homopolar state for velocity between u^- and u^+ is discussed in detail in Ref. [12] and is explained on the basis of the occurrence of two characteristic velocities u^- and u^+ in the coupled system. However, possibly due to the difficulty of reaching "in flight" the range of velocities between u^- and u^+, we did not observe this state in our stacks.

Turning to the fluxon-antifluxon state, we can calculate the corresponding singular branch in the IV-characteristics (the so called first Zero Field Step, ZFS1) using a simple energetic approach [13]. Differentiating with respect to the time the energy Eq. (5.3) and using Eqs. (5.1) we get

$$\frac{dH}{dt} = \int_{-\infty}^{\infty}\left[\gamma_A\varphi_t - \gamma_B\psi_t - \alpha\varphi_t^2 - \alpha\psi_t^2\right]dx. \tag{5.6}$$

Following the classical approach [13] we assume that the dominant perturbation is in the velocity and we assume also the existence of a stationary velocity u that makes the energy Eq. (5.5) stationary. The relevant velocity of our state is then found inserting the fluxon-antifluxon solution in Eq. (5.6) with $dH/dt = 0$ (power balance) to have [2]

$$\frac{\gamma_A+\gamma_B}{2} = \frac{4\alpha}{\pi}\frac{u}{u^-}\frac{1}{\sqrt{1-\left(\frac{u}{u^-}\right)^2}}, \tag{5.7}$$

where $u^- = \sqrt{1+\varepsilon}$. By noticing that the measured voltage of the ZFS1 is proportional to the stationary velocity, $V_{ZFS1} = u\Phi_0/L$, and that γ_A and γ_B are proportional to the physical bias currents, $\gamma_{A,B} = I_{A,B}/J_0LW$, the Eq. (5.7) really describes the IV-curve of the ZFS1.

In the non-tunneling limit and in the absence of perturbations ($\sin\phi = \sin\psi = \gamma_A = \gamma_B = \alpha = \eta = 0$) Eqs. 5.1 reduce to equations for two coupled transmission lines

$$\begin{aligned}\varphi_{xx} - \varphi_{tt} - \varepsilon\psi_{xx} &= 0,\\ \psi_{xx} - \psi_{tt} - \varepsilon\varphi_{xx} &= 0.\end{aligned} \tag{5.8}$$

Looking for plane wave solutions $\varphi = Ae^{i(kx-\omega t)}$, $\psi = Be^{i(kx-\omega t)}$ we get two characteristic electromagnetic waves velocities $u^+ = \sqrt{1-\varepsilon}$ (in physical units: $c^+ = \bar{c}\sqrt{1-\varepsilon}$) for in-phase mode ($A = B$) and $u^- = \sqrt{1+\varepsilon}$ [$c^- = \bar{c}\sqrt{1+\varepsilon}$] for out-of-phase mode ($A = -B$). Since all of the current singularities exhibit asymptotic voltage proportional to the velocity of the electromagnetic waves in the junctions (Swihart velocity), all the current singularities accounting for the different regimes already known from single junctions [Zero Field Steps (ZFS's), Fiske Steps (FS's) and Flux Flow Steps (FFS's)] will exhibit two characteristic voltage spacings in the stack of two junctions.

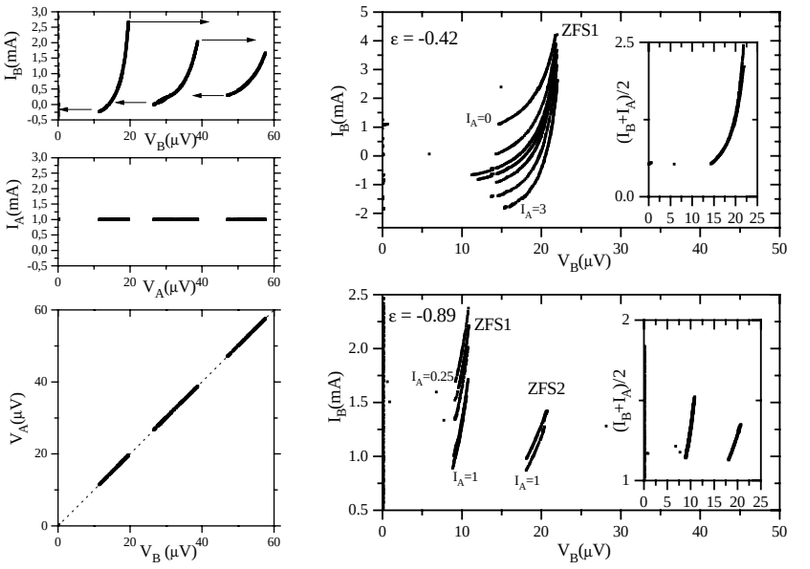

FIGURE 2. To the left, from top to the bottom: Three ZFS's recorded in junction B for junction A at $I_A = 1$mA; voltage of junction A while B is swept on these steps; voltage in junction A as a function of the voltage in junction B. The stack has $\varepsilon = -0.56$ and $l_A \approx l_B = 13$. To the right, top: IV-curves of the ZFS1 recorded in a stack with $\varepsilon = -0.42$, $l_A \approx l_B = 15$ for different bias currents in junction A. The curves are replotted in the inset using the mean current axis. To the right, bottom: Here the curves refer to the stack with $\varepsilon = -0.89$, $l_A \approx l_B = 10$ and also ZFS2 is considered.

1.2 Experiments on stacks of two long Josephson junctions

As was noted above, the antipolar (fluxon-antifluxon) state is energetically favorable in the two-stack junctions, so ZFS's with asymptotic velocity c^- are expected to appear in the IV-characteristics of the stack when we have no applied magnetic field. The ZFS1 accounts for an oscillatory motion of one fluxon-antifluxon pair, the ZFS2 for two oscillating pairs, and so on. The experimental evidence [4, 2] for three of such a ZFS's in a $Nb/AlO_x/Nb/AlO_x/Nb$ stack is shown in the upper left corner of Fig. 2. On these steps, the voltage of the junction A, biased with constant current, follows exactly the voltage of junction B (voltage polarities are referred to the intermediate electrode), indicating that the fluxon and antifluxons constituting the pairs are really bounded. In the upper right corner of Fig. 2 are shown the IV-curves of a ZFS1 recorded from junction B using the current in junction A as a parameter: as the current in junction A is increased, the curves shift toward the bottom; all the curves essentially fall on the top of each other if represented with a "mean current" axis defined

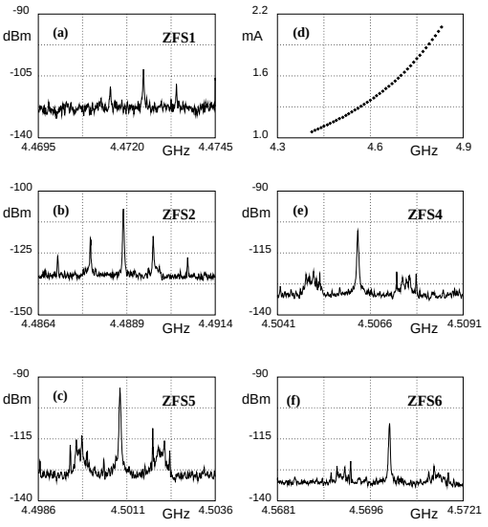

FIGURE 3. Power spectra of the radiation received from the ZFS's. The radiation is recorded at voltages: (a): 18.5 μV, (b): 37.1 μV, (c): 92.2 μV, (e): 72.0 μV, (f): 112.0 μV. In (d) is shown the $current - frequency$ curve of the upper branch of the ZFS1.

by $I_s = (I_A + I_B)/2$. For a stack with $\varepsilon = -0.89$ (lower left corner in Fig. 2) we plotted also ZFS2 and replotted it using the mean current axis in the inset. The result suggests that also higher order steps are driven by a force proportional to $I_s = (I_A + I_B)/2$.

Though Eq. (5.7) was obtained assuming a perfect fluxon-antifluxon state and infinite-length junctions, it describes the principal features of the experimental IV-curve summarized in Fig. 2. In fact Eq. (5.7) predicts u^- (c^- in physical units) as the asymptotic velocity of the step and the shift of the IV-curve if the bias current of one of the two junctions is used as a parameter. Moreover, from Eq. (5.7) it is evident that the force driving the state is the average of the bias currents, or in other words that the curves must fall on the top of each other if represented with the mean current axis (insets in Fig. 2).

In Fig. 3, we report the power spectra of the radiation received [2, 5] from the ZFS's in a stack with $\varepsilon = -0.56$. Apart from the spurious spectral components (due to an instrumental effect), the fundamental frequency of the emitted signal is found to satisfy the same frequency-voltage relation as the ZFS's in single overlap junctions

$$f_{ZFSN} = \frac{1}{2}\Phi_0^{-1}\frac{V_{ZFSN}}{N} = \frac{1}{2}483.6\frac{MHz}{\mu V}V_{ZFS1}, \tag{5.9}$$

where N is the order of the step. We remark that when we are on the ZFS's, the a.c. signals generated by the two junctions are opposite in sign. So, we could conclude that the net signal generated at the edge of the stack should be the sum of nearly opposite signals and consequently too small to be detected. Nevertheless, we received appreciable radiation from the

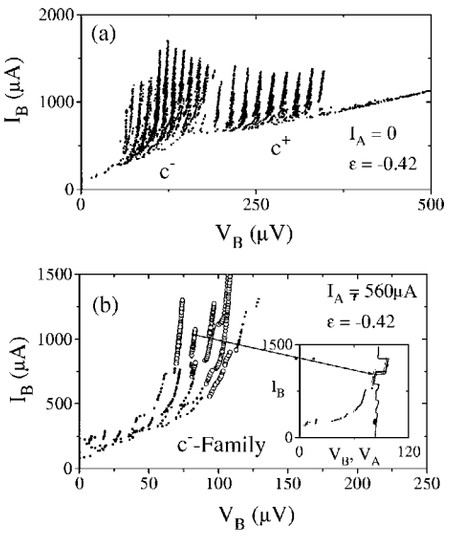

FIGURE 4. (a) Resonant flux-flow steps in junction B while junction A is unbiased. (b) Voltage locked states on the resonant flux-flow steps of the c^- family recorded in junction B while junction A is biased with a positive constant current. The marked branches (circles) indicate where the junctions are voltage-locked. In the inset is shown a particular step together with the voltage of junction A (steeper line) versus current in junction B.

ZFS's. The explanation for the experimental fact documented in Fig. 3 has been ascribed to the impulsive nature of the signals, joined to a suitable time delay between the component signals generated by the non identical length of the junctions in the stack [2].

Figure. 4(a) shows the resonant flux-flow steps (cavity modes, or Fiske steps) recorded in one junction of the stack while the other is unbiased. The plots are obtained superimposing the IV- characteristics recorded for different values of the applied magnetic field. They exhibit two families of Fiske steps characterized by a different voltage spacing: this correspond to the splitting [18, 17] of the Swihart velocity $\bar{c}$ of the single junction into the velocities c^- and c^+ in the stack. Biasing in series the stack voltage locking with opposite (relatively to the intermediate electrode) polarities [18] is observed in the c^- and in the c^+ family of steps. However, biasing one junction of the stack with a constant current and sweeping the other in the resonant flux flow regime, it is possible to record up to four voltage locked states [3], or in other words it is possible to excite four different Fiske modes, with equal or opposite voltage polarities, both in the c^- and in the c^+ family. An experimental example of voltage locking with equal voltage polarities in the c^- family is reported in Fig. 4(b). As noted above, in-phase oscillations are associated to the steps of the c^+ family and out-of-phase oscillations to the steps of the c^- family. All of these voltage locked modes have been theoretically studied in Ref. [11].

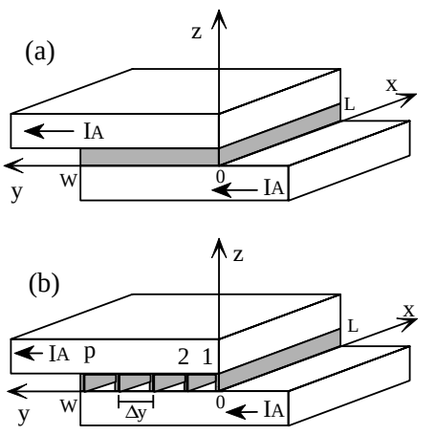

FIGURE 5. (a) Sketch of a two-dimensional Josephson junction. (b) Parallel array of p long Josephson junctions. The "holes" between the junctions model the regions without tunneling separating the junctions in the experimental device.

2 Parallel arrays of Josephson junctions

Coherence should be expected if the interaction between fluxons is attractive. Due to their physical nature, homopolar fluxons localized in the same plane will statically repel each other; on the contrary, they will *attract* each other if they are localized in parallel planes. Following this intuitive physical consideration, we here consider a structure, consisting of a parallel array of long Josephson junctions, where the condition for the observation of coherence is satisfied.

2.1 The physical system and its model

To model the parallel array of Fig. 5(b), we can start from the description of the two-dimensional Josephson junction in Fig. 5(a), that is modelled as [10]

$$\begin{aligned} \varphi_{xx} + \varphi_{yy} - \varphi_{tt} &= \sin\varphi + \alpha\varphi_t \\ \varphi_x(0) = \varphi_x(l) &= \eta^e \equiv H^e/H_c \\ \varphi_y(0) &= -\eta^T - \chi \equiv -H^T/H_c - I_A/2\lambda_J J_0 L \\ \varphi_y(w) &= -\eta^T + \chi \equiv -H^T/H_c + I_A/2\lambda_J J_0 L \end{aligned} \tag{5.10}$$

where φ is the phase difference between the two superconducting electrodes and where we have accounted for an externally applied magnetic field $\mathbf{H_a} = (H^T, H^e, 0)$ as well as for the field generated by the bias current I_A. The parallel array of long junctions (that are assumed one-dimensional for the sake of simplicity) in Fig. 5(b) becomes the continuous system of Fig. 5(a) in the limit

$$p \to \infty,\ \Delta y \to 0;\ p\Delta y = W. \tag{5.11}$$

For a finite separation between the junctions Δy and for a finite number of junctions p, we can think of the system in Fig. 5(b) as the $y-$discretized version of the continuous system in Fig.5(a). Thus, the model for the parallel array of long Josephson junctions will be obtained by discretizing the model Eqs.(5.10) in the $y-$direction

$$\varphi(x,y,t) \to \varphi(x,[n-1]\Delta y,t) \equiv \varphi_n(x,t).$$

Hence, our model is [6]

$$\begin{gathered}
\varphi_{ntt} = \varphi_{nxx} - \sin\varphi_n - \alpha\varphi_{nt} + \tfrac{1}{\beta}(\varphi_{n+1} - 2\varphi_n + \varphi_{n-1}) \qquad 2 \le n \le p-1 \\
\varphi_{1tt} = \varphi_{1xx} - \sin\varphi_1 - \alpha\varphi_{1t} + \tfrac{2}{\beta}(\varphi_2 - \varphi_1) + p\gamma_b + 2\eta^T/\sqrt{\beta} \\
\varphi_{ptt} = \varphi_{pxx} - \sin\varphi_p - \alpha\varphi_{pt} + \tfrac{2}{\beta}(\varphi_{p-1} - \varphi_p) + p\gamma_b - 2\eta^T/\sqrt{\beta} \\
\varphi_{nx}(0) = \varphi_{nx}(l) = \eta^e,
\end{gathered} \tag{5.12}$$

where we have defined $\beta \equiv (\Delta y/\lambda_J)^2$ and γ_b is the bias current per junction, normalized to the critical current of the single junction.

In the absence of perturbations ($\alpha = \eta^e = \eta^T = \gamma_b = 0$) Eqs.(5.12) simplify to a system of "elastically" coupled sine-Gordon equations

$$\varphi_{ntt} = \varphi_{nxx} - \sin\varphi_n + \frac{1}{\beta}(\varphi_{n+1} - 2\varphi_n + \varphi_{n-1}) \tag{5.13}$$

with the Hamiltonian

$$H = \sum_m \int_0^l \left[\frac{1}{2}(\varphi_{mx})^2 + \frac{1}{2}(\varphi_{mt})^2 + p - \cos\varphi_m + \frac{1}{2\beta}(\varphi_m - \varphi_{m-1})^2\right] dx. \tag{5.14}$$

Some analytical calculations can easily be carried out if we assume infinite length junctions. In this case, two homopolar (fluxon-fluxon) or two antipolar (fluxon-antifluxon) solitons separated by ξ along the x-direction, moving with translational velocity u and localized in two contiguous junctions are described as

$$\begin{aligned}
\varphi_m &= 4\arctan\left[\exp(\gamma\left(x - \tfrac{\xi}{2} - ut\right)\right] \\
\varphi_{m-1} &= 4\arctan\left[\exp(\sigma\gamma\left(x + \tfrac{\xi}{2} - ut\right)\right],
\end{aligned} \tag{5.15}$$

where $\sigma = \pm 1$ and $\gamma = 1/\sqrt{1-u^2}$. The interaction energy between the two solitons is found inserting these solutions in the nearest-neighbours interaction energy

$$E_I(\xi,u) \equiv E_{m,m-1} = \frac{1}{2\beta}\int_{-\infty}^{\infty} dx\,(\varphi_m - \varphi_{m-1})^2. \tag{5.16}$$

Numerical evaluation of the integral in Eq. (5.16) shows that $E_I(\xi)$ has a quasi-parabolic behavior in the static regime ($\gamma = 1$) and such a behavior

is conserved in the relativistic regime for $\xi \ll 1$. However, the interaction energy is found attractive (repulsive) for two homopolar (antipolar) fluxons. Similar results are found for long but finite junctions. The quasi-parabolic behavior observed for $\gamma\xi < 1$ can be recovered if we expand to order ξ^2 the integrand in Eq. (5.16). An approximate analytic form for the interaction energy and the interaction force is then found as

$$\begin{aligned} E_I(\xi,\gamma) &\approx \sigma \tfrac{8\gamma}{2\beta}\xi^2 \\ F_I(\xi,\gamma) &\equiv -\tfrac{\partial E_I}{\partial \xi} = -\sigma \tfrac{8\gamma}{\beta}\xi, \end{aligned} \tag{5.17}$$

where we recall that $\beta = \Delta y^2$. The situation is opposite with respect to the one in which the solitons are in the same plane. Here the solitons are in parallel planes and, as we should intuitively expect, we find attractive interaction for homopolar ($\sigma = 1$) solitons and repulsive interaction for antipolar ($\sigma = -1$) solitons. Obviously, this kind of interaction helps the coherent motion of homopolar solitons arranged in more or less ordered rows ("strings").

Neglecting the tunneling ($\alpha = \sin\varphi_n = 0$), our array becomes a system of coupled transmission lines described by

$$\varphi_{nxx} + \frac{1}{\beta}(\varphi_{n+1} - 2\varphi_n + \varphi_{n-1}) - \varphi_{ntt} = 0. \tag{5.18}$$

The dispersion relation of the structure is obtained substituting a plane wave solution

$$\varphi_n(x) = Ae^{i(k_x x + (n-1)\sqrt{\beta}k_y - \omega t)}$$

in (5.18) and cavity modes resonances are then found imposing open circuit boundary conditions, with the result

$$\omega_{j,m} = \sqrt{\left(\frac{j\pi}{l}\right)^2 + \frac{4}{\beta}\sin^2\frac{m\pi}{2(p-1)}} \tag{5.19}$$

where $m,\ j$ are integers. These resonances are practically excited by a magnetic field and are manifested as current singularities in the IV-characteristics. The voltages of occurrence are found [1] matching the Josephson oscillation frequency with the cavity frequency (5.19). So, if we apply only a magnetic field along the y-direction, the cavity modes will have voltages $V_j = j\Phi_0\overline{c}/2L$, i.e., a series of singularities spaced half the voltage spacing of the ZFS's. If we apply a field only in the x-direction, the cavity modes will have the voltages

$$V_m = \overline{c}\frac{\Phi_0}{\Delta y}\left|\sin\frac{m\pi}{2(p-1)}\right| \tag{5.20}$$

i.e., a series of singularities not evenly spaced, and limited in voltage.

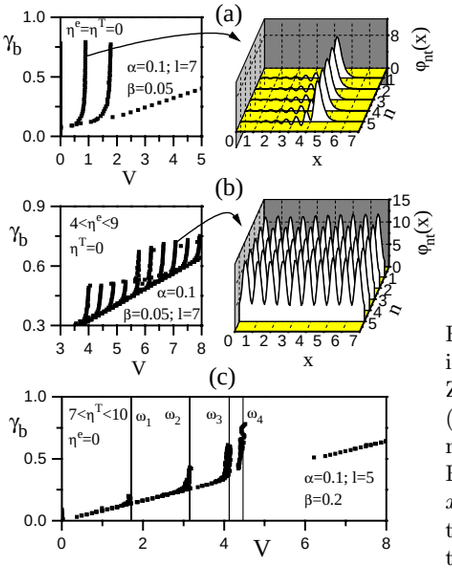

FIGURE 6. (a) Numerical IV-characteristics of ZFS1 and ZFS2. (b) Cavity modes resonances (Fiske steps) excited by a variable magnetic field in the y-direction. (c) Resonances excited by a field in the x-direction. The vertical bars are the voltage positions predicted by the discrete dispersion relation.

2.2 Numerical and experimental results on five-junctions parallel arrays

Due to the attractive interaction, we expect that coherent oscillatory motion of fluxons in zero magnetic field can take place in our parallel array, i.e., we can expect to observe a ZFS1 due to a string of p fluxons (one per junction) in oscillatory motion, a ZFS2 due to a string of $2p$ fluxons (two per junction) in oscillatory motion etc. ... These coherent ZFS's should appear in the IV- characteristics at the same voltages as in the single junction, and with a current amplitude p times the current of the single junction. Due to the coherent motion, the emitted radiation should have the same fundamental frequency Eq. (5.9) of the single junction. Numerical simulations [6] of model Eqs. (5.12) confirm this picture, as shown in Fig. 6(a).

In Fig. 7(a) there are shown the first three ZFS's recorded [6] in an array of five $(600 \times 20)\,\mu\text{m}^2$ junctions, separated by 20μm. The coupling parameter for this array was $\beta = 0.05$. The observed voltage spacing is $39\mu V$, the amplitude of the current is five times the expected amplitude of the ZFS of the single junction. With this voltage spacing we expect, from the Eq. (5.9), electromagnetic radiation in the X-band when the array is polarized on these ZFS's. The signals detected when the array is polarized on the ZFS3 and on the ZFS1 are shown in Fig. 7(a). The signal levels are relatively high, though we have not used any impedance matching circuit, and the linewidth is quite narrow. So, also taking into account the numerical results of Fig. 6(a), we take the results in Fig. 7(a) as an experimental evidence of coherent oscillatory motion of homopolar fluxons in a parallel

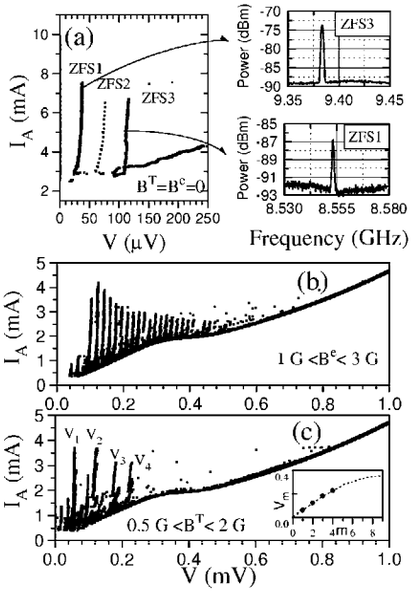

FIGURE 7. (a) ZFS's recorded in a five-junctions array and microwave emission recorded when the array is biased on the ZFS3 at $V = 116.4\mu V$ or on the ZFS1 at $V = 35.3\mu V$. (b) Cavity modes resonances excited by a magnetic field applied along the y-direction of the array. (c) Voltage-limited resonances series excited by a magnetic field applied along the x-direction of the array. In the inset the asymptotic voltages of the resonances are fitted with the theoretical functional form.

array of Josephson junctions.

The Fiske steps excited by a field in the y-direction are also found possible in a coherent state, as it is shown in the numerical results in Fig. 6(b). Fig. 7(b) shows such singularities recorded from our array. The voltage spacing is $19.5\mu V$, *i.e.*, one half the voltage spacing of ZFS's in Fig. 7(a). The picture is very similar to that of a single junction with current amplitude of the steps five times the amplitude of the single junction. So, taking also into account the numerical results, we can interpret the results in Fig. 7(a) as the experimental evidence for in-phase cavity modes excited by fluxon chains in a coherent unidirectional motion.

The experimental series of resonances excited by a field in the x-direction of the array is shown in Fig. 7(c). The series is qualitatively in agreement with numerical results shown in Fig. 6(c), with results of similar experiments in unidimensional arrays of small Josephson junctions, and with relation (5.20).

3 Triangular cells of long Josephson junctions

The most remarkable feature of a triangular cell of long Josephson junctions is that [19, 20], operating junctions ϕ_1 and ϕ_2 [see Fig. 8(a)] in the flux flow mode, a.c. power can be transferred to junction ψ_1 by mean of an axial flux Φ in the loop. When Φ is close to $\Phi_0/2$, junction ψ_1 is expected

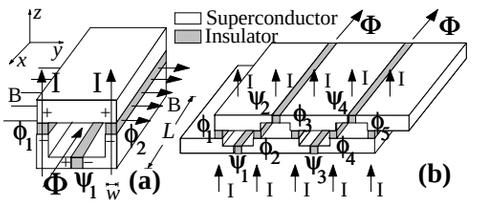

FIGURE 8. (a) Triangular cell of long Josephson junctions. (b) Array of triangular cells.

[20] to generate, at zero d.c. voltage, a nearly sinusoidal a.c. signal with doubled amplitude with respect to the junctions ϕ_1 and ϕ_2. Similar features are exhibited by the "transverse" junctions (ψ_1, ψ_3, ... or ψ_2, ψ_4....) in the array shown in Fig. 8(b). Moreover, these transverse junctions also add their signals and impedances, making the array capable of large output voltage and impedance with negligible power emitted into higher harmonics.

3.1 The model

If we neglect the self-inductance of the loop and the mutual inductance between the junctions, the coupling between the junctions in the triangular cell shown in Fig. 8(a) originates from the quantization of the axial magnetic flux Φ linked to the cell [1]

$$\psi_1(x,t) = \phi_1(x,t) - \phi_2(x,t) + 2\pi\Phi/\Phi_0 - 2\pi n, \tag{5.21}$$

where $\Phi_0 = h/2e$ is the flux quantum. Assuming the three junctions to be identical and taking into account Eq. (5.21), we can easily derive the coupled equations that model the system as

$$\begin{gathered}
\phi_{1,xx} - \phi_{1,tt} = \alpha\phi_{1,t} - \beta\phi_{1,xxt} - \gamma_b \\
+\tfrac{2}{3}\sin\phi_1 + \tfrac{1}{3}\sin\phi_2 + \tfrac{1}{3}\sin\left(\phi_1 - \phi_2 + 2\pi\Phi/\Phi_0\right), \\
\phi_{2,xx} - \phi_{2,tt} = \alpha\phi_{2,t} - \beta\phi_{2,xxt} - \gamma_b \\
+\tfrac{2}{3}\sin\phi_2 + \tfrac{1}{3}\sin\phi_1 + \tfrac{1}{3}\sin\left(\phi_2 - \phi_1 - 2\pi\Phi/\Phi_0\right), \\
\phi_{1,x}(0) + \beta\phi_{1,xt}(0) = \eta, \\
\phi_{2,x}(l) + \beta\phi_{2,xt}(l) = \eta,
\end{gathered} \tag{5.22}$$

where γ_b is the normalized bias current, α accounts for the ohmic dissipation in the barrier and β for the surface impedance of the electrodes, l is the normalized length of the junctions, and η is the normalized magnetic field applied in the y direction [Fig. 8(a)].

Some analytical traveling wave solutions $\varphi_k(\xi) = \varphi_k[\gamma(x - ut)]$ of the unperturbed ($\alpha = \beta = \gamma_b = \eta = 0$) system have been reported [19] for two values of the axial flux, i.e., $\Phi = 0$ and $\Phi = \Phi_0/2$. The stable (lower energy) solutions $\vec{K} = [\phi_1(\xi), \phi_2(\xi)]$, generally named Kinks in the following, are summarized in Table 5.1. In the table we reported also the fluxon content of the waves constituting the Kink, defined as $\Phi_k = 2\pi^{-1}\int_{-\infty}^{+\infty}\phi_{k,x}dx$. As

TABLE 5.1. Basic stable Kinks.

$\Phi = 0 \rightarrow \phi(\xi) = 4\tan^{-1}[\exp(\xi)]$				
Kink	ϕ_1	Φ_1	ϕ_2	Φ_2
(**EF**)	$\phi(\xi)$	$+1$	0	0
(**EB**)	0	0	$\phi(\xi)$	$+1$
(**EC**)	$\phi(\xi)$	$+1$	$\phi(\xi)$	$+1$
$\Phi = \Phi_0/2 \rightarrow \phi(\xi) = 2\tan^{-1}\left\{3^{-1/2}\tanh\left[8^{-1/2}(\xi)\right]\right\}$				
Kink	ϕ_1	Φ_1	ϕ_2	Φ_2
(**ab**)	$2\phi(\xi) + 3\pi$	$+2/3$	$\phi(\xi) + 2\pi$	$+1/3$
(**bc**)	$\phi(\xi) + 4\pi$	$+1/3$	$2\phi(\xi) + 3\pi$	$+2/3$
(**cd**)	$-\phi(\xi) + 4\pi$	$-1/3$	$\phi(\xi) + 4\pi$	$+1/3$

is seen in Table 5.1 the basic Kinks (**EF**), (**EB**), (**EC**) and their anti-Kinks (**FE**), (**BE**), (**CE**), obtained substituting ξ with $-\xi$ in the Kinks definitions, consist of components with an integer fluxon content. Conversely, at $\Phi = \Phi_0/2$ the basic Kinks have a fractional fluxon content.

3.2 Numerical and experimental results

In the presence of perturbations only the (**EC**) Kink (see Table 5.1) is found stable in simulations [7] at $\Phi = 0$. The stationary regime consists of a (**EC**) Kink travelling toward one edge of the cell where it is reflected as an anti-Kink (**CE**) and again as (**EC**) at the other edge, so establishing an oscillatory motion. In other words, at $\Phi = 0$ the dynamics of the ZFS1 in the triangular cell is equivalent to the dynamics of the ZFS1 in single overlap junction, with the role of the single fluxon played now by a pair of bounded fluxons (one in the junction ϕ_1, the other in the junction ϕ_2). The analogy with single junctions is recovered also for the higher order ZFS's, with a bunch (**EC**)-(**EC**) (two fluxons in each junction) accounting for the ZFS2, a bunch (**EC**)-(**EC**)-(**EC**) for the ZFS3, and so on. The singularities in the IV-characteristics of the cell corresponding to these dynamical states are similar to the ones of the single junction: a series of steps spaced by $\Delta V = 2\pi/l$ (in physical units $\Delta V = \Phi_0\bar{c}/L$), as it is shown in Fig. 9.

In the presence of an axial flux $\Phi = \Phi_0/2$ the initial (**EC**) Kink (a fluxon in each junction) evolves into an excitation consisting of a (**bc**)-(**ab**) bunch. Hence, there is again a total flux quantum traveling in both biased junctions. The excitation is accelerated toward one edge of the cell where it is reflected as (**ba**)-(**cb**), and so forth. In other words, we observe an oscillation similar to the one accounting for the ZFS1 at $\Phi = 0$, but now the flux quantum is carried by the (**bc**)-(**ab**) excitation. As is seen in Fig. 9 the current step in the IV-characteristics corresponding to this dynamical state at $\Phi = \Phi_0/2$ has the same asymptotic voltage but a critical current much smaller than the ZFS1 at $\Phi = 0$.

Higher order steps exhibit a more complex dynamics. The ZFS2 at $\Phi =$

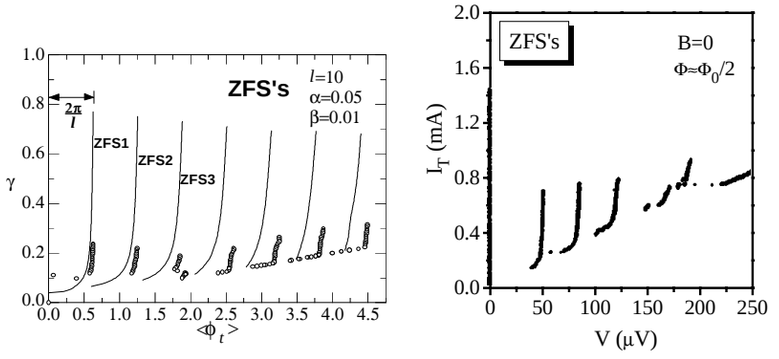

FIGURE 9. To the left: Numerically obtained ZFS's at $\Phi = 0$ (continuous line) compared with the ZFS's at $\Phi = \Phi_0/2$ (open circles). To the right: ZFS's recorded in a device with $l = 7$ at $\Phi \approx \Phi_0/2$.

$\Phi_0/2$ consist of a (**bc**)-(**ab**) and a (**ba**)-(**cb**) counter-oscillating in the junctions, in a fashion similar to the *symmetric* mode [15] of the single overlap junctions. However, both in the single junction and in the cell at $\Phi = 0$, the excitations carrying more flux quanta tend to oscillate bunched in the junction (*bunch* mode [15]). Here, instead, the higher order steps are almost exclusively related to *symmetric* modes. In some cases, the cell exhibits an hybrid behavior, as the one encountered in the ZFS3. Here, a transition is observed from a symmetric mode to a bunch mode when the bias exceeds a critical value. The current step in the IV-characteristics is broken at the critical value of the bias corresponding to the transition, as shown in Fig. 9.

To the right of Fig. 9 are shown steps recorded at $\Phi \approx \Phi_0/2$ and $B = 0$ in a device that we fabricated. The junctions in the cell have normalized length $l = 7$. Apart from the fourth step, that accounts for an external resonance, the steps at $\Phi = 0$ were found similar to the ones shown here, but with a higher critical current. This is in qualitative agreement with the numerical results shown to the left of Fig. 9. Moreover, the voltage spacing of the first three steps was found to be two times the voltage spacing of the Fiske steps observed in magnetic field. So, we can take the steps in Fig. 3.2 as the experimental evidence [7] for the discussed ZFS's in the triangular cell.

In Fig. 3.2 are shown the numerically obtained IV-characteristics of the cell for values of the magnetic field that induce resonant flux flow steps. As it is seen in this figure, both for $\Phi = 0$ and for $\Phi = \Phi_0/2$ the IV-curves consist of a step broken in sub-steps separated by the characteristics voltage spacing $\Delta_{FS} = \pi/l$ (in physical units: $\Delta_{FS} = \Phi_0\bar{c}/2L$) of the cavity modes (Fiske steps). The voltage signals generated at one edge of the junctions are, however, different in the two cases. At $\Phi = 0$, the signal generated by the ψ_1 junction is found to be zero. At $\Phi = \Phi_0/2$ the signal generated

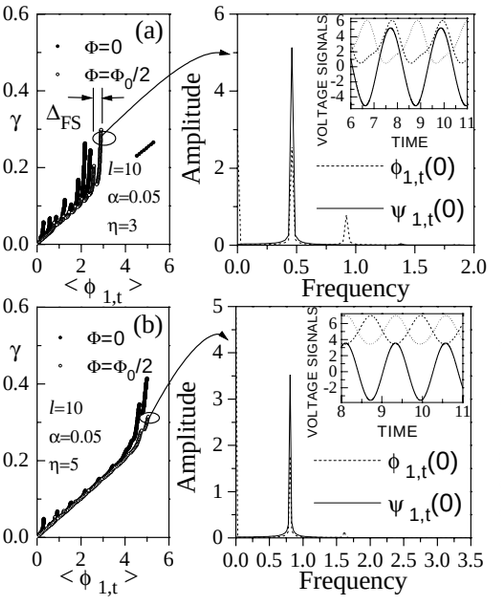

FIGURE 10. (a) Numerically obtained IV-curves of the resonant flux flow steps at $\Phi = 0$ and $\Phi = \Phi_0/2$ and analysis of the voltage signals (at $\Phi = \Phi_0/2$) generated at one edge of the cell by the junctions when biased at the marked point. (b) Same as in (a), but here $\eta = 5$.

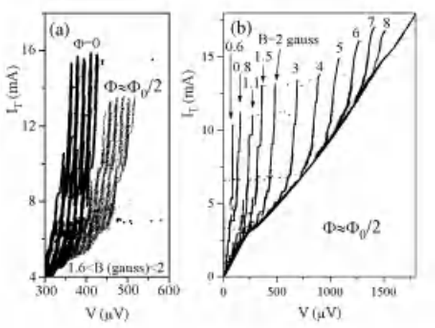

FIGURE 11. (a) Resonant flux flow steps recorded in a device with $l = 23$ at $\Phi = 0$ and $\Phi \approx \Phi_0/2$. (b) Steps recorded in the device at $\Phi \approx \Phi_0/2$ for different magnetic fields.

by the ψ_1 junction is, instead, other than zero and exhibits an enhanced spectral purity, as it is better seen in the signal analysis at $\Phi = \Phi_0/2$ shown in Fig. 3.2.

In the simulations, we observed this property also at the values of the magnetic field that induce the pure flux flow regime. We also noted that the flux flow branches in the IV-characteristics at $\Phi = \Phi_0/2$ start at slightly higher voltage (but the signal frequency was the same as in the $\Phi = 0$ case). Moreover, for quite large field values, the critical current of the flux flow state was found lower in the $\Phi_0/2$ case than in the $\Phi = 0$ case, as it is also seen in Fig. 3.2(b).

Experimental [8] resonant flux flow steps recorded in a device with $l = 23$ at $\Phi = 0$ and $\Phi \approx \Phi_0/2$ are shown in Fig. 3.2(a). Here on the current axis is reported the total current I_T flowing in the device, that is parallel biased as shown in Fig. 8. As it is seen, the characteristics voltage spacing of the Fiske steps, here $\Delta_{FS} = 15\mu$V, occurs for both the values of the axial flux, with the resonant flux flow steps at $\Phi \approx \Phi_0/2$ slightly lower in current. Moreover, the family of resonant flux flow steps at $\Phi \approx \Phi_0/2$ is also found slightly shifted toward higher voltages, in agreement with the qualitative indications of the numerical simulations [see Fig. 3.2(b)]. In Fig. 3.2(b) we show some steps recorded at $\Phi \approx \Phi_0/2$ for different values of the magnetic field. For magnetic field values larger than $B = 4$ gauss, also the pure flux flow regime is exhibited by the device, up to a voltage value of 1500 μV. From the Josephson frequency-voltage relation $f = \Phi_0^{-1}V$, this says that the device might be able to generate a signal of appreciable power and good spectral purity up to about 700 GHz when operated in the flux flow regime.

4 References

[1] A. Barone and G. Paternó. *Physics and Applications of the Josephson Effect.* John Wiley & Sons, New York (1982).

[2] G. Carapella. Fluxon-antifluxon state in stacked Josephson junctions. *Phys. Rev. B*, **59**:1407 (1999).

[3] G. Carapella and G. Costabile. Mutual phase-locking on resonant flux flow steps in stacked long Josephson junctions. *Appl. Phys. Lett.*, **71**:3409 (1997).

[4] G. Carapella, G. Costabile, A. Petraglia, and J. Mygind. Phase locked fluxon-antifluxon states in stacked Josephson junctions. *Appl. Phys. Lett.*, **69**:1300 (1996).

[5] G. Carapella, G. Costabile, N. F. Pedersen, and J. Mygind. Microwave radiation from zero field singularities in stacks of two long Josephson junctions. *to be published in IEEE Trans. Appl. Sup.* (1999).

[6] G. Carapella, G. Costabile, and P. Sabatino. Coherent motion of homopolar solitons in parallel arrays of long Josephson junctions. *Phys. Rev. B*, **58**:15094 (1998).

[7] G. Carapella, G. Costabile, and P. Sabatino. Zero field and flux flow excitations in a triangular cell of long Josephson junctions. *submitted to Phys. Rev. B* (1998).

[8] G. Carapella, G. Costabile, and P. Sabatino. Flux flow steps in a triangular cell of long Josephson junctions. *Appl. Phys. Lett.*, **74**:90 (1999).

[9] G. Carapella, G. Costabile, S. Sakai, and N. F. Pedersen. Maximum supercurrent in two Josephson junctions stacks: Theory and experiment. *Phys. Rev. B*, **58**:6497 (1998).

[10] J. C. Eilbeck, P. S. Lomdahl, and O. H. Olsen. Comparison between one-dimensional and two-dimensional models for Josephson junctions of overlap type. *J. Appl. Phys.*, **57**:861 (1985).

[11] N. Grønbech-Jensen, J. Blackburn, and M. R. Samuelsen. Phase locking between Fiske and flux flow modes in coupled sine-Gordon systems. *Phys. Rev. B*, **53**:12364 (1996).

[12] N. Grønbech-Jensen, D. Cai, A. R. Bishop, A. W. C. Lau, and P. S. Lomdahl. Bunched fluxons in coupled Josephson junctions. *Phys. Rev. B*, **50**:6352 (1994).

[13] D. W. McLaughlin and A. Scott. Perturbation analysis of fluxon dynamics. *Phys. Rev. A*, **18**:1652 (1978).

[14] R. D. Parmentier. In K. Lonngren and A. C. Scott, editors, *Solitons in Action*. Academic, New York (1978).

[15] S. Pnevmatikos and N. F. Pedersen. In P. L. Christiansen, J. C. Eilbeck, and R. D. Parmentier, editors, *Future Directions of Nonlinear Dynamics in Physical and Biological Systems*, volume 312 of NATO Advanced Study Institute of *B: Physics*. Plenum, New York (1993).

[16] S. Sakai, P. Bodin, and N. F. Pedersen. Fluxons in thin film superconductor-insulator superlattices. *J. Appl. Phys.*, **73**:2411 (1993).

[17] S. Sakai, A. V. Ustinov, H. Kohlstedt, A. Petraglia, and N. F. Pedersen. Theory and experiment on electromagnetic wave propagation velocities in stacked superconducting tunnel structures. *Phys. Rev. B*, **50**:12905 (1994).

[18] A. V. Ustinov, H. Kohlstedt, and C. Heiden. Possible phase locking of vertically stacked Josephson flux flow oscillators. *Appl. Phys. Lett.*, **65**:1457 (1994).

[19] S. P. Yukon and N. C. H. Lin. Coupled long Josephson junctions and the n sine-Gordon equation. *IEEE Trans. Magn.*, **27**:2736 (1991).

[20] S. P. Yukon and N. C. H. Lin. ˙ In R. D. Parmentier and N. F. Pedersen, editors, *Nonlinear Superconducting Devices and High-Tc Materials*. World Scientific, Singapore (1995).

6
Stacked Josephson Junctions

N.F. Pedersen

1 Introduction

The topic of fluxons in Josephson junctions appears often and in many contexts within superconductivity. The potential applications and the appearance of high-T_c superconductivity are some of the reasons for this recently increased interest. Several comprehensive reviews of fluxons (solitons) in long Josephson junctions [8], [2], [10], [9] exist, of which particularly the most recent [9] gives an excellent overview of the state of the art for low-T_c niobium-type long Josephson junctions. In recent years the properties of long Josephson junctions stacked on top of each other has been studied quite intensively. This is because such systems model existing low-T_c superconducting multilayers. In addition these models have revealed surprising dynamical properties. The appearance of high-T_c ceramic superconductors has also had quite an impact on the topic of fluxons in superconductors, in particular stacked long Josephson junctions. Here the fluxons play a very important role, for example, in the BSCCO type of material with a layered structure and big anisotropy. Dynamic properties of fluxons are also relevant for the material properties such as the bulk current density and the behaviour near the transition temperature. Some of these high-T_c phenomena we will try to include in our description below.

2 Short summary of fluxon properties

An excellent mathematical description of fluxons in long Josephson junctions may be found, for example, in Ref. [9]. We will thus only summarize the most important properties here, and rely on Ref. [8], [2], [10], [9] for details. Fluxon propagation in a long quasi-one-dimensional Josephson junction is assumed to be described by the perturbed sine-Gordon equation [8], [2], [10], [9] for the quantum mechanical phase difference, $\phi(x,t)$,

$$-\phi_{xx} + \phi_{tt} + \alpha\phi_t + \sin\phi = \gamma \; . \tag{6.1}$$

Here, subscripts x and t denote differentiation with respect to normalized space and time, respectively. The sine-Gordon equation with its boundary

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 121−136, 2000.

conditions may appear in different ways depending on the geometry of the junction [8], [2], [10], [9]. The most often used geometry is the so-called overlap geometry described by Eq. (6.1) together with boundary conditions $\phi_x = 0$ at both ends $x = 0$ and $x = L$ of the junction. Another frequently used geometry is the annular geometry obtained by joining the two ends of the junction ($x = 0$ and $x = L$) (see Ref. [10]). For this geometry the boundary conditions are $\phi(x = 0) = \phi(x = L) + 2p\pi$. The dimensionless parameter α describes the damping, with typical values in the range $0.01 < \alpha < 0.1$ for low-T_c systems. The parameter γ describes the normalized bias current which is a controllable parameter in experiments, with typical values in the range $0 < \gamma < 1$. The time t is measured in units of $1/\omega_0$, where ω_0 is the plasma frequency, and length x is measured in units of the Josephson penetration depth λ_j (see Ref. [8], [2], [10], [9]). Typical values of ω_0 and λ_j for low- and high-T_c superconductors are 10^{11} and 10^{12} rad/s and 10^{-5} and 10^{-6} m, respectively. Particularly for high-T_c systems these parameters are not well known, as there is not yet a well established theory for the superconductivity itself.

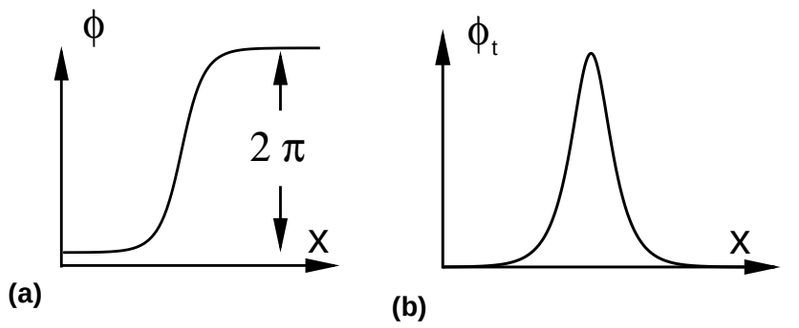

FIGURE 1. Appearance of a fluxon in a long Josephson junction as a 2π kink in the phase difference ϕ(x). (1b) as a voltage pulse $\phi_t(x)$.

The simplest solution to Eq. (6.1) is the fluxon or soliton which is a localized 2π kink in the phase difference ϕ moving with a velocity u. This velocity is determined as a balance between the losses α and the bias γ. The velocity u is measured in units of c, where c is the velocity of light in the junction, often called the Swihart velocity. With a perturbational approach u is approximately given by the expression [8], [2], [10], [9]

$$u = 1/\sqrt{(1 + (4\alpha/\pi\gamma)^2)}. \tag{6.2}$$

and thus may assume values $0 < u < 1$. Note that for low values of the bias

current the velocity is proportional to the bias $u \sim \gamma$, while for large values of γ/α a saturation occurs such that the velocity approaches the velocity of light in the junction, c. This is often referred to as relativistic behaviour. The qualitative appearance of such kink solutions is shown in Fig. 1 for the case of the simple single-fluxon solution. The time derivative of the phase ϕ_t is the voltage, and the fluxon solution of Eq. (6.1) appears as a voltage pulse moving along the long Josephson junction with a velocity given by Eq. (6.2). The simplified picture presented here can be improved by taking boundary effects, multisoliton solutions and fluxon collisions into account; however, for these more elaborate descriptions the reader is referred to Ref. [8], [2], [10], [9]

3 Stacked junctions

A new superconducting system that is described by special fluxon solutions is the long stacked Josephson junction which is sketched in Fig. 2. It consists of N alternating superconducting and isolating layers here shown in the overlap geometry for $N = 4$.

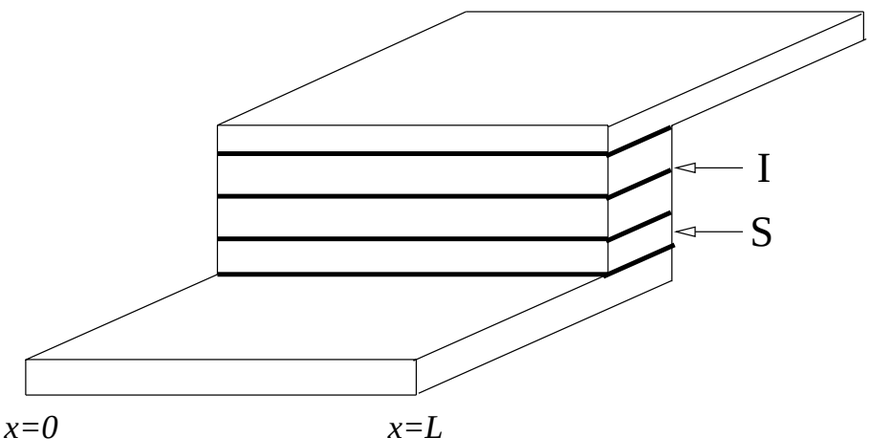

FIGURE 2. An example of a long Josephson stack: the fourfold stack of the overlap geometry. In a typical experiment the magnetic field H is applied in the plane of the tunnel barriers and the bias current flows across them.

The equations for such a stacked junction system may be obtained from a generalization of the Josephson junction equations using the London and Maxwell equations [15]. A simplified version of the stacked junction equations may be written [15] (in unnormalized units)

$$\frac{\hbar}{2e\mu_0}\frac{\partial^2}{\partial x^2}\begin{pmatrix}\phi_{1,0}\\ \vdots\\ \vdots\\ \phi_{i,i-1}\\ \vdots\\ \vdots\\ \phi_{N,N-1}\end{pmatrix} = \tag{6.3}$$

$$\begin{pmatrix} d'_{1,0} & s_1 & 0 & \cdots & & & \\ s_1 & d'_{2,1} & s_2 & 0 & \cdots & & \\ & & \ddots & & & & \\ & 0 & s_{i-1} & d'_{i,i-1} & s_i & 0 & \\ & & & & \ddots & & \\ & & & & & & s_{N-1} \\ & & & & 0 & s_{N-1} & d'_{N,N-1} \end{pmatrix} \begin{pmatrix} J^Z_{1,0}\\ \vdots\\ \vdots\\ J^Z_{i,i-1}\\ \vdots\\ \vdots\\ J^Z_{N,N-1}\end{pmatrix}.$$

Here $\phi_{i,i-1}$ is the phase difference between superconducting layers i and $i-1$, and we have defined the effective magnetic thickness $d'_{i,i-1}$

$$d'_{i,i-1} = d_{i,i-1} + \lambda_i \coth\left(\frac{t_i}{\lambda_i}\right) + \lambda_{i-1}\coth\left(\frac{t_{i-1}}{\lambda_{i-1}}\right), \tag{6.4}$$

and the parameter s_i describing the coupling between layers by

$$s_i = -\frac{\lambda_i}{\sinh t_i/\lambda_i}. \tag{6.5}$$

Here λ_i is the London penetration depth λ_L of the superconducting layer i, and t_i is the thickness of this layer, i runs from zero to N, $d_{i,i-1}$ is the thickness of the isolating layer between superconducting layers i and $i-1$. From the usual Josephson junction model we now introduce the capacitive, resistive and Josephson currents between S-layers i and $i-1$, and we define the total current of them, by $J^Z_{i,i-1}$

$$J^Z_{i,i-1} \equiv \frac{\hbar}{2e}C_{i,i-1}\frac{\partial^2\phi_{i,i-1}}{\partial t^2} + \frac{\hbar}{2e}G_{i,i-1}\frac{\partial\phi_{i,i-1}}{\partial t} + J_{i,i-1}\sin\phi_{i,i-1} \tag{6.6}$$

Here $C_{i,i-1}$ is the unit area capacitance, $G_{i,i-1}$ the unit area conductivity and $J_{i,i-1}$ the dc maximum Josephson current density between S-layers i and $i-1$.We note here that an external bias current may be included in Eq. (6.3) if necessary, as described in Ref. [15].

4 Fluxon solutions, selected examples

In the following we will illustrate some of the consequences of the general equations derived in the previous section. The numerical part of the examples are selected so that realistic low-T_c junction parameters are used, and some of the predicted effects have already been observed in experiments in Niobium thin film systems (Ref. [14] gives several examples). With other parameter values, high-T_c systems of the BSCCO type may be simulated.

4.1 The coherent 2-fluxon mode

The equations of the previous section have solutions in which fluxons in different layers move coherently (or phase-locked). This is a very interesting type of motion, which has been studied in planar systems [7],[5]. For simplicity let us assume three superconducting layers with the bias current of the overlapping geometry as in Fig. 2. We assume that the top and bottom layers have the same properties, and the two insulating layers have equal properties, i.e. $N = 3$ superconducting layers, $d_{1,0} = d_{2,1} \equiv d$, $t_2 = t_0$, $\lambda_2 = \lambda_0$ (but may be different from λ_1). The two tunnel junctions have the same properties described by C, G, and J. From Eq. (6.5) we obtain after some calculations (see Ref. [15])

$$\frac{\hbar}{2e\mu_0}\begin{pmatrix} \dfrac{\partial^2\phi_{1,0}}{\partial x^2} \\ \\ \dfrac{\partial^2\phi_{2,1}}{\partial x^2} \end{pmatrix} = -I_B(d'+s)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} d' & s \\ \\ s & d' \end{pmatrix}\begin{pmatrix} \dfrac{\hbar C}{2e}\dfrac{\partial^2\phi_{1,0}}{\partial t^2} + \dfrac{\hbar G}{2e}\dfrac{\partial\phi_{1,0}}{\partial t} + J\sin\phi_{1,0} \\ \\ \dfrac{\hbar C}{2e}\dfrac{\partial^2\phi_{2,1}}{\partial t^2} + \dfrac{\hbar G}{2e}\dfrac{\partial\phi_{2,1}}{\partial t} + J\sin\phi_{2,1} \end{pmatrix}, \quad (6.7)$$

$$d' = d + \lambda_0\coth\left(\frac{t_0}{\lambda_0}\right) + \lambda_1\coth\left(\frac{t_1}{\lambda_1}\right), \qquad s_1 = -\frac{\lambda_1}{\sinh(t_1/\lambda_1)}. \quad (6.8)$$

Let us now look for a coherent mode with the property $\phi_{1,0}(x,t) = \phi_{2,1}(x,t) \equiv \phi(x,t)$. In that case Eq. (6.7) becomes

$$\frac{\hbar}{2e\mu_0}\frac{\partial^2\phi}{\partial x^2} = (d'+s)\left(\frac{\hbar C}{2e}\frac{\partial^2\phi}{\partial t^2} + \frac{\hbar G}{2e}\frac{\partial\phi}{\partial t} + J\sin\phi - I_B\right). \quad (6.9)$$

This equation is almost the same as the perturbed sine-Gordon equation (6.1) known from conventional long Josephson junctions [9]. The main difference is a change of the length scale due to the factor $(d'+s)$ in Eq. (6.9).

If we go through the normalizations as for example done in Ref. [10], we find that the effective Josephson penetration depth $\lambda_J^{(2)}$ becomes

$$\lambda_J^{(2)} = (\hbar/2e\mu_0(d'+s)J)^{\frac{1}{2}} \tag{6.10}$$

and the velocity of light in the barrier is

$$\overline{c}^{(2)} = \frac{1}{\sqrt{\varepsilon\mu_0}}\,(d/(d'+s))^{\frac{1}{2}}\,. \tag{6.11}$$

Compared to the single junction soliton case we note that

$$\frac{\lambda_J^{(2)}}{\lambda_J^{(1)}} = \frac{\overline{c}^{(2)}}{\overline{c}^{(1)}} = \sqrt{\frac{d'}{d'+s}}\,, \tag{6.12}$$

where superscript (1) refer to the single junction soliton case. Since $s < 0$ (Eq. (6.8)) we note that $\overline{c}^{(2)}$ is larger than $\overline{c}^{(1)}$, i.e. we may exceed the velocity of light in the single junction case.

We also note that if we arrange a bias situation such that we extract a bias current I_{BI}, from the center layer such that $I_{BI} = -2I_B$ (see Fig. 4) the solutions discussed above become antisymmetric in the sense that now a soliton and an antisoliton move together in a coherent mode. The equation of motion, the penetration depth and the velocity of light can be obtained from Eqs. (6.9), (6.10) and (6.11) by substituting $-s$ for s. It should be noted that this coherent soliton - antisoliton state is not possible in conventional single long Josephson junctions.

Above we have concentrated on the coherent modes of the $N = 3$ superconducting layer system. Of course other modes than the coherent ones are possible. An example will be shown below.

For $N > 3$ we can follow the same procedure as used above. A complication of course is the boundaries created by the top and bottom layers. For $N \gg 1$ it may be justified to neglect effects from these two layers and assume that all the intervening layers are identical. In that case we may obtain coherent solutions of the same nature as for the $N = 3$ case discussed above. For the N layer system ($N \gg 1$) each of the $N-1$ coherent solitons will then obey the equation

$$\frac{\hbar}{2e\mu_0}\frac{\partial^2\phi}{\partial x^2} = (d'+2s)\left(\frac{\hbar C}{2e}\frac{\partial^2\phi}{\partial t^2} + \frac{\hbar G}{2e}\frac{\partial\phi}{\partial t} + J\sin\phi - I_B\right), \tag{6.13}$$

and the characteristic length and velocity are now

$$\lambda_J^{(N-1)} = \left(\ \hbar/2e\mu_0(d'+2s)J\ \right)^{\frac{1}{2}} \tag{6.14}$$

$$\overline{c}^{(N-1)} = \frac{1}{\sqrt{\varepsilon\mu_0}}\left(\ d/(d'+2s)\ \right)^{\frac{1}{2}}\,. \tag{6.15}$$

4.2 The two modes of the two fluxon case

We have made numerical simulations corresponding to the two fluxon coherent state described in Eq. (6.9). The normalizations are performed using the usual one soliton case with the velocity of light given by $\overline{c}^{(1)}$ and the Josephson penetration depth $\lambda_J^{(1)}$. The parameters are: $N = 3$ (2 junctions), $\alpha = 0.1$, L=5 and the coupling parameter S= -0.3 where the loss parameter α is defined by $\alpha = G(\hbar/2eCJ)^{1/2}$. Three fluxons are trapped in each of the annular junctions A and B, which are biased in series [14].

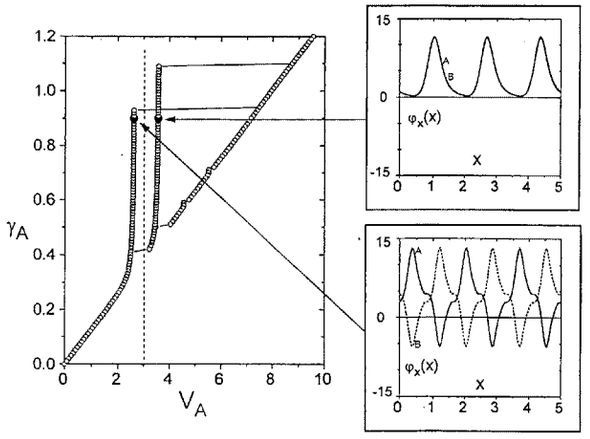

FIGURE 3. Numerical calculation of the IV curve for a twofold annular stack. Parameters are described in the text. The insets shows the waveforms for the two branches corresponding to junctions A and B.

Figure 3 shows the static IV curve for the coherent 2 junction mode. Also shown is another mode found during the simulations. For this mode the motion is symmetric in the sense that when the fluxon in junction (0.1) is at $x = L$ then the fluxon in junction (2.1) is at $x = 0$. The waveforms of the in phase and out of phase modes are shown in the insets. For the coherent mode in Fig. 3 the asymptotic value of the voltage is approximately 1.20 times higher that of the single junction case (dashed line). This factor is in good agreement with the theoretical expression. $\overline{c}^{(2)}/\overline{c}^{(1)} = 1/\sqrt{1+S}$.

We note that the losses appear to play a bigger role for the symmetric mode. Even though the solitons are in different lines, they interact with each other, and "pseudo collisions" take place. The line shape for the in phase mode (coherent) is shown in the upper inset in Fig. 3. Here it should be noted that the two solitons are fully identical, i.e., there is no phaseshift between them, as is usually the case for coherent modes in other comparable systems [6].

The behavior resembles qualitatively the system of two solitons in one junction with surface losses. Also here both the bunched and the symmetric modes exist, although their stability ranges are distinctly different from Fig. 3. However in both cases high bias seems to favor the coherent or bunched mode.

5 Stacked junction plasma oscillation solutions

This topic has very recently become a very active area in connection with high-T_c BSCCO type superconductors (see for example Ref. [3]). For a single small area Josephson junction the derivation of the plasma resonance was made in Ref. [1], [4], [11]. Kirchhoff's law for the Josephson junction equivalent circuit leads to

$$C_J dV/dt + V/R + i_0 \sin\varphi = i_{dc} + i_{rf} \sin\omega t \tag{6.16}$$

with $\partial\phi/\partial t = (2e/\hbar)V$, where all variables and parameters have their standard meaning. Assuming $i_{dc} < i_0$ and i_{rf}/i_0 small, we may assume $\varphi = \varphi_0 + \varphi_1$, with $\sin\varphi_0 = i_{dc}/i_0$ and $\mid \varphi_1 \mid \ll 1$. After some algebra we find [1], [4], [11], [13]

$$\varphi_1 = \frac{2ei_{rf}}{\hbar C_j \sqrt{(\omega^2 - \omega_p^2)^2 + (\omega/CR)^2}} \sin(\omega t + \Theta) \tag{6.17}$$

and the normalized rf-voltage amplitude, V_{rf}, is given by

$$V_{rf} = (\hbar\omega/2e)\varphi_1. \tag{6.18}$$

The plasma frequency ω_p is given by

$$\omega_p = \sqrt{2ei_0 \cos\varphi_0/\hbar C_j} \tag{6.19}$$

where $\sin\varphi_0 = i_{dc}/i_0$ and the maximum plasma frequency $\omega_{p0} = \sqrt{2ei_0/\hbar C_J}$. The Q of the plasma resonance is given by $Q{=}\omega_p C_J R$. We note that the plasma oscillation is a longitudinal oscillation of Cooper pairs across the barrier. Alternatively we may describe it as an inductance-and-capacitance resonance between the Josephson inductance $L_j = \hbar/(2ei_0 \cos\phi_0)$ and the capacitance C_J. It is typically in the microwave frequency range.

Let us now for the stacked junction system (Ref. [12]) consider a small rf bias current in addition to the dc bias current, similarly to the single junction case, Eq. (6.16). In the present paper, however, we will assume a space dependent rf current such that

$$I_B = I_{dc} + I_{rf} \cos(kx - \omega t) \tag{6.20}$$

A spatially dependent bias is physically realistic. For instance, without a ground plane a current concentration near the edges may occur in the superconducting electrodes. In such cases, the rf bias may be expanded in its Fourier components. When the rf term is small and the system becomes linear, the solution for the system is a superposition of the solutions for each Fourier element and thus considering the case of equation (6.20) is general and sufficient. The derivation below follows that of Ref. [12].

For the phase differences $\phi_{i,i-1}$ between the superconducting layers i and $i-1$ we will - in analogy with the derivation in Ref. [11] assume the form

$$\phi_{i,i-1} = \phi^{(0)}_{i,i-1} + \phi^{(1)}_{i,i-1} \tag{6.21}$$

where $\phi^{(0)}_{i,i-1}$ is a dc term and $\mid \phi^{(1)}_{i,i-1} \mid \ll 1$ is assumed to vary at the frequency ω. Accordingly we find

$$\sin\phi_{i,i-1} \cong \sin\phi^{(0)}_{i,i-1} + \cos(\phi^{(0)}_{i,i-1})\phi^{(1)}_{i,i-1}. \tag{6.22}$$

Inserting the bias current, Eq. (6.20), in Eq. (6.3) and applying the approximation Eq. (6.22) we obtain equations for the terms at zero frequency and the terms at the angular frequency ω.

The equation for the dc terms gives

$$I_{dc} = J_{i,i-1}\sin\phi^{(0)}_{i,i-1} \tag{6.23}$$

which determines the dc part of the phases $\phi^{(0)}_{i,i-1}$. For the remaining terms at angular frequency ω we get the following coupled linear equations

$$\frac{\hbar}{2e\mu_0}\partial_{xx}\begin{pmatrix} \phi^{(1)}_{1,0} \\ \vdots \\ \vdots \\ \phi^{(1)}_{i,i-1} \\ \vdots \\ \vdots \\ \phi^{(1)}_{N,N-1} \end{pmatrix} = \tag{6.24}$$

$$\begin{pmatrix} d'_{1,0} & s_1 & & & & & \\ s_1 & d'_{2,1} & s_2 & & & 0 & \\ & \ddots & \ddots & \ddots & & & \\ & & s_{i-1} & d'_{i,i-1} & s_i & & \\ & & & \ddots & \ddots & & \\ & 0 & & & \ddots & \ddots & s_{N-1} \\ & & & & & s_{N-1} & d'_{N,N-1} \end{pmatrix} \begin{pmatrix} \Delta J^{Z(1)}_{1,0} \\ \vdots \\ \vdots \\ \Delta J^{Z(1)}_{i,i-1} \\ \vdots \\ \vdots \\ \Delta J^{Z(1)}_{N,N-1} \end{pmatrix}$$

where

$$\begin{aligned} \Delta J^{Z(1)}_{i,i-1} = \quad & \tfrac{\hbar}{2e} C_{i,i-1} \partial_{tt} \phi^{(1)}_{i,i-1} + \tfrac{\hbar}{2e} G_{i,i-1} \partial_t \phi^{(1)}_{i,i-1} + \\ & J_{i,i-1} \left(\cos \phi^{(0)}_{i,i-1} \right) \phi^{(1)}_{i,i-1} - I_{rf} \cos(kx - \omega t). \end{aligned} \tag{6.25}$$

For mathematical convenience, $I_{rf} \cos(kx - \omega t)$ is here replaced with the complex form $(I_{rf}/2) \exp[j(kx - \omega t)]$, where $j^2 = -1$. As is well known, the final physical quantities can be obtained by a summation of the solution in this case and the solution in the case where the bias has its complex conjugate form. Then the solution of Eq. (6.24) is

$$\begin{pmatrix} H_1 - k^2 & s_1 \Omega_{2,1} & & & & \\ s_1 \Omega_{1,0} & H_2 - k^2 & s_2 \Omega_{3,2} & 0 & & \\ \ddots & \ddots & \ddots & & & \\ & s_{i-1} \Omega_{i-1,i-2} & H_i - k^2 & s_i \Omega_{i+1,i} & & \\ & \ddots & \ddots & \ddots & & \\ & 0 & & \ddots & & \\ & & & & H_N - k^2 \end{pmatrix}$$

$$\times \begin{pmatrix} A_{1,0} \\ \vdots \\ \vdots \\ A_{i,i-1} \\ \vdots \\ \vdots \\ A_{N,N-1} \end{pmatrix} = -\frac{2e\mu_0}{\hbar} \cdot \frac{I_{rf}}{2} \begin{pmatrix} d'_{1,0} + s_1 \\ d'_{2,1} + s_1 + s_2 \\ \vdots \\ d'_{i,i-1} + s_{i-1} + s_i \\ \vdots \\ \vdots \\ d'_{N,N-1} + s_{N-1} \end{pmatrix}, \tag{6.26}$$

$$\Omega_{i,i-1} \equiv \mu_0 C_{i,i-1} \omega^2 - \frac{2e\mu_0}{\hbar} J_{i,i-1} \cos \phi^{(0)}_{i,i-1} + i\omega\mu_0 G_{i,i-1} \,, \tag{6.27}$$

and H_i=$d'_{i,i-1}\Omega_{i,i-1}$.

In the following we will simplify the problem by assuming identical layers. This permits us to drop unnecessary subscripts in the equation above, and by introducing $H = d'\Omega$ and $S = s/d'$, we get

$$\begin{pmatrix} H-k^2 & SH & & & & & \\ SH & H-k^2 & & & & 0 & \\ & \ddots & \ddots & \ddots & & & \\ & & SH & H-k^2 & SH & & \\ & & & \ddots & \ddots & \ddots & \\ & 0 & & & \ddots & \ddots & SH \\ & & & & & SH & H-k^2 \end{pmatrix} \times \begin{pmatrix} A_{1,0} \\ \vdots \\ \vdots \\ A_{i,i-1} \\ \vdots \\ \vdots \\ A_{N,N-1} \end{pmatrix} = -\frac{1}{\lambda_{exc}^2} \begin{pmatrix} 1+S \\ 1+2S \\ \vdots \\ 1+2S \\ \vdots \\ 1+2S \\ 1+S \end{pmatrix} \tag{6.28}$$

where

$$\lambda_{exc}^{-2} = \frac{2e\mu_0 d}{\hbar}\frac{I_{rf}}{2} \tag{6.29}$$

defines a length scale for the applied rf bias current.

First we find the solution to Eq. (6.28) with no applied rf current, i.e. having the right hand side of Eq. (6.28) equal to zero. In that case we find ([12]),

$$k^2 = H\left\{1 - 2S\cos\left(\frac{m\pi}{N+1}\right)\right\}, \quad \text{for m=1,2,...N,} \tag{6.30}$$

where N is the number of stacks, and m is the mode number. The solution $A_{i,i-i}$ for mode number m is

$$A_{i,i-1}^m = \sqrt{\frac{2}{N+1}}\sin\left[\frac{i\,(N+1-m)\,\pi}{N+1}\right]. \tag{6.31}$$

From Eq. (6.30), we get explicitly the $k-\omega$ plasma dispersion relation for each mode m,

$$\omega^2 - \omega_p^2 - [c_m^{(N)}]^2 k^2 + j\omega G/C = 0 \tag{6.32}$$

where

$$\omega_p = \sqrt{2eJ\cos\phi^{(0)}/\hbar C}, \tag{6.33}$$

$$c_m^{(N)} = \frac{c_0}{\left[1 - 2S\cos(m\pi/(N+1))\right]^{1/2}} \tag{6.34}$$

with $c_0 = (\mu_0 d' C)^{-1/2}$. We find from Eq. (6.32) that k is imaginary due to the presence of losses, and thus a mode which is once excited is damped with time. In the lossless limit, we have $k - \omega$ dispersion curves, $\omega^2 = \omega_p^2 + \left[c_m^{(N)}\right]^2 k^2$. Figure 4 shows the dispersion curves for the case of 3 junctions. At $\omega \gg \omega_p$, $d\omega/dk$ approaches asymptotically the Swihart-type velocities, $c_m^{(N)}$, in the case of an N junction stack, i.e., the characteristic velocity of electromagnetic waves in the N-stack system with no tunnel currents. As will be shown below, in the presence of an rf bias current, plasma resonances may appear on some selected modes of these $k - \omega$ dispersion curves.

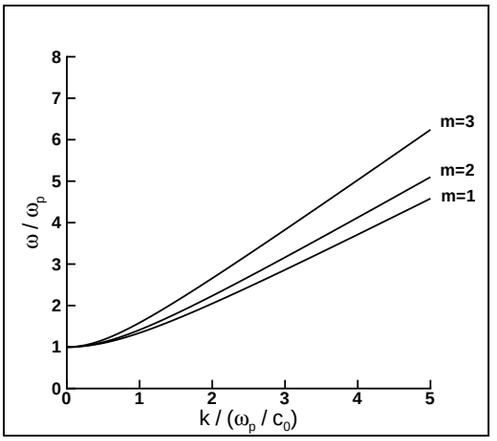

FIGURE 4. Plasma dispersion curves for 3 junctions. S = -0.4.

As the simplest non-trivial case let us first study the case $N = 2$ with an rf excitation of the form $I_{rf}\cos(kx - \omega t)$. Equation (6.20) gives a solution of the form

$$\begin{Bmatrix} \phi_{1,0}^{(1)} \\ \phi_{2,1}^{(1)} \end{Bmatrix} = \frac{2eI_{rf}}{\hbar C} \frac{\cos(kx - \omega t + \theta)}{\sqrt{\left(\omega^2 - \omega_p^2 - c_+^2 k^2\right) + (\omega G/C)^2}} \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \tag{6.35}$$

$$\text{with} \quad \tan\theta = \frac{\omega G/C}{\left(\omega^2 - \omega_p^2 - c_+^2 k^2\right)}, \text{ and } c_2^{(2)} \equiv c_+ = \frac{c_0}{\sqrt{1+S}}.$$

We notice that Eq. (6.35) has a form analogous to Eq. (6.17) for the single junction case. Note also that $\phi_{1,0}^{(1)}$ and $\phi_{2,1}^{(1)}$ are in phase and identical, and their magnitudes are enhanced on the plasma dispersion curve corresponding to the c_+ mode, i.e., the plasma resonance takes place on the c_+ mode. On the other hand on the c_- mode with the characteristic velocity, $c_1^{(2)} \equiv c_- = c_0/\sqrt{1-S}$, and where the opposite phase relation $(\phi^{(1)})_{1,0} = -\phi_{2,1}^{(1)})$ is found, a resonance does not appear, because, for reasons of symmetry, it is not being exited by I_{rf}, applied in series through the two junctions.

In the case of three junction stack ($N = 3$), the solution of Eq. (6.27) becomes ([12]), (note that $A = \phi^{(1)} \propto V_{rf}$)

$$A_{1,0} = A_{3,2} = -\frac{1}{\lambda_{exc}^2} \cdot \frac{(1+S)(H^2 - k^2) + (1+2S)SH}{\left[(1+\sqrt{2}S)H - k^2\right] \cdot \left[(1-\sqrt{2}S)H - k^2\right]}, \quad (6.36)$$

$$A_{2,1} = -\frac{1}{\lambda_{exc}^2} \cdot \frac{(1+2S)(H^2 - k^2) - 2(1+S)SH}{\left[(1+\sqrt{2}S)H - k^2\right] \cdot \left[(1-\sqrt{2}S)H - k^2\right]}. \quad (6.37)$$

We change the terms appearing in the denominators in Eq. (6.36) and (6.37) to the more explicit forms

$$\left(1+\sqrt{2}S\right)H - k^2 = c_1^{-2}\left(\omega^2 - \omega_p^2 - c_1^2 k^2\right) + i\omega G/C \quad (6.38)$$

and

$$\left(1-\sqrt{2}S\right)H - k^2 = c_3^{-2}\left(\omega^2 - \omega_p^2 - c_3^2 k^2\right) + i\omega G/C \quad (6.39)$$

$$\text{with} \quad c_1^{(3)} \equiv c_1 = \frac{c_0}{\sqrt{1-\sqrt{2}S}} \quad \text{and} \quad c^{(3)} \equiv c_3 = \frac{c_0}{\sqrt{1+\sqrt{2}S}}.$$

These equations show that resonances take place on the plasma dispersive curves of the c_1 and c_3 modes. (See Fig. 3.) The c_2 mode does not appear as a resonance, which is also understandable from symmetry considerations. Equation (6.36) also assures $\phi_{1,0}^{(1)} = \phi_{3,2}^{(1)}$, in any case, i.e. the oscillation of the first and third junctions are completely in phase and identical.

In order to determine the relationship between A_{10} and A_{21} we take the ratio A_{21}/A_{10}. Let us first consider the case where the $k-\omega$ relationship is on the c_3 mode curve, i.e., $\omega^2 = \omega_p^2 + c_3^2 k^2$. In this case we find [12],

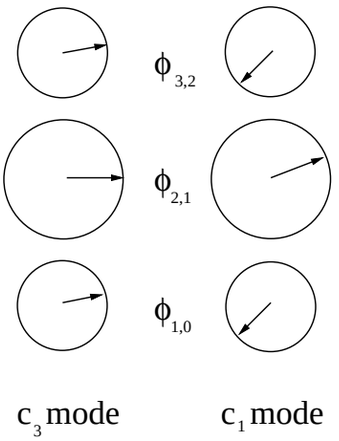

FIGURE 5. Schematic drawing of the complex junction voltages for the two resonance modes of a three-stack

$$\frac{A_{2,1}}{A_{1,0}} = \frac{-\sqrt{2}k^2c_0^2\left(1+\sqrt{2}\right)S + j\omega G/C}{-k^2c_0^2\left(1+\sqrt{2}\right)S + j\omega G/C}. \tag{6.40}$$

From this equation and recalling $S < 0$, we find that A_{21} and A_{10} changes with time almost in phase and $\mid A_{21} \mid > \mid A_{10} \mid$ that is shown in Fig. 5 (left part). Due to the losses, the phase of A_{10} is slightly delayed with respect to the phase of A_{21}. In the lossless limit, we see that A_{21} and A_{10} are perfectly in phase and $\mid A_{21} \mid / \mid A_{10} \mid = \sqrt{2}$.

Similarly, at the resonance condition of the c_1 mode($\omega^2 = \omega_p^2 + c_1^2k^2$), we obtain

$$\frac{A_{2,1}}{A_{1,0}} = \frac{-\sqrt{2}k^2c_0^2\left(\sqrt{2}-1\right)S + j\omega G/C}{-k^2c_0^2\left(\sqrt{2}-1\right)S + j\omega G/C}. \tag{6.41}$$

Again we find that $\mid A_{21} \mid > \mid A_{10} \mid$, but A_{21} and A_{10} changes with time almost in an anti-phase manner as shown in Fig. 5 (right part). In the lossless limit, $\mid A_{21} \mid / \mid A_{10} \mid = \sqrt{2}$, and A_{21} has the perfectly opposite phase of A_{32}. From Fig. 5, a large voltage amplitude across the stack may be expected as for the c_3 resonance condition.

For the case of N-stacks we find that the highest phase velocity $c_N^{(N)}$ mode always corresponds to all junctions oscillating in phase as the plasma resonance. The general picture may be understood from Fig. 6. Possible dispersion modes with stack plasma resonances are found as every second mode by counting downwards from the $c_N^{(N)}$ mode, which has the highest velocity.

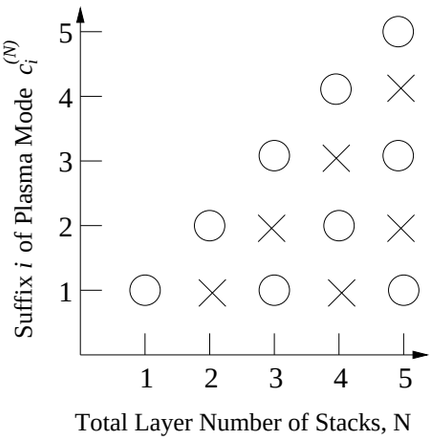

FIGURE 6. Resonance modes of an N-stack. Circles represent allowed modes and crosses, prohibited modes.

6 Conclusion

A system consisting of alternating layers of superconducting and insulating thin films has been considered. We assume the dimensions in the plane to be large so that the system resembles stacked long Josephson junctions. An implicit assumption is that the film thickness is not large compared to the London penetration depth. This condition can be readily met in niobium-nitride or niobium systems, as well as in high-T_c BSCCO systems.

Our analysis shows a variety of interesting behavior, of particular interest is the mode of $N-1$ coherent solitons in a structure with N superconducting layers. This mode may have practical interest in connection with phase-locking of Josephson oscillators. Other types of motion that we observed in numerical simulations are unique to the three dimensional structure, and do not exist in the one or two dimensional sine-Gordon systems that have mostly been considered until now.

7 References

[1] P. W. Anderson. In E. R. Caianiello, editor, *Lectures on the Many - Body Problem, Vol. II*, p. 113, New York, 1964. Academic.

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

[3] In M. Tachiki and T. Yamashita, editors, *Proceedings of the First International Symposium on Intrinsic Josephson effects and THz plasma oscillations*

in high Tc superconductors, North-Holland, 1997.

[4] A. J. Dahm, A. Denenstein, T. F. Finnegan, D. N. Langenberg, and D. J. Scalapino. *Phys. Rev. Lett.*, **20**, 859E, 1968, Erratum ibid **20**, 1020, 1968.

[5] A. Davidson, N. Grønbech-Jensen, and N. F. Pedersen. *IEEE Trans. on Magn.*, **27**, 3347, 1991.

[6] P. S. Lomdahl, O. H. Sørensen, and P. L. Christiansen. *Phys.Rev. B*, **25**, 5737, 1982.

[7] R. Monaco, S. Pagano, and G. Costabile. *Phys. Lett. A*, **131**, 122, 1988.

[8] D. W. McLaughlin and A. C. Scott. *Phys.Rev. A*, **18**, 1652, 1978.

[9] R. D. Parmentier. In H. Weinstock and R. W. Ralston, editors, *The New Superconducting Electronics*, p. 469, Dordrecht, 1993. Kluwer.

[10] N. F. Pedersen. In S. E. Trullinger, V. E. Zakharov, and V. L. Pokrovsky, editors, *Solitons*, p. 469, Amsterdam, 1986. Elsevier.

[11] N. F. Pedersen, T. F. Finnegan, and D. N. Langenberg. *Phys. Rev. B*, **6**, 4151, 1972.

[12] N. F. Pedersen and S. Sakai. *Phys. Rev. B*, **58**, 2820, 1998.

[13] N. F. Pedersen, M. R. Samuelsen, and K. Saermark. *J. Appl Phys.*,**44**, 5120, 1973.

[14] N. F. Pedersen and A. V. Ustinov. *Supercond. Sci. Technol.*, **8**, 389, 1995.

[15] S. Sakai, P. Bodin, and N. F. Pedersen. *J. Appl. Phys.*, **73**, 2411, 1993.

7

Dynamics of Vortices in Two-Dimensional Magnets

Franz G. Mertens
Alan R. Bishop

ABSTRACT Theories, simulations and experiments on vortex dynamics in quasi-two-dimensional magnetic materials are reviewed. These materials can be modelled by the classical two-dimensional anisotropic Heisenberg model with XY (easy-plane) symmetry. There are two types of vortices, characterized by their polarization (a second topological charge in addition to the vorticity): Planar vortices have Newtonian dynamics (even-order equations of motion) and exhibit strong discreteness effects, while non-planar vortices have non-Newtonian dynamics (odd-order equations of motion) and smooth trajectories. These results are obtained by a collective variable theory based on a generalized travelling wave ansatz which allows a dependence of the vortex shape on velocity, acceleration etc.. An alternative approach is also reviewed and compared, namely the coupling of the vortex motion to certain quasi-local spinwave modes.
The influence of thermal fluctuations on single vortices is investigated. Different types of noise and damping are discussed and implemented into the microscopic equations which yields stochastic equations of motion for the vortices. The stochastic forces can be explicitly calculated and a vortex diffusion constant is defined. The solutions of the stochastic equations are compared with Langevin dynamics simulations. Moreover, noise-induced transitions between opposite polarizations of a vortex are investigated.
For temperatures above the Kosterlitz-Thouless vortex-antivortex unbinding transition, a phenomenological theory, namely the vortex gas approach, yields central peaks in the dynamic form factors for the spin correlations. Such peaks are observed both in combined Monte Carlo- and Spin Dynamics-Simulations and in inelastic neutron scattering experiments. However, the assumption of ballistic vortex motion appears questionable.

1 Introduction

During the past 15 years an increasing interest in two-dimensional (2D) magnets has developed. This is a result of (i) the investigation of a wide class of well-characterized quasi-2D magnetic materials that allow a detailed experimental study of their properties (inelastic neutron scatter-

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 137−170, 2000.

ing, nuclear magnetic resonance, etc.), and (ii) the availability of high-speed computers with the capabilities for simulations of large lattices. Examples of these materials are (1) layered magnets [1], like K_2CuF_4, Rb_2CrCl_4, $(CH_3NH_3)_2CuCl_4$ and $BaM_2(XO_4)_2$ with $M = Co, Ni, \ldots$ and $X = As, P, \ldots$; (2) $CoCl_2$ graphite intercalation compounds [2], (3) magnetic lipid layers [3], like $Mn(C_{18}H_{35}O_2)_2$. For the first class of these examples, the ratio of inter- to intraplane magnetic coupling constants is typically 10^{-3} to 10^{-6}. This means that the behavior is nearly two-dimensional as concerns the magnetic properties. For the second class, the above ratio of the coupling constants can be tuned by choosing the number of intercalated graphite layers. For the third class, even monolayers can be produced, using the Langmuir-Blodgett method.

Many of the above materials have an "easy-plane" or XY-symmetry. This means that the spins prefer to be oriented in the XY-plane, which is defined as the plane in which the magnetic ions are situated. The simplest model for this symmetry is the 2D classical anisotropic Heisenberg Hamiltonian (see Ref. [4])

$$H = -J \sum_{<m,n>} [S_x^m S_x^n + S_y^m S_y^n + (1-\delta) S_z^m S_z^n] \,, \tag{7.1}$$

where $< m, n >$ labels nearest neighbors of a 2D lattice; usually a square lattice is used. The subscripts x, y and z stand for the components of the classical spin vector $\mathbf{S}$; the spin length S is set to unity by the redefinition $J \to J/S^2$. J is the magnetic coupling constant, both ferro- and antiferromagnetic materials were investigated. The anisotropy parameter δ lies in the range $0 < \delta \leq 1$; note that $\delta = 1$ corresponds to the XY-model, not to the so-called planar model where the spins are strictly confined to the XY-plane; this confinement is possible only if one is not interested in the dynamics. Instead of the exchange anisotropy in the Hamiltonian (7.1), one can also use an on-site anisotropy term $(S_z^m)^2$, which yields similar results.

There are two kinds of excitations: Spin waves which are solutions of the linearized equations of motion, and vortices which are topological collective structures. The vortices are responsible for a topological phase transition [5, 6] at the Kosterlitz-Thouless transition temperature T_{KT}. Below T_{KT}, vortex-antivortex pairs are thermally excited and destroyed; above T_{KT} these bound pairs dissociate and the density of free vortices increases with temperature. It must be noted that an order-disorder phase transition is not possible according to the Mermin-Wagner-Theorem [7]; the reason is that all long-range order is destroyed by the long-wave linear excitations in all 1D and 2D models with continuous symmetries.

Both theory [8] and Monte Carlo simulations [9] showed that T_{KT} is only very weakly dependent on the anisotropy δ, except for δ extremely close to 0, when $T_{\mathrm{KT}} \to 0$. For the materials mentioned above, coupling constants were estimated from fits to spin-wave theory, and δ values are in the range

$0.01 - 0.6$, where T_{KT} is still close to its value for $\delta = 1$.

However, the out-of-plane structure of the vortices (i. e., the structure of the S_z components) depends crucially on δ, while the in-plane structure (S_x and S_y components) remains the same [10]. For $\delta > \delta_c (\approx 0.297$ for a square lattice [11]), static vortices have null S_z components, they are termed "in-plane" or planar vortices in the literature. For $\delta < \delta_c$ there are "out-of-plane" or "non-planar" vortices which exhibit a localized structure of the S_z components around the vortex center. This structure falls off exponentially with a characteristic length, the vortex core radius [10]

$$r_v = \frac{1}{2}\sqrt{\frac{1-\delta}{\delta}}\,, \tag{7.2}$$

in units of the lattice constant a. The core radius increases with diminishing δ, allowing a continuous crossover to the isotropic Heisenberg model ($\delta = 0$) where the topological excitations are merons and instantons [12], rather than vortices.

Compared to the vortices in classical fluids [13] and superfluids [14], there is an important difference: The vortices of the easy-plane Heisenberg model carry *two* topological charges, instead of one: (1) The vorticity $q = \pm 1, \pm 2, \ldots$ which is defined in the usual way: The sum of changes of the azimuthal angle $\phi = \arctan(S_y/S_x)$ on an arbitrary closed contour around the vortex center yields $2\pi q$; if the center is not inside the contour the sum is zero. In the following only the cases $q = +1$ (vortex) and $q = -1$ (antivortex) will be considered, the cases $|q| > 1$ would be relevant only for high temperatures. (2) The polarization or polarity p: For the non-planar vortices $p = \pm 1$ indicates to which side of the XY-plane the out-of-plane structure points, while $p = 0$ for the planar vortices (Fig. 1).

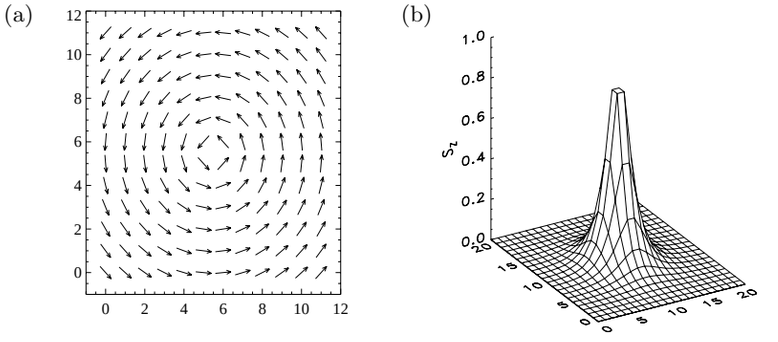

FIGURE 1. (a) In-plane structure of a static planar vortex ($q = +1$). (b) Out-of-plane structure of a static non-planar vortex with polarization $p = +1$.

Both topological charges [1] are crucial for the vortex dynamics, which is governed by two different forces: (1) 2D Coulomb-type forces $\mathbf{F}$, which are proportional to the product of the vorticities of two vortices and inversely proportional to their distance (assuming that the distance is larger than $2r_v$, such that the out-of-plane structures do not overlap), (2) a "gyrocoupling" force [16, 17]

$$\mathbf{F}_G = \dot{\mathbf{X}} \times \mathbf{G} \tag{7.3}$$

where $\dot{\mathbf{X}}$ is the vortex velocity, and $\mathbf{X}(t)$ is the trajectory of the vortex center.

The force (7.3) is formally equivalent to the Magnus force in fluid dynamics and to the Lorentz force on an electric charge. The gyrocoupling vector [18]

$$\mathbf{G} = 2\pi qp\, \mathbf{e}_z\,, \tag{7.4}$$

where $\mathbf{e}_z$ is the unit vector in z-direction, does not represent an external field but is an intrinsic quantity, namely a kind of self-induced magnetic field which is produced by the localized S_z-structure and carried along by the vortex. For antiferromagnets $\mathbf{G}$ vanishes.

Since the gyrovector (7.4) contains the product of both topological charges, the dynamics of the two kinds of vortices is completely different:
(1) For planar vortices $\mathbf{G} = \mathbf{0}$, therefore they have a Newtonian dynamics

$$M\ddot{\mathbf{X}} = \mathbf{F}\,, \tag{7.5}$$

where the vortex mass M will be defined in section 2.2. However, in the simulations the trajectories are not smooth due to strong discreteness effects [19].
(2) Non-planar vortices have smooth trajectories if the diameter $2r_v$ of the out-of-plane structure is considerably larger than the lattice constant [19]; i. e., if δ is not close to δ_c. For steady state motion, the dynamics is described by the Thiele equation [16, 17]

$$\dot{\mathbf{X}} \times \mathbf{G} = \mathbf{F}\,, \tag{7.6}$$

which was derived from the Landau-Lifshitz equation (section 2) for the spin vector $\mathbf{S}^m(t)$. This equation is identical to Hamilton's equations with Hamiltonian (1.1).

However, for arbitrary motion the Thiele equation (1.5) is only an approximation because a rigid vortex shape was assumed in the derivation. If a velocity dependence of the shape is allowed [20, 21], an inertial term $M\ddot{\mathbf{X}}$ appears on the l. h. s. of Eq. (7.6). However, this second-order equation could not be confirmed by computer simulations [23, 24]. The reason will

[1] From the viewpoint of homotopy groups, q and $Q = -\frac{1}{2}qp$ are π_1- and π_2-topological charges, respectively; see Ref. [15].

be discussed in section 2; in fact, a collective variable theory reveals that the dynamics of non-planar vortices can only be described by odd-order equations of motion. Therefore the dynamics is *non-Newtonian*.

Another interesting topic is the dynamics under the influence of thermal fluctuations (section 3). Here white noise and damping are implemented in the Landau-Lifshitz equation. The same type of collective variable theory as in section 2 then yields stochastic equations of motion for the vortices. The stochastic forces on the vortices can be calculated and the solutions of the equations of motion are compared with Langevin dynamics simulations. In this way, the diffusive vortex motion can be well understood.

For somewhat higher temperatures (still below T_{KT}) the thermal noise can induce transitions of non-planar vortices from one polarization to the opposite one. Theoretical estimates of the transition rate are compared with Langevin dynamics simulations (section 3.3).

For the temperature range above T_{KT}, so far only a phenomenological theory exists, the vortex-gas approach (section 4). It contains only two free parameters (the density of free vortices and their r.m.s. velocity). The theory predicts "central peaks" in the dynamic form factors for both in-plane and out-of-plane spin correlations. Such peaks are observed both in computer simulations and in inelastic neutron scattering experiments, and many of the predicted features are confirmed. However, a recent *direct* observation of the vortex motion in Monte Carlo simulations has revealed that a basic assumption of the theory, namely a *ballistic* vortex motion, might not be valid. (In section 4 we confine ourselves to the ferromagnetic case, although easy-plane antiferromagnets were also investigated).

The conclusion in section 5 will discuss the ingredients of a theory which can fully explain all observations above T_{KT}.

2 Collective variable theories at zero temperature

2.1 Thiele equation

The spin dynamics is given by the Landau-Lifshitz equation, which is the classical limit of the Heisenberg equation of motion for the spin operator $\mathbf{S}^m$,

$$\frac{\mathrm{d}\mathbf{S}^m}{\mathrm{d}t} = -\mathbf{S}^m \times \frac{\partial H}{\partial \mathbf{S}^m} \tag{7.7}$$

where H is the Hamiltonian, in our case that of the anisotropic Heisenberg model (7.1). The meaning of Eq. (7.7) is that the spin vector $\mathbf{S}^m$ precesses in a local magnetic field $\mathbf{B}$, with cartesian components $B_\alpha = -\partial H/\partial S^m_\alpha$. In spin dynamics simulations (see below), the Landau-Lifshitz equation is integrated numerically for a large square lattice, typically with 72×72 lattice points. As initial condition a vortex is placed on the lattice and the trajectory $\mathbf{X}(t)$ of the vortex center is monitored.

This trajectory is then compared with theory, i. e. with the solution of an equation of motion for the vortex. The standard procedure to obtain this equation consists in taking the continuum limit $\mathbf{S}^m(t) = \mathbf{S}(\mathbf{r}, t)$ where $\mathbf{r}$ is a vector in the XY plane, and in developing a *collective variable theory*. The simplest version makes the travelling wave ansatz [16, 17]

$$\mathbf{S}(\mathbf{r}, t) = \mathbf{S}(\mathbf{r} - \mathbf{X}(t)) \,, \tag{7.8}$$

where the functions S_α on the r. h. s. describe the vortex shape. (Strictly speaking, these functions should bear an index to distinguish them from the functions on the l. h. s.)

As the equation of motion is expected to contain a force, the following operations are performed with Eq. (7.7) which yield force densities

$$\mathbf{S}\left(\frac{\partial \mathbf{S}}{\partial X_i} \times \frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t}\right) = -\mathbf{S}\left(\frac{\partial \mathbf{S}}{\partial X_i} \times \left[\mathbf{S} \times \frac{\delta H}{\delta \mathbf{S}}\right]\right) = -S^2 \frac{\delta H}{\delta \mathbf{S}} \frac{\partial \mathbf{S}}{\partial X_i} = -S^2 \frac{\partial \mathcal{H}}{\partial X_i} \tag{7.9}$$

with $i = 1, 2$ and Hamiltonian density $\mathcal{H}$. According to the ansatz (7.8),

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{\partial \mathbf{S}}{\partial X_j} \dot{X}_j \tag{7.10}$$

is inserted, with summation over repeated indices. Integration over $\mathbf{r}$ then yields the equation of motion

$$\mathbf{G}\dot{\mathbf{X}} = \mathbf{F} \,, \tag{7.11}$$

with the force

$$F_i = -\int \mathrm{d}^2 r \, \frac{\partial \mathcal{H}}{\partial X_i} \,, \tag{7.12}$$

and the gyrocoupling tensor

$$G_{ij} = \int \mathrm{d}^2 r \, \mathbf{S} \frac{\partial \mathbf{S}}{\partial X_i} \times \frac{\partial \mathbf{S}}{\partial X_j} = \int \mathrm{d}^2 r \, \left\{ \frac{\partial \phi}{\partial X_i} \frac{\partial \psi}{\partial X_j} - \frac{\partial \phi}{\partial X_j} \frac{\partial \psi}{\partial X_i} \right\} . \tag{7.13}$$

The expression on the right was obtained by introducing the two canonical fields

$$\phi = \arctan(S_y / S_x), \; \psi = S_z \tag{7.14}$$

instead of the three spin components S_α with the constraint $S = 1$. The static vortex structure is [10]

$$\phi_0 = q \arctan \frac{x_2'}{x_1'} \tag{7.15}$$

$$\psi_0 = p \left[1 - a_1^2 \left(\frac{r'}{r_v} \right)^2 \right] \quad \text{for } r' \ll r_v \tag{7.16}$$

$$\psi_0 = p a_2 \sqrt{\frac{r_v}{r'}} \mathrm{e}^{-r'/r_v} \quad \text{for } r' \gg r_v \tag{7.17}$$

where the constants a_1 and a_2 can be used for a matching, and

$$\mathbf{r}' = \mathbf{r} - \mathbf{X}\,. \tag{7.18}$$

The integral then yields

$$G_{ij} = G\epsilon_{ij}\,,\; G = 2\pi qp \tag{7.19}$$

where ϵ_{ij} is the antisymmetric tensor. Interestingly, only the value p of the S_z component at the vortex *center* enters the final result [18], i. e., the out-of-plane vortex structure needs not to be explicitly known here.

As $\mathbf{G}$ is antisymmetric, Eq. (7.11) is identical to the Thiele equation (7.6) with gyrovector $\mathbf{G} = G\mathbf{e}_z$. Since $p = 0$ and thus $G = 0$ for planar vortices, the Thiele equation is incomplete in this case. Obviously there must be a non-vanishing term on the l. h. s., this will be an inertial term (section 2.2). Moreover, the ansatz (7.8), and thus the Thiele equation, is only valid for steady-state motion when the vortex shape is rigid (in the moving frame). This includes not only translational motion with constant velocity V_0 but also rotational motion with constant angular velocity ω_0. Both types of motion can be obtained by considering two non-planar vortices at a distance $2R_0$ which drive each other by their Coulomb force $F = 2\pi q_1 q_2/(2R_0)$. For a certain velocity this force is compensated by the gyrocoupling force (7.3). In fact, since each vortex carries two charges, there are four physically different scenarios which represent stationary solutions of the Thiele equation. They fall into two classes: If the gyrovectors of the two vortices are parallel (i. e., $q_1p_1 = q_2p_2$), a vortex-vortex or vortex-antivortex pair rotates with ω_0 on a circle with radius R_0 around each other, where

$$\omega_0 R_0^2 = \frac{1}{2}\,\frac{q_2}{p_1} = \frac{1}{2}\,\frac{q_1}{p_2}\,. \tag{7.20}$$

If the gyrovectors are antiparallel, the pair performs a parallel translational motion with velocity V_0 and distance $2R_0$, where $V_0R_0 = q_2/(2p_1) = -q_1/(2p_2)$.

Both kinds of scenarios also appear for vortices in (super)fluids [13, 14]. However, in these contexts there are only two physically different situations: vortex-vortex rotation and vortex-antivortex translation.

2.2 Vortex mass

The above scenarios for magnetic vortices were not tested by computer simulations until 1994, and the results were surprising [22, 23]: Using a square or circular system with free boundary conditions, two of the four scenarios (vortex-vortex rotation and vortex-antivortex translation) were confirmed, but not the two other ones: For vortex-antivortex rotation and vortex-vortex translation the observed trajectories showed pronounced oscillations around the trajectories predicted by the Thiele equation (7.6) or (7.11), see Fig. 2.

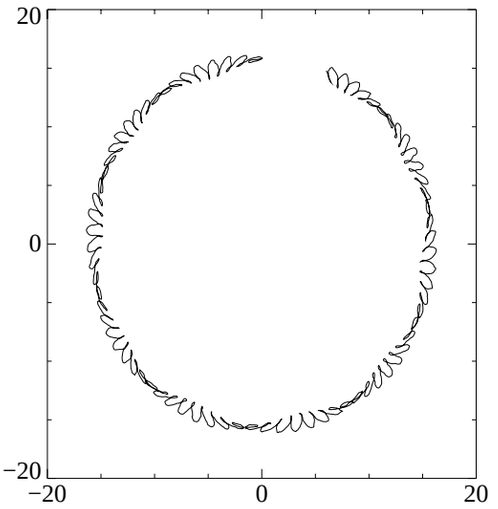

FIGURE 2. Vortex-antivortex rotation on a circular system with radius $L = 36$ and free boundary conditions. Only the trajectory of the vortex center is plotted, the center of the antivortex is always opposite. The mean trajectory is a circle with radius $R_0 = 15.74$.

In fact, such oscillations had already been predicted [20, 21] by assuming that the vortex shape in the travelling wave ansatz (7.8) depends explicitly on the velocity $\dot{\mathbf{X}}$. This leads to an additional term $(\partial \mathbf{S}/\partial \dot{X}_j)\,\ddot{X}_j$ in Eq. (7.10) and thus to an inertial term in the Thiele equation

$$\mathbf{M}\ddot{\mathbf{X}} + \mathbf{G}\dot{\mathbf{X}} = \mathbf{F}\,, \tag{7.21}$$

where $\mathbf{M}$ is the mass tensor

$$M_{ij} = \int \mathrm{d}^2 r\, \mathbf{S} \frac{\partial \mathbf{S}}{\partial X_i} \times \frac{\partial \mathbf{S}}{\partial \dot{X}_j} = \int \mathrm{d}^2 r \left\{ \frac{\partial \phi}{\partial X_i} \frac{\partial \psi}{\partial \dot{X}_j} - \frac{\partial \phi}{\partial \dot{X}_j} \frac{\partial \psi}{\partial X_i} \right\}. \tag{7.22}$$

This integral can be easily evaluated if the vortex is placed at the center of a circular system with radius L and free boundary conditions. The dominant contribution stems from the outer region $a_c \le r \le L$, with $a_c \simeq 3r_v$, where the velocity dependent parts of the vortex structure are [10, 24]

$$\psi_1 = \frac{q}{4\delta r^2}(x_2 \dot{X}_1 - x_1 \dot{X}_2)\,. \tag{7.23}$$

$$\phi_1 = p(x_1 \dot{X}_1 + x_2 \dot{X}_2) \tag{7.24}$$

Together with the static parts ϕ_0 from Eq. (7.15) and ψ_0, which falls off exponentially outside the core (7.2), the rest mass M is obtained

$$M_{ij} = M\delta_{ij}\,, \quad M = \frac{\pi q^2}{4\delta} \ln \frac{L}{a_c} + C_M\,. \tag{7.25}$$

The constant C_M stems from the inner region $0 \le r \le a_c$, where the vortex structure is not well known due to the discreteness, but C_M is not important for discussion. There is also a velocity dependent contribution to the mass which is negligible because the vortex velocities in the simulations are always much smaller than the spin wave velocity, which is the only characteristic velocity of the system.

The above vortex mass is consistent with other results in the literature: Generalizing the momentum of solitons in 1D magnets [25], the vortex momentum

$$\mathbf{P} = -\int \mathrm{d}^2 r\, \psi \nabla \phi \tag{7.26}$$

is defined and can be shown to be a generator of translations [21]. Then $\mathbf{P} = M\dot{\mathbf{X}}$ results, but $\mathbf{P}$ is not a canonical momentum because the Poisson bracket $\{P_1, P_2\} = G$ does not vanish. For the kinetic energy $M\dot{\mathbf{X}}^2/2$ is obtained [10], therefore this energy and the rest energy $E = \pi \ln(L/a_c) + C_E$ both show the same logarithmic dependence on the system size L.

For a continuum model of a 2D ferromagnet with uniaxial symmetry and a magnetostatic field, a slightly different vortex momentum was defined [26], namely

$$\mathbf{P}_{PT} = \int \mathrm{d}^2 r\, \mathbf{r} \times \mathbf{g}\,, \tag{7.27}$$

where $\mathbf{g} = \nabla\phi \times \nabla\psi$ represents the gyrovector density, cf. Eq. (7.13), which is related to the gyrovector as described below Eq. (7.19). The definition (7.27) depends on the choice of origin of the system and is not a generator of translations. Nevertheless, if the time derivative of $\mathbf{P}_{PT}$ is set equal to the force $\mathbf{F}$ on the vortex, the generalized Thiele equation (7.21) is again obtained [21]. Therefore the vortex dynamics is qualitatively the same, as will be discussed now.

Eq. (7.21) has the same form as that for an electric charge e in a plane with a perpendicular magnetic field $\mathbf{B}$ and an in-plane electric field $\mathbf{E}$. I. e., the vortex motion is completely analogous to the Hall effect: The trajectory is a cycloid with frequency

$$\omega_c = \frac{G}{M}\,, \tag{7.28}$$

cf. the cyclotron frequency $eB/(Mc)$, where c is the speed of light. It is possible to transform to a frame where the force $\mathbf{F}$ vanishes and the vortex rotates (i. e., guiding center coordinates [27, 26, 21]).

The cycloidal trajectories can explain *qualitatively* the oscillations observed in the simulations (Fig. 2). Moreover, the fact that oscillations are

observed only for two of the four scenarios for two vortices (see above) can also be explained: A generalization of the ansatz (7.8) to the case of two vortices yields two coupled equations of motion with a 2-vortex mass tensor [28, 23]. In the case of vortex-vortex rotation and vortex-antivortex translation there are cancellations in this mass tensor which prevent oscillations, in agreement with the simulations.

However, a *quantitative* comparison of Eq. (7.21) and the simulations reveals two severe discrepancies [23, 24]:
(1) The mass $M = G/\omega_c$ which is obtained from Eq. (7.28) by inserting the observed frequency, turns out to be much larger than predicted by Eq. (7.25). Moreover, the dependence on the system size L is linear, while Eq. (7.25) predicts a logarithmic dependence.
(2) Instead of the *one* frequency ω_c of Eq. (7.28), *two* frequencies $\omega_{1,2}$ are observed in the spectrum of the oscillations (Fig. 3). As ω_1 and ω_2 are close to each other, a pronounced beat is observed in the trajectories (Fig. 2): The cycloidal frequency is $(\omega_1+\omega_2)/2$, but the shape of the cycloids changes slowly with $(\omega_2 - \omega_1)/2$.

Both discrepancies have nothing to do with 2-vortex effects, because they also occur in 1-vortex simulations [24]. Here the vortex is driven by image forces, in analogy to electrostatics. The most convenient geometry is a circular system with radius L and free boundaries. In this case there is only one image vortex, which has opposite vorticity but the same polarization [22, 29]. The radial coordinate of the image is L^2/R, where R is the vortex coordinate. The vortex trajectory can be fitted very well to a superposition of two cycloids. This observation will be explained in the next section.

2.3 Hierarchy of equations of motion

Recently a general collective variable theory was developed for nonlinear coherent excitations in classical systems with *arbitrary* Hamiltonians [24]. In this theory the dynamics of a single excitation is governed by a hierarchy of equations of motion for the excitation center $\mathbf{X}(t)$. The *type* of the excitation determines on which levels the hierarchy can be truncated consistently: "Gyrotropic" excitations are governed by odd-order equations and thus do not have Newtonian dynamics, e. g. Galileo's law is not valid. "Non-gyrotropic" excitations are so-to-speak the normal case, because they are described by even-order equations, i. e. by Newton's equation in the first approximation.

Examples of the latter class are kinks in 1D models like the nonlinear Klein-Gordon family and the planar vortices of the 2D anisotropic Heisenberg model (7.1). The non-planar vortices of this model represent the simplest gyrotropic example. 3D models have not been considered so far.

The basis of the above collective variable theory is a "*generalized travelling wave ansatz*" for the canonical fields in the Hamiltonian. For a spin

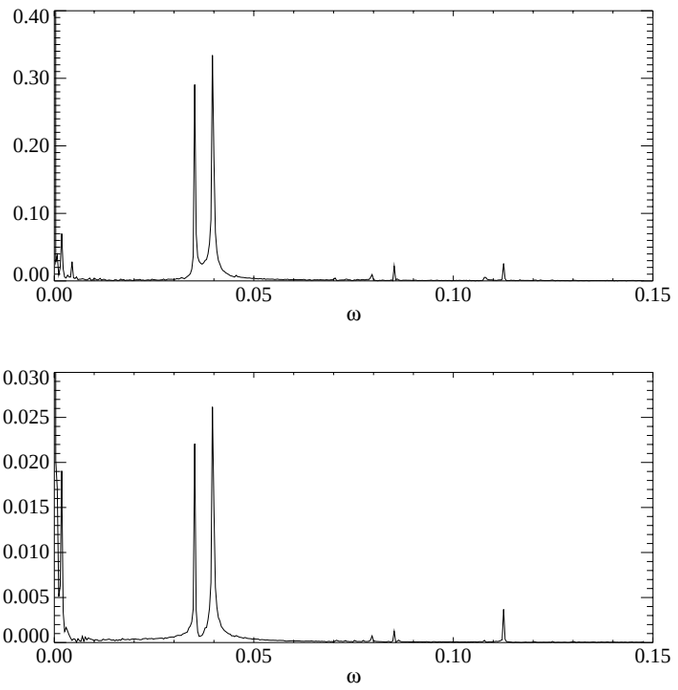

FIGURE 3. Upper panel: Fourier spectrum of the radial displacements $r(t) = R(t) - R_0$ of the vortex in Fig. 2 from the mean trajectory; the integration time is 20000 in units $(JS)^{-1}$. Lower panel: Spectrum of the azimuthal displacements $\varphi(t) = \phi(t) - \omega_0 t$ from the mean trajectory, where ω_0 is the angular velocity.

system, as considered in this review, this ansatz reads

$$\mathbf{S}(\mathbf{r}, t) = \mathbf{S}(\mathbf{r} - \mathbf{X}, \dot{\mathbf{X}}, \ddot{\mathbf{X}}, \ldots, \mathbf{X}^{(n)}) \,. \tag{7.29}$$

This generalization of the standard travelling wave ansatz (7.8) yields an $(n+1)^{\text{th}}$-order equation of motion for $\mathbf{X}(t)$, because Eq. (7.10) is replaced by

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{\partial \mathbf{S}}{\partial X_j} \dot{X}_j + \frac{\partial \mathbf{S}}{\partial \dot{X}_j} \ddot{X}_j + \cdots + \frac{\partial \mathbf{S}}{\partial X_j^{(n)}} X_j^{(n+1)} \,. \tag{7.30}$$

The same procedure as described below Eq. (7.10) then yields the $(n+1)^{\text{th}}$-order equation.

The case $n = 1$ corresponds to the second-order equation (7.21). In one dimension the antisymmetric tensor $\mathbf{G}$ from Eq. (7.13) vanishes and (7.21) reduces to a Newtonian equation. The same is true for the 2D planar

vortices where $G_{ij} = 0$ because $p = 0$ in Eq. (7.19).
The case $n = 2$ yields the third-order equation

$$\mathbf{A}\dddot{\mathbf{X}} + \mathbf{M}\ddot{\mathbf{X}} + \mathbf{G}\dot{\mathbf{X}} = \mathbf{F}(\mathbf{X}) \tag{7.31}$$

with the third-order gyrotensor

$$A_{ij} = \int \mathrm{d}^2 r\, \mathbf{S} \frac{\partial \mathbf{S}}{\partial X_i} \times \frac{\partial \mathbf{S}}{\partial \ddot{X}_j} = \int \mathrm{d}^2 r \left\{ \frac{\partial \phi}{\partial X_i} \frac{\partial \psi}{\partial \ddot{X}_j} - \frac{\partial \phi}{\partial \ddot{X}_j} \frac{\partial \psi}{\partial X_i} \right\} . \tag{7.32}$$

For the above 1D models and the 2D planar vortices nothing changes because A_{ij} turns out to be zero (below). But the 2D non-planar vortices are the first gyrotropic example.

For the calculation of the integral (7.32) the acceleration dependence of the outer region of the vortex is needed [24]

$$\psi_2 = \frac{p}{4\delta} \left(x_1 \ddot{X}_1 + x_2 \ddot{X}_2 \right) \tag{7.33}$$

$$\phi_2 = \frac{q}{8\delta} \left(x_2 \ddot{X}_1 - x_1 \ddot{X}_2 \right) \ln \frac{r}{\mathrm{e}L} . \tag{7.34}$$

Together with the static parts ϕ_0 and ψ_0 this yields

$$A_{ij} = A \epsilon_{ij} , \quad A = \frac{G}{16\delta} \left(L^2 - a_c^2 \right) + C_A , \tag{7.35}$$

where the constant C_A stems from the inner region.

In the simulations the dynamic in-plane structures ϕ_1 and ϕ_2 cannot be clearly observed because it is difficult to subtract the static structure ϕ_0 which varies drastically with the vortex position. However, the dynamic out-of-plane structures, ψ_1 and ψ_2, can be observed and distinguished by looking at specific points of the trajectory: e. g., at the turning points in Fig. 4 the acceleration is maximal while the velocity is small. Figs. 5, 6 confirm the structure of ψ_2 in the outer region, they also show that this structure oscillates as a whole, with frequency $(\omega_1 + \omega_2)/2$. An alternative interpretation of this oscillating structure in terms of certain "quasilocal" spinwaves will be given in section 2.4.

As ψ_2 and ϕ_2 depend linearly on $\ddot{X}_j$ and as A_{ij} contains derivatives with respect to $\ddot{X}_j$, the leading contribution to A_{ij} is independent of velocity and acceleration. There are higher-order terms which are negligible, though; cf. the discussion of the mass. For these reasons the l. h. s. of the equation of motion (7.31) is linear.

The radial Coulomb force $\mathbf{F}$ on the r. h. s. can be linerarized by expanding in the small radial displacement $r(t) = R(t) - R_0$ from the mean trajectory which is a circle of radius R_0

$$F(R) = F(R_0) + F'(R_0) r . \tag{7.36}$$

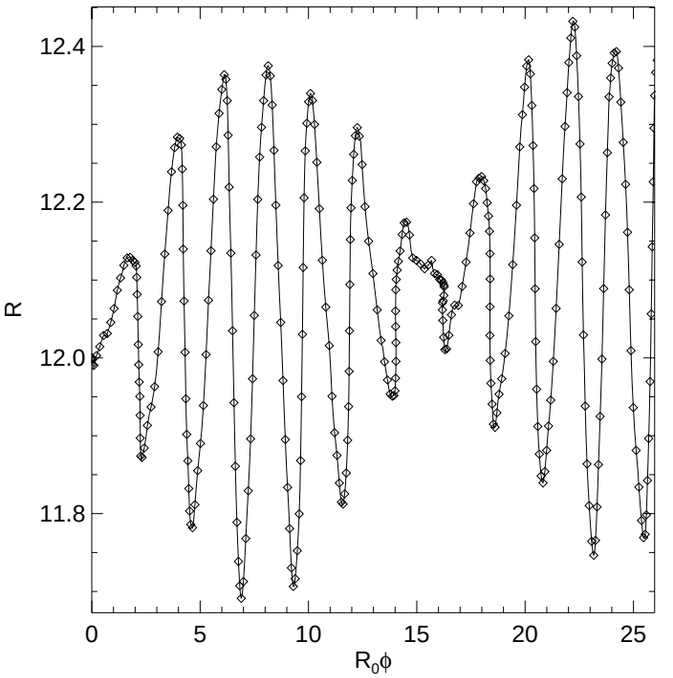

FIGURE 4. First part of the trajectory of a vortex with $q = p = 1$ on a circular system with radius $L = 36$ and free boundary conditions. The small diamonds ($\diamond$) mark the position of the vortex in time intervals of $10(JS)^{-1}$.

The third-order equation of motion (7.31) then has the following solutions [24]: A stationary solution and a superposition of two cycloids

$$\begin{aligned} r(t) &= a_1 \cos\omega_1 t + a_2 \cos\omega_2 t \\ R_0\varphi(t) &= b_1 \sin\omega_1 t + b_2 \sin\omega_2 t \end{aligned} \tag{7.37}$$

where $\varphi = \phi - \omega_o t$ is the azimuthal displacement. The results for $\omega_{1,2}$ can be considerably simplified for $R_0 \ll L$, which is the case in the simulations. The frequencies $\omega_{1,2}$ form a weakly-split doublet. The mean frequency depends on A, but not on M:

$$\bar{\omega} = \sqrt{\omega_1\omega_2} = \sqrt{G/A} \sim 1/L \tag{7.38}$$

for large L. Contrary to this, the splitting of the doublet

$$\Delta\omega = \omega_2 - \omega_1 = M/A \sim \frac{\ln L}{L^2} \tag{7.39}$$

is proportional to the mass.

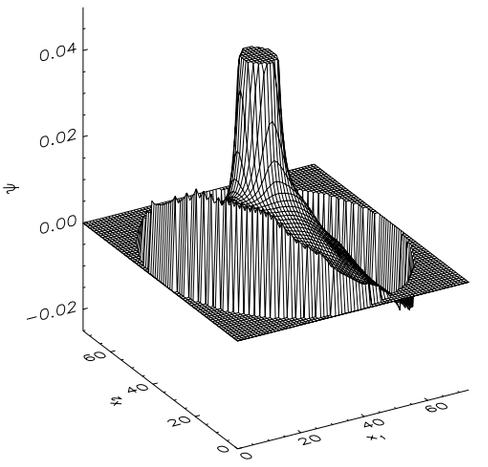

FIGURE 5. Out-of-plane structure of the vortex at the 7^{th} turning point of the trajectory in Fig. 4. Here the acceleration has a maximum and points in the radial direction, while the velocity is small and points in the azimuthal direction.

As $G = 2\pi qp$ is known from Eq. (7.19), the last two equations are sufficient to determine M and A by using the frequencies $\omega_{1,2}$ observed in the simulations. In Ref. [24] the data for lattice sizes $L = 24 \ldots 72$ were extrapolated for $R_0 \to 0$ by using several runs for small R_0/L, because the theoretical predictions (7.25) and (7.35) were made for $R_0 = 0$. For the anisotropy $\delta = 0.1$ the data for A are well represented by $A = CL^\alpha + A_0$ with $\alpha = 2.002,\ C = 4.67$ and $A_0 = 40$. This agrees rather well with the L^2-term in (7.35); the constant C_A cannot be tested. However, $M \approx 15$ is nearly independent of L, in contrast to the logarithmic dependence in Eq. (7.25). This can be connected to the observation that the velocity dependent part ψ_1 seems to approach an L-independent constant at the boundary, in contrast to the predicted $1/r$-decay in Eq. (7.23).

A very interesting point is the discussion of the size dependence of the different terms in the equation of motion (7.31). As every time derivative of X_i contributes a factor of $\omega_{1,2} = \mathcal{O}(1/L)$, the orders of the terms are $A\dddot{X}_i = \mathcal{O}(1/L), M\ddot{X}_i = \mathcal{O}(\ln L/L^2)$ and $G\dot{X}_i = \mathcal{O}(1/L)$. Therefore the strong third-order term cannot be neglected when the weak second-order term is retained. This is the reason for the two severe discrepancies between the simulations and the predictions of the second-order equation (7.21), which were discussed in section 2.2. However, the neglection of both the second- and third-order terms represents a consistent first approximation,

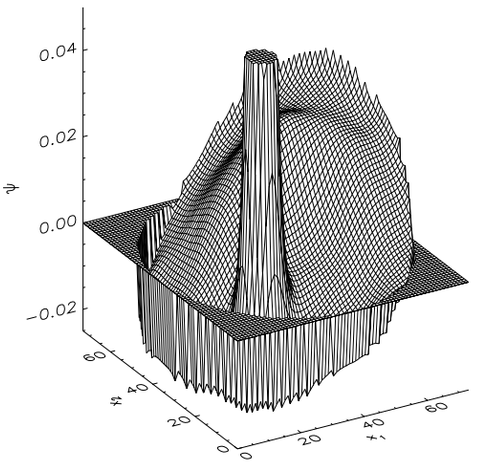

FIGURE 6. Same as in Fig. 5, but at the 8^{th} turning point, where the accelaration points in the negative radial direction.

namely, the Thiele equation (7.11).

The next consistent approximation in the hierarchy is the third-order equation (7.31), as discussed above. An investigation of even higher-order terms in Ref. [24] confirms the conjecture that only the odd-order equations of the hierarchy represent valid approximations for gyrotropic excitations (For non-gyrotropic ones all members of the hierarchy are even).

The odd higher-order equations predict additional frequency doublets $\omega_{3,4}$, $\omega_{5,6}$ etc. These frequencies normally vanish in the background of the spectrum. However, they become visible if the simulations are specially designed: For a vortex in the center of a quadratic system with antiperiodic boundary conditions, all image forces cancel exactly and only the small pinning forces remain. For this configuration two additional doublets with strongly decaying amplitudes were indeed observed [24].

Finally we mention that cycloidal vortex trajectories have been found not only in field-theoretic models for magnets [26, 30], but also for (1) neutral and (2) charged superfluids:
(1) In the Ginzburg-Landau theory, which describes vortex motion in thin, neutral, superfluid films, the usual assumption of incompressibility was dropped [31]. A moving vortex then exhibits a density profile in the region outside the vortex core. This profile is similar to the velocity dependent out-of-plane structure (7.23) of the magnetic vortices, and the consequences are also similar: There are small-amplitude cycloids [32] superimposed on the

trajectories in both 2-vortex scenarios (vortex-vortex rotation and vortex-antivortex translation, see end of section 2.1).
(2) For the dynamics of vortices in a charged superfluid the same kind of superimposed cycloids was found, again for both 2-vortex scenarios [33]. Here a field-theoretic model was used, where a charged scalar field is minimally coupled to an electromagnetic field and a ϕ^4-potential is included; this is proposed as a phenomenological model for a superconductor.

However, compared to the dynamics of magnetic vortices as reviewed in this article, there is an important difference: For all the models discussed above, there is only *one* cycloidal frequency; correspondingly the vortex dynamics is governed by second-order equations of motion.

It will be interesting to study many other physical contexts in which topological vortex structures appear (e. g., dislocations in solids and flux lattices, vortex filaments in compressible fluids and complex Ginzburg-Landau models). We expect that the details of the vortex dynamics differ, depending on the order of the equations of motion, for instance (see beginning of this section).

2.4 An alternative approach: coupling to magnons

The spirit of the *generalized travelling wave ansatz* (7.29) differs considerably from the well-known *standard ansatz*, which reads for a spin system

$$\mathbf{S}(\mathbf{r},t) = \mathbf{S}_0(\mathbf{r} - \mathbf{X}(t)) + \boldsymbol{\chi}(\mathbf{r},t)\,. \tag{7.40}$$

Here $\mathbf{S}_0$ represents the *static* structure of a single nonlinear coherent excitation (a vortex in our case) and $\boldsymbol{\chi}$ represents a magnon field (or meson field for other kinds of systems). Since $\mathbf{S}_0$ is static, but the shape of the excitation usually depends on the velocity, a part of the dynamics obviously must be taken over by the spin waves.

Based on earlier work, this concept has very recently been carried through [40]. First, the magnon modes in the presence of a single static vortex were obtained by a numerical diagonalization for relatively small, discrete systems with fixed (Dirichlet) boundary conditions (BC) [34, 35, 36]. Analytical investigations were done for planar vortices in antiferromagnets [37, 39] and ferromagnets [38, 39]. These articles demonstrated nontrivial properties of the eigenmodes, e. g., the presence of quasi-local (resonance type) [34, 35] or truly local [36] modes. Moreover, the relevance of these modes for the vortex dynamics was shown, in particular the transition between planar and non-planar vortices was investigated [34, 35].

However, for non-planar vortices all this was based on numerical diagonalization. But very recently, a general theory of vortex-magnon coupling was developed for arbitrarily large systems with circular symmetry and general BC [40]. The S-matrix for vortex-magnon scattering was calculated and expressed by Bessel and Neumann functions. Using the S-matrix, general formulas for the eigenfrequencies were obtained, as a function of the

parameters and size of the system, and for different BC, namely for Dirichlet (fixed) and Neumann (free) BC.

There is a very good agreement between the frequencies of the two lowest quasi-local modes and the frequencies $\omega_{1,2}$, which were observed for the vortex trajectories in spin dynamics simulations for the discrete system (section (2.2). The error is only 0.8 % for the mean frequency $\bar{\omega} = \sqrt{\omega_1 \omega_2}$ and 4 % for the splitting $\Delta\omega = \omega_2 - \omega_1$. This demonstrates very clearly that the *cycloidal vortex motion* is accompanied by certain magnon modes, namely by *quasi-local modes*. These modes are both extended (like the other magnon modes) and local, because they exhibit a localized structure around the vortex center (Fig. 7). The structure of these modes is qualitatively similar to the dynamic vortex structure, which oscillates with the rythm of the cycloidal motion (Figs. 5, 6). However, only the out-of-plane part ψ of this structure can be clearly observed, and only in the outer region (see section 2.3).

As the frequencies $\omega_{1,2}$ in the cycloidal vortex motion are the same as the frequencies of the lowest quasi-local modes, and as the latter ones were calculated analytically, the parameters A and M in the third-order equation of motion (7.31) can also be calculated analytically [40]. The results can be expressed by the first root x_1 of the equation $aJ_1(x) + bxJ_1'(x)r_v/L = 0$, where J_1 is a Bessel function. The values for a and b define the boundary conditions: $a = 0, b = 1$ for free BC and vice versa for fixed BC.

The general result

$$A = \frac{\pi}{2\delta x_1^2} L^2 \tag{7.41}$$

in the limit of large L specializes to $A = 4.634L^2$ for $\delta = 0.1$ and free BC, which agrees perfectly with the result $4.67L^{2.002}$ from the simulations (section 2.3). For fixed BC $A = 1.07L^2$ is about four times smaller. The general result for the vortex mass M is much more complicated than Eq. (7.41) and can be found in Ref. [40]. The numerical value $M = 14.74$ for $\delta = 0.1$ and free BC agrees well with the value 15 from the simulations (section 2.3). However, both results are independent of the system size L, in contrast to the logarithmic L-dependence (7.25) predicted by the integral (7.22) in the collective variable theory, cf. the discussion in section 2.3. For fixed BC, $M = 7.661$ is obtained, about half of the above value for free BC.

The fact that both the mass M and the factor A in the third-order gyrocoupling term depend strongly on the boundary conditions appears very natural because the vortices are not localized excitations like solitons in 1D; the in-plane vortex structure falls off as $1/r$.

It is important to note that the higher-lying quasi-local modes could be calculated by the same methods and are expected to agree with the higher-order doublets $\omega_{3,4}, \omega_{5,6}, \ldots$ appearing in the spectrum of the trajectories. Each additional doublet corresponds to taking into account two additional orders in the generalized travelling wave ansatz (7.29), as discussed in section 2.3.

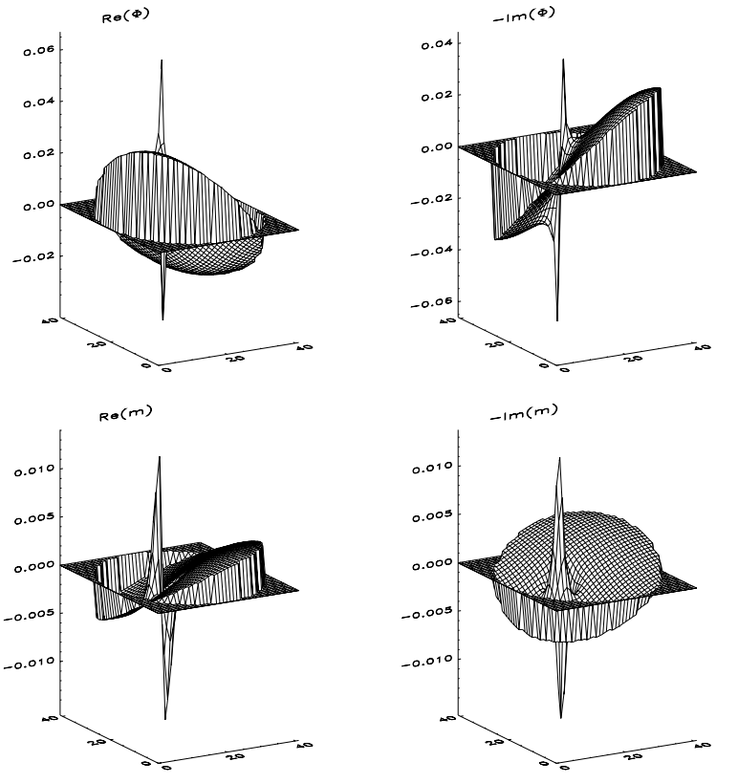

FIGURE 7. One of the two low-lying quasi-local spin wave modes in a complex representation, obtained by a numerical diagonalization for a circular system with a vortex in its center, from [29] (In this reference m corresponds to our ψ).

Finally, we mention the recent development of a very general collective variable theory for an arbitrary Hamiltonian $H[\phi, \psi]$ supporting nonlinear coherent excitations, without making any approximations [29, 41]. This theory starts with the standard ansatz for $\phi(\mathbf{r}, t)$ and $\psi(\mathbf{r}, t)$, cf. Eq. (7.40), but imposes constraints between the meson fields and the functions $\phi_0(\mathbf{r}; X_1, \ldots, X_m)$ and $\psi_0(\mathbf{r}; X_1, \ldots, X_m)$, which describe the shape of the excitation. Here the $X_i(t)$ are m collective variables for the position and the internal modes of the excitation (if there are any). There are then $2m$ constraints in order to preserve the correct number of degrees of freedom. Consequently, the mathematical formalism is based on the classical limit of Dirac's quantum mechanics for constraint systems. The equations of motion for the collective variables are a generalization of Thiele's equation (7.11). The *rank* of the gyrocoupling matrix $\mathbf{G}$ is used to classify the excitations: In the case of vanishing $\mathbf{G}$ the excitations have an effective

mass and exhibit Newtonian dynamics; in the case of regular **G**, the excitations behave like charged, massless particles in an external magnetic field, similar to the gyrotropic excitations defined in section 2.3. The above theory is a generalization of earlier work of Tomboulis [42], Boesch et al. [43], and Willis et al. [44], which applies only to "standard" Hamiltonians (i. e., consisting of the sum of kinetic and potential energy terms). This generalization is necessary, e. g., in the case of magnetic systems which cannot be modelled by standard Hamiltonians.

3 Effects of thermal noise on vortex dynamics

3.1 Equilibrium and non-equilibrium situations

As already mentioned in the introduction, below T_{KT} vortex-antivortex pairs appear and vanish spontaneously due to thermal fluctuations. But these pairs do not move and therefore give no contribution to the dynamic correlation functions. Above T_{KT}, some of the pairs unbind and the free vortices can move. In section 4, this situation will be investigated by a phenomenological theory, namely the vortex gas approach.

On the other hand, the effect of thermal fluctuations on *single* vortices can be studied by putting a vortex in a thermal bath. This is a non-equilibrium situation, in fact it takes a long time until equilibrium is reached: The vortex very slowly approaches the boundary where it annihilates together with an image antivortex; during this process spin waves are radiated which are eventually thermalized.

The vortex motion with thermal noise is a random walk process, where a *vortex diffusion constant* can be defined. This offers the possibility to develop an ab-initio theory for the dynamic form factors. This can be compared with the phenomenological vortex gas approach, which assumes, however, a ballistic motion (section 4).

3.2 Collective variable theory and Langevin dynamics simulations

In principle, the generalization of a collective variable theory to finite temperatures is a straightforward procedure consisting of 4 steps:
(1) Thermal noise and damping (because of the fluctuation-dissipation theorem) are introduced into the microscopic equations.
(2) A travelling wave ansatz for a nonlinear coherent excitation is made which yields equations of motion with stochastic forces acting on the excitation.
(3) The solutions of these stochastic ode's are used to calculate the variances of the trajectory, which contain as a factor an effective diffusion

constant for this excitation. The dependences of this constant on the temperature and other parameters can be discussed.
(4) The predicted variances and the effective diffusion constant are compared with the same quantities as observed in Langevin dynamics simulations.

Step 1: Introduction of thermal noise

This step is problematic, although there are many papers in which the problems are either not appreciated or hidden. There are basically two major problems:
(a) The microscopic equations can often be written in different ways which are equivalent. In the case of spin systems, the Landau-Lifshitz equation for the spin $\mathbf{S}$ is equivalent to the Hamilton equations for ϕ and ψ. However, if noise is implemented, e. g., by an additive term, the results may be *qualitatively different* (see below).
(b) Typically, either *additive* or *multiplicative* noise can be used, but many papers do not give a reason why one of the two types was chosen. However, the results are usually *qualitatively different* for the two types of noise.

In the case of the vortex dynamics, both major problems have been investigated in a preprint [45]: Additive noise in the Hamilton equations yields a vortex diffusion constant D_V which depends *logarithmically* on the system size, while in the multiplicative case D_V is *independent* of the system size. Additive noise in the Landau-Lifshitz equation [46, 47] yields a diffusion constant with the same logarithmic term as in the case of additive noise in the Hamilton equations, but the small constants that appear in addition to the logarithmic term differ for the two cases because the vortex core gives different contributions.

Unfortunately, the additive noise in the Landau-Lifshitz equation has an unphysical feature, namely the spin length is not conserved. In the collective variable theory this problem is overcome by a constraint, while in the simulations a renormalization of the spin length is necessary after every time step[2].

However, taking additive noise and using such tricks are not really necessary, because there is a better noise term that eventually leads to the same results, but which is well motivated on physical grounds and which conserves the spin length [45, 48]: In the Landau-Lifshitz equation (7.7) each spin $\mathbf{S}^m$ precesses in a local magnetic field $\mathbf{B}$ which is the gradient of the energy with respect to the spin components. This local field is the only way through which the spin $\mathbf{S}^m$ can feel any changes in its environment. Adding a thermal noise term to the local field thus accounts for the interac-

[2]Technically, this is achieved by adding a Lagrange parameter multiplying the constraint [46]. This means that a multiplicative noise term appears besides the additive one.

tion of the spin with magnons, phonons and any other thermally generated excitations. The stochastic Landau-Lifshitz equation then reads

$$\frac{\mathrm{d}\mathbf{S}^m}{\mathrm{d}t} = -\mathbf{S}^m \times \left[\frac{\partial H}{\partial \mathbf{S}^m} + \mathbf{h}^m(t)\right] - \epsilon \mathbf{S}^m \times \frac{\mathrm{d}\mathbf{S}^m}{\mathrm{d}t}\,. \tag{7.42}$$

Because of the cross product with $\mathbf{S}^m$ this noise is *multiplicative*. In Ref. [45] Gaussian white noise is used:

$$< h_\alpha^m > \;=\; 0 \tag{7.43}$$

$$< h_\alpha^m(t) h_\beta^n(t') > \;=\; D\delta^{mn}\delta_{\alpha\beta}\delta(t-t')\,, \tag{7.44}$$

where $D = 2\epsilon k_B T$ is the diffusion constant and α, β denote cartesian components. Following Refs. [16, 17, 18], a Gilbert damping with damping parameter ϵ was chosen in Eq. (7.42), chiefly because it is isotropic, in contrast to the Landau-Lifshitz damping [49].

Strictly speaking, the three equations (7.42) do not really represent Langevin equations, because *all* the components of $d\mathbf{S}^m/dt$ appear in each equation due to the cross product. To properly introduce the noise, one first has to group all the time derivatives on the l. h. s. of the equation, and only then one can introduce independent white noise terms for each spin component. But this procedure eventually produces only a renormalization of the damping parameter ϵ in the order of ϵ^2. In the simulations values of ϵ in the order of 10^{-3} were used, and therefore the correction factor can even be neglected.

Another, even more important issue is the interpretation of the stochastic differential equation (7.42). As the noise is *multiplicative*, Ito and Stratonovič interpretations do not coincide, in contrast to the additive noise case [50, 51]. In principle, when thinking of thermal excitations interacting with the spins, there would be a finite correlation time which would lead to a colored noise term. Taking white noise means taking the limit of zero correlation time, and therefore the stochastic Landau-Lifshitz equation (7.42) must be interpreted in the *Stratonovič sense*.

As noted in the beginning of this subsection, the remaining three steps of the stated procedure are straightforward. Therefore we present here only the results for the case of non-planar vortices.

Step 2: Stochastic equation of motion

The generalized travelling wave ansatz (7.29) up to order $n = 2$ yields the third-order stochastic ode

$$\mathbf{A}\dddot{\mathbf{X}} + \mathbf{M}\ddot{\mathbf{X}} + \mathbf{G}\dot{\mathbf{X}} = \mathbf{F}(\mathbf{X}) + \mathbf{F}^{\mathrm{mult}}(t) \tag{7.45}$$

The tensors $\mathbf{A}, \mathbf{M}$ and $\mathbf{G}$ are the same as in Eqs. (7.19), (7.25) and (7.35), except that all the vanishing components in these three expressions are now replaced by damping terms proportional to ϵ. Thus the damping appears

at every order in Eq. (7.45). Without $\mathbf{F}^{\rm mult}$, the solution is obtained in complete analogy to section 2.3: Two cycloids are superimposed on an outward spiral on a circle (Fig. 8). In the simulations the purpose of the damping is to dissipate the energy which is supplied to the system by the kicks of the noise. Therefore the range of ϵ (for a given system size L) must be determined in which the cycloidal frequencies $\omega_{1,2}$ in the trajectories are not influenced by the damping [46]. The result is the condition $\epsilon L \ll 6$.

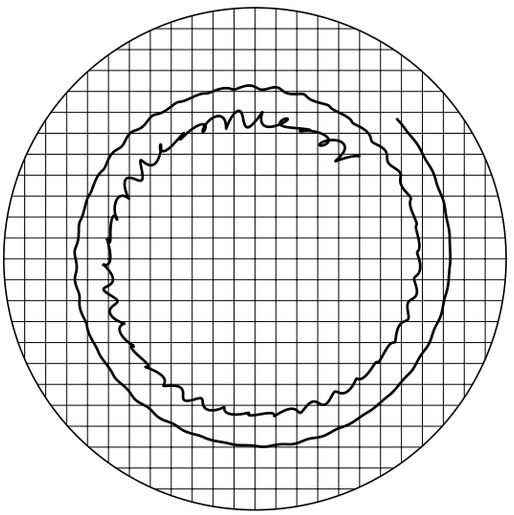

FIGURE 8. Schematic sketch of the vortex motion as governed by the Landau-Lifshitz equation with Gilbert damping. The plot is approximate and does not correspond to an actual simulation.

The stochastic force in (7.45)

$$F_i^{\rm mult}(t) = \frac{1}{S^2}\int {\rm d}^2 r \frac{\partial \mathbf{S}}{\partial X_i}\mathbf{h}(\mathbf{r},t)\,, \tag{7.46}$$

which stems from the multiplicative noise in Eq. (7.42), has zero mean and the correlation function [45]

$$< F_i^{\rm mult}(t) F_j^{\rm mult}(t') >= D\delta_{ij}\delta(t-t')\int {\rm d}^2 r \frac{\partial \mathbf{S}}{\partial X_i}\frac{\partial \mathbf{S}}{\partial X_j}\,. \tag{7.47}$$

Putting the vortex on the center of a circular system of radius L with free BC, the leading contribution to the variance is obtained [47] by using the static vortex structure (7.15) to (7.17)

$$\mathrm{Var}\left(F_i^{\rm st}\right) = D_V \delta(t-t') \tag{7.48}$$

with the vortex diffusion constant

$$D_V = D\pi \left(\ln \frac{L}{a_c} + C \right) . \tag{7.49}$$

Here C stems from the inner vortex region $0 \leq r \leq a_c$ and is obtained by numerical integration [47].

The variance (7.48) is a remarkable result for two reasons:
(a) *The stochastic forces acting on the vortex represent additive white noise* with an effective diffusion constant. The point is that this was shown by going from the microscopic level, where *multiplicative* white noise was implemented, to the level of the collective variables. This approach is much more satisfying than assuming *ad hoc* a noise term on the collective variable level.
(b) The mean and variance of the stochastic force (7.46) turn out to be the *same* as those of the stochastic force

$$F_i^{\text{add}} = \frac{1}{S^2} \int \mathrm{d}^2 r \left(\mathbf{S} \times \frac{\partial \mathbf{S}}{\partial X_i} \right) \boldsymbol{\eta}(\mathbf{r}, t) \, , \tag{7.50}$$

which resulted from starting with additive white noise $\boldsymbol{\eta}(\mathbf{r}, t)$ in the Landau-Lifshitz equation; this noise has already been discussed in step 1. This equivalence of the two forces means that in the case of the vortices additive noise and the multiplicative noise in Eq. (7.42) have the *same* effect on the vortex dynamics. This is a nontrivial result, because normally a qualitative difference is expected (cf. the discussion in step 1). Qualitatively, this result can be understood by noting that the renormalization of the spin length in the case of additive noise introduces effectively a multiplicative noise term, in addition to the additive one (cf. footnote in step 1).

Step 3: Variances of the vortex trajectory

The inhomogeneous stochastic third-order ode (7.45) can be solved [47] by a standard Green function formalism, because the l. h. s. is linear and the deterministic force $\mathbf{F}(\mathbf{X})$ on the r. h. s. can be linearized by expanding in the displacement $\mathbf{x}(t) = \mathbf{X}(t) - \mathbf{X}^0(t)$ from the mean trajectory $\mathbf{X}^0(t)$ on which the vortex is driven by $\mathbf{F}$, cf. Eq. (7.36). Knowing $\mathbf{x}(t)$, the variance matrix

$$\sigma_{ij}^2(t) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle . \tag{7.51}$$

can be calculated; here i, j denote polar coordinates R and ϕ. Each element of the matrix (7.51) turns out to consist of 36 terms. In order to facilitate the discussion several simplifications can be made which give the following results [47]:

$$\sigma_{RR}^2(t) = \frac{D_V}{(2\pi)^2} \left[t + \frac{1}{4\beta} \left(1 - \mathrm{e}^{-2\beta t} \right) - \frac{2}{\bar{\omega}} \mathrm{e}^{-\beta t} \sin \bar{\omega} t + \frac{1}{4\bar{\omega}} \mathrm{e}^{-2\beta t} \sin 2\bar{\omega} t \right] , \tag{7.52}$$

with $\bar{\omega} = \sqrt{\omega_1 \omega_2}$ and $\beta \simeq \epsilon/5$ is the damping constant for the vortex motion. For large times, $t \gg 1/\beta$, only the first term remains and the variance is linear in time: the standard random walk result. Interestingly, this result is identical to the one which is obtained by omitting the second- and third-order terms in the stochastic equation of motion (7.45). This obviously means that these two terms have the following effects:
(a) They produce the oscillatory parts in Eq. (7.52), which are naturally connected to the cycloidal oscillations in the vortex trajectories,
(b) for small times, $t \gg 1/\beta$, the slope of σ^2_{RR}, averaged over the oscillations, is larger by a factor of 3/2 compared to late times.

However, the first effect cannot be observed in the simulations because the oscillations are hidden in strong discreteness effects (see next step). For this reason we discuss only the long-time behavior of the other elements of the variance matrix

$$\sigma^2_{R\phi}(t) = \frac{1}{2} \frac{D_V}{(2\pi)^2} \frac{kF_0'}{2\pi} t^2 \tag{7.53}$$

$$\sigma^2_{\phi\phi}(t) = \frac{D_V}{(2\pi)^2} \left[t + \frac{1}{3} \left(\frac{kF_0'}{2\pi} \right)^2 t^3 \right] \tag{7.54}$$

with $k = 1 - F_0/(F_0' R_0)$ and $F_0 = F(R_0)$, $F_0' = F'(R_0)$. The quadratic and cubic time dependences are *non-standard* results which appear in addition to the standard linear dependence because the deterministic force field $F(R) = F_0 + F_0'(R - R_0)$ is *inhomogeneous*. F is a radial force which drives the vortex in the azimuthal direction, due do the gyrocoupling force (7.3). Therefore, only the ϕ-components of σ^2 are affected: $\sigma^2_{R\phi}$ acquires a factor $kF_0'/(2\pi) \cdot t$, while $\sigma^2_{\phi\phi}$ acquires it twice.

Step 4: Langevin dynamics simulations

The stochastic Landau-Lifshitz equation (7.42) was solved numerically for a large circular lattice with free BC and one vortex driven by a radial image force [45, 46, 47]. The mean trajectory is an outward spiral. Therefore a small damping parameter ϵ was chosen in order to allow long integration times.

There are several qualitatively distinct temperature regimes. For $0 \leq T < T_3 \approx 0.05$ (in dimensionless units, where $T_{\mathrm{KT}} \approx 0.70$ for the XY-model [52, 53]), two frequencies are observed in the oscillations around the mean trajectory which can be identified with the cycloidal frequencies $\omega_{1,2}$ in section 2.3. As these frequencies are constant in the whole regime, the third-order equation of motion (7.45) with *temperature independent* parameters can in fact describe the vortex dynamics.

For $T_3 < T < T_1 \approx 0.3$ the above two frequencies cannot be observed any longer due to large fluctuations , thus the first-order equation of motion is sufficient here. However, in some of the runs the vortex suddenly changed its direction of motion; this will be explained in the next subsection.

Finally, for $T > T_1$, a single vortex theory is no longer adequate because here the probability for the spontaneous appearance of vortex-antivortex pairs in the neighborhood of the single vortex becomes too large.

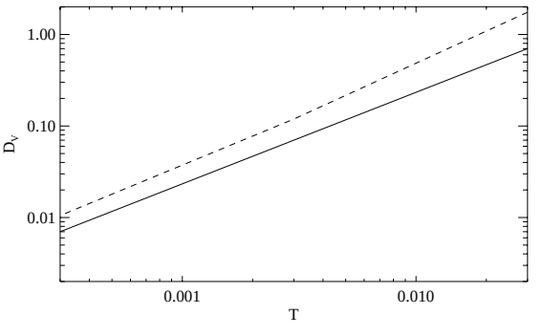

FIGURE 9. Vortex diffusion constant D_V as a function of temperature, for $\epsilon = 0.002$ and $L = 24$. Solid line: Theoretical results from Eq. (7.49); dashed line: Adjusted D_V from fitting the theoretical curves for $\sigma^2(t)$ to the simulation data.

The linear, quadratic, and cubic time dependences in Eqs. (7.52) - (7.54) are well confirmed by the simulations. Therefore the factor D_V can be fitted and turns out to differ from the predicted vortex diffusion constant (7.49) only by a *nearly temperature independent* factor of about 1.8, see Fig. 9 (The constant C in (7.49) was obtained by a numerical integration over the vortex core [46]). This agreement is amazingly good, taking into account that the simulations were performed for a discrete lattice while the theory works in the continuum limit and uses additional approximations, like the expansion (7.36) of the Coulomb force.

Although discreteness effects are hardly visible in the trajectories of non-planar vortices at zero temperature [19], see also Fig. 2, these effects are very important at finite temperatures: Fig. 10 clearly demonstrates that the *variance* of the trajectory shows a pronounced peak whenever the vortex center moves over a ridge of the Peierls-Nabarro potential. Due to this effect, the predicted cycloidal oscillations (7.52) in the variance cannot be observed.

3.3 Noise-induced transitions between opposite polarizations

As already mentioned above, a vortex in a thermal bath can suddenly change its direction of motion on its outward spiral. A closer inspection shows that this occurs because, opposite to the vorticity q, the polarization p is not a constant of motion for a *discrete* system: The out-of-plane vortex

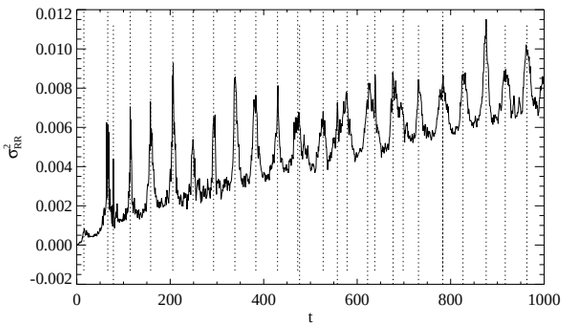

FIGURE 10. Variance of the radial vortex coordinate averaged over 2000 realizations, for $T = 0.003$, $\epsilon = 0.002$, and $L = 24$. The dashed vertical lines indicate the times at which the vortex center moves over ridges of the Peierls-Nabarro periodic potential.

structure can flip to the other side of the lattice plane due to the stochastic forces acting on the core spins. Then the direction of the gyrovector (7.4) is reversed and according to the Thiele equation (7.6) the direction of motion is reversed as well (The same holds for the third-order equation (7.31) because $\mathbf{A} \sim \mathbf{G}$ in Eq. (7.35)).

In a preprint [54] the cores of both planar and non-planar vortices are described by a discrete Hamiltonian, similar to the one that was used for the study of the instability at $\delta = \delta_c$ [34, 11]. Using the stochastic Landau-Lifshitz equation (7.42) the Fokker-Planck equation is derived. Its stationary solution exhibits two maxima for the two possible polarizations of the non-planar vortex and a saddle point for the planar vortex, if the anisotropy parameter δ lies in a certain temperature dependent range. The rate κ for the transition from one polarization to the opposite one is calculated in analogy to Langer's instanton theory [55, 56], using the fact that for $\delta \to \delta_c$ there is a soft mode among the normal modes which were obtained numerically for a system with one vortex [35]. Taking into account only the four innermost spins of the core, a very simple result is obtained

$$\kappa = \frac{1}{2\pi} \sqrt{(1-\delta)^2 - (1-\delta_c)^2}\, \mathrm{e}^{-\frac{\Delta E}{k_B T}} \tag{7.55}$$

where ΔE is the energy difference between the planar and the non-planar vortex. For $\delta = 0.1$, the average transition times $\tau = \kappa^{-1}$ are 100440, 4837, and 1060 for $T = 0.1$, 0.15, and 0.2 respectively. Despite of the crude model for the vortex core, these values agree rather well with the transition times from Langevin dynamics simulations: $\tau = 92516$, 4016, and 825 for the above temperatures, with statistical errors of 22%, 6%, and 10%, respectively.

4 Dynamics above the Kosterlitz-Thouless transition

4.1 Vortex-gas approach

This is a phenomenological theory that is based on the following assumptions: Above T_{KT}, the *free* vortices form a *dilute* gas and move either *diffusively* [57, 58, 18] or *ballistically* [59]. In the former case, only spin *autocorrelation* functions were calculated which lead to dynamic form factors without wave vector dependence. Therefore we review only the latter case [59]:

The density n_v and the r. m. s. velocity $\overline{v}$ are the only free parameters. According to Kosterlitz and Thouless [6]

$$n_v \approx \frac{1}{(2\xi)^2}\,, \tag{7.56}$$

where ξ is the static spin correlation length which diverges with an essential singularity for $T \to T_{\mathrm{KT}}$ from above. Therefore the vortex gas is in fact dilute if T is close enough to T_{KT}.

The vortex density is homogeneous only on the average, locally the distribution is expected to be *random*. Therefore the net force is zero on the average and the distribution around zero is Gaussian, which yields a *Maxwellian* velocity distribution. This also holds for the non-planar vortices because the velocity is proportional to the cross product of the net force and the gyrovector, due to the Thiele equation (7.6).

However, the assumption of a ballistic vortex motion is problematic (section 4.3). Under this assumption the dynamic spin-spin correlation functions

$$S_{\alpha\alpha}(\mathbf{r},t) = \langle S_\alpha(\mathbf{r},t) S_\alpha(0,0)\rangle\,, \alpha = x,y,z\,, \tag{7.57}$$

and their space-time Fourier transforms, namely the dynamic form factors $S_{\alpha\alpha}(\mathbf{q},\omega)$, can be calculated analytically [59].

There is an important difference between in-plane correlations ($\alpha = x, y$) and out-of-plane ones ($\alpha = z$): As the in-plane vortex structure is *not* localized (it has no spatial Fourier transform), $S_{xx} = S_{yy}$ are only *globally sensitive* to the presence of the vortices and the characteristic length scale is their average distance 2ξ. The dynamic form factor exhibits a "*central peak*", i. e. a peak around $\omega = 0$, with the (squared) *Lorentzian* form

$$S_{xx}(\mathbf{q},\omega) = \frac{1}{2\pi^2}\,\frac{\gamma^3\xi^2}{\omega^2 + \gamma^2\left[1 + (\xi q)^2\right]^2}\,, \tag{7.58}$$

where $\gamma = \sqrt{\pi}\overline{v}/(2\xi)$. Here the core structure did not enter the result, because the theory was worked out on a length scale much larger than the core radius r_v. The integrated intensity $I_x(q)$ of (7.58) is *inversely*

proportional to the density n_v, therefore the motion of the vortices actually *destroys* correlations.

For the out-of-plane correlations S_{zz} the situation is completely different: The non-planar vortices have statically the localized S_z structure (7.16), (7.17) which has a spatial Fourier transform, namely the static form factor $f(q)$. Therefore S_{zz} is *locally sensitive* to the vortices, i. e., to their size and shape. Consequently

$$S_{zz}(\mathbf{q},\omega) = \frac{n_v}{4\pi^{5/2}\overline{v}} \frac{|f(q)|^2}{q} \exp\left[-\frac{\omega^2}{(\overline{v}q)^2}\right] \tag{7.59}$$

contains $f(q)$ and is proportional to the density. This *Gaussian* central peak simply reflects the Maxwellian velocity distribution.

For the planar vortices the situation is more complicated because they do not have a static S_z structure but only a dynamic one, namely Eqs. (7.23), (7.24). Therefore the form factor is velocity dependent which yields a more complicated result for S_{zz} [10], containing the same Gausssian as in Eq. (7.59). However, the intensity of this peak is much smaller than that of the peak (7.59), because the dynamic S_z components are about two orders of magnitude smaller than the static ones [24].

4.2 Comparison with simulations and experiments

In combined Monte Carlo (MC) and Spin Dynamics (SD) simulations many features of the predicted dynamic form factors (7.58) and (7.59) were confirmed:
(1) *In-plane correlations*: Both for the XY-model ($\delta = 1$) [59] and for the weakly anisotropic case ($\delta < \delta_c$) [60], the observed $S_{xx}(\mathbf{q},\omega)$ exhibits a central peak for $T > T_{\mathrm{KT}}$ (but not for $T < T_{\mathrm{KT}}$, as expected). The statistical errors were too large to decide about the shape. However, the width $\Gamma_x(q)$ and the integrated intensity $I_x(q)$ of the Lorentzian (7.58) can be fitted to the data, in this way the free parameters ξ and $\overline{v}$ are determined. $\xi(T)$ agrees rather well with the static correlation length [6]. $\overline{v}(T)$ increases with T and then saturates; except of a factor of about 2, this agrees with a result of Huber [58, 18] who assumed a diffusive vortex motion and calculated the velocity autocorrelation function.

Central peaks were also observed in inelastic neutron scattering experiments: For $BaCo_2(AsO_4)_2$ [61] and Rb_2CrCl_4 [62] the measurement was performed for only a few q-values. The reported widths of the peaks differ from the predicted $\Gamma_x(q)$ by factors of about 7 and 2.5, respectively. However, in this comparison one has to take into account that the theory has neglected many features of the real quasi-2D materials: e. g., the lattice structure, a pronounced in-plane anisotropy in the case of $BaCo_2(AsO_4)_2$, and quantum effects.

For $CoCl_2$ graphite intercalation compounds the q-dependence of the

central peak width was measured [2] and agrees qualitatively with $\Gamma_x(q)$ from the peak (7.58).

(2) *Out-of-plane correlations*: In MC-SD simulations, $S_{zz}(\mathbf{q},\omega)$ exhibits a central peak only together with a spin wave peak which sits on its shoulder (in contrast to S_{xx} where the spin waves are strongly softened for small q).

For the XY-model ($\delta = 1$), which bears planar vortices, the intensity of the central peak is expected to be small (see end of section 4.1). Early simulations with large statistical errors [10] had difficulties to substract the spin wave peak and reported only upper bounds for the width $\Gamma_z(q)$ and the intensity $I_z(q)$ of the central peak. Recent simulations with much better statistics investigated the range $\delta_c < \delta \leq 1$. Some papers did not find a central peak [63, 53] others did [64].

However, for the weakly anisotropic case ($\delta < \delta_c$) with non-planar vortices a central peak was indeed found [63]. Experimentally, $S_{zz}(\mathbf{q},\omega)$ was first not accessible because of intensity problems. Only the use of spin-polarized neutron beams made it possible to observe a central peak in Rb_2CrCl_4 and to distinguish it from the spin wave peak and from other contributions [65]. The measured width is practically equal to the width of the Gaussian (7.59). However, such a good agreement seems to be accidental because this material exhibits a breaking of the rotational symmetry in the XY-plane which is described by a more complicated Hamiltonian [66]. For the same reason it is not clear whether this material actually belongs to the case $\delta < \delta_c$.

4.3 Vortex motion in Monte Carlo simulations

The vortex gas approach assumes for single free vortices a diffusive [57, 58] or ballistic [59] motion. Very recently this was tested in MC-simulations by monitoring the position of each vortex in the system (free or bound in a pair) as a function of time [67, 68, 69]. The surprising result is that a single vortex very seldom moves freely over a larger distance. Normally the vortex travels only one or a few lattice spacings until it annihilates with the antivortex of a pair which meanwhile appeared spontaneously in the neighborhood. Another possibility is that the single vortex docks on a pair or a cluster of pairs and after a while one of the vortices leaves the cluster. These results confirm suggestions that vortices cannot move freely for more than a few lattice spacings, which were made by computing the vortex density-density correlation function [70, 67].

However, the interpretation of these results is not clear at all. One possibility is that the vortex gas approach is not valid [67, 68], but then the striking qualitative agreement with the central peaks in the simulations and experiments is not understood. We favour another possibility: Probably only the *effective* vortex motion is important for the dynamic correlation function $S_{xx}(\mathbf{r},t)$ (7.57). This would mean that it does not matter if a vortex is annihilated with the antivortex of a pair in the considered time

interval $[0, t]$, because the vortex of that pair continues the travel instead of the original vortex. Only effective lifetimes are seriously affected. This picture is strongly supported by the mechanism [59] which yields the central peak $S_{xx}(\mathbf{q}, \omega)$ in Eq. (7.58): Looking at the orientation of one particular spin in the XY-plane, one sees that its direction is *reversed* after the vortex center has gone over this spin or its neighborhood. Thus S_x and S_y have *changed their signs* after the vortex is gone, and therefore

$$S_{xx}(\mathbf{r}, t) \sim \langle (-1)^{N(\mathbf{r},t)} \rangle \,. \tag{7.60}$$

Here $N(\mathbf{r}, t)$ is the number of vortices which pass an arbitrary nonintersecting contour connecting $(\mathbf{0}, 0)$ and $(\mathbf{r}, t)$. Obviously, it does not matter if those vortices are replaced by other vortices during their travel (except maybe if the replacement happens just when the contour is passed). The evaluation of Eq. (7.60) leads to the central peak (7.58) [59].

5 Conclusion

The vortex dynamics at zero temperature can be understood by a collective variable theory: a generalized travelling wave ansatz that allows for deformations of the vortex shape due to velocity, acceleration etc., leading to equations of motion. In the case of non-planar vortices, the trajectories exhibit a superposition of cycloids, which is fully confirmed by spin dynamics simulations.

This collective variable theory can be generalized to finite temperatures, yielding stochastic equations of motion. The vortex motion is diffusive and agrees well with Langevin Dynamics simulations. Moreover, the rate of noise-induced transitions between vortex states with opposite polarization is calculated and agrees with the simulations.

Above the Kosterlitz-Thouless transition temperature, however, there is so far no theory that can fully explain the central peaks that were observed both in inelastic neutron scattering experiments and in combined Monte Carlo and spin dynamics simulations. A qualitative agreement is achieved by a phenomenological vortex gas theory, but one of its assumptions, namely ballistic vortex motion, is questionable. Probably, both the diffusive character of the vortex motion and annihilation and creation processes must be incorporated into a theory that can fully explain the above facts. It is likely that the situation is similar for all Kosterlitz-Thouless phase transitions, among many others.

Acknowledgments: Many thanks for a very fruitful collaboration go to (in historical order): Gary Wysin (Kansas State University, USA), Elizabeth Gouvêa (Belo Horizonte, Brazil), Armin Völkel (Xerox, Palo Alto,

USA), Hans-Jürgen Schnitzer (Aachen, Germany), Angel Sánchez, Francisco Domínguez-Adame, Esteban Moro (Madrid, Spain), Niels Grønbech-Jensen (Los Alamos, USA), Boris Ivanov, Yuri Gaididei (Kiev, Ukraine), Alex Kovalev (Kharkov, Ukraine) and Till Kamppeter (Bayreuth, Germany). We are grateful to Frank Göhmann (Stony Brook, USA), Dimitre Dimitrov, Grant Lythe, and Roman Sasik (Los Alamos, USA) for critical readings of the manuscript and constructive comments.

6 References

[1] K. Hirakawa, H. Yoshizawa, J. D. Axe, and G. Shirane, Suppl. J. Phys. Soc. Jpn. **52**, 19 (1983); L. P. Regnault, J. P. Boucher, J. Rossat-Mignod, J. Bouillot, R. Pynn, J. Y. Henry, and J. P. Renard, Physica B+C **136B**, 329 (1986); M. T. Hutchings, P. Day, E. Janke, and R. Pynn, J. Magn. Magn. Mater. **54-57**, 673 (1986); S. T. Bramwell, M. T. Hutchings J. Norman, R. Pynn, and P. Day, J. de Phys. **49**, C8-1435 (1988); L. P. Regnault, C. Lartigue, J. F. Legrand, B. Farago, J. Rossat-Mignod, and J. Y. Henry, Physica B **156-7**, 298 (1989).

[2] D. G. Wiesler, H. Zabel, and S. M. Shapiro, Physica B **156-7**, 292 (1989); D. G. Wiesler, H. Zabel, and S. M. Shapiro, Z. Physik B **93**, 277 (1994).

[3] M. Pomerantz, Surf. Sci. **142**, 556 (1984); D. I. Head, B. H. Blott, and D. Melville, J. Phys. C **8**, 1649 (1988).

[4] M. Steiner and A. R. Bishop, in *Solitons*, eds. S. E. Trullinger, V. E. Zakharov, and V. L. Pokrovsky, North Holland, Amsterdam (1986).

[5] V. L. Berezinskii, Sov. Phys.-JETP **34**, 610 (1971).

[6] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); J. M. Kosterlitz, *ibid* **7**, 1046 (1974).

[7] N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966).

[8] S. Hikami and T. Tsuneto, Prog. Theor. Phys. **63**, 387 (1980).

[9] C. Kawabata and A. R. Bishop, Solid State Commun. **60**, 169 (1986).

[10] M. E. Gouvêa, G. M. Wysin, A. R. Bishop, and F. G. Mertens, Phys. Rev. B **39**, 11840 (1989).

[11] G. M. Wysin, Phys. Lett. A **240**, 95 (1998).

[12] A. A. Belavin and A. M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. **22**, 503 (1975) [JETP Lett.. **22**, 245 (1975)].

[13] G. K. Batchelor, *An introduction to fluid dynamics*, Cambridge University Press, Cambridge (1967).

[14] R. J. Donnely, *Quantized Vortices in Helium II*, Cambridge University Press, Cambridge (1991).

[15] B. A. Ivanov and A. K. Kolezhuk, Low Temp. Phys. **21**, 275 (1995).

[16] A. A. Thiele, Phys. Rev. Lett. **30**, 230 (1973).

[17] A. A. Thiele, J. Appl. Phys. **45**, 377 (1974).

[18] D. L. Huber, Phys. Rev. B **26**, 3758 (1982).

[19] A. R. Völkel, F. G. Mertens, A. R. Bishop, and G. M. Wysin, Phys. Rev. B **43**, 5992 (1991).

[20] G. M. Wysin and F. G. Mertens, in *Nonlinear Coherent Structures in Physics and Biology*, Lecture Notes in Physics, Vol. 393, eds. M. Remoissenet and M. Peyrard, Springer, Berlin (1991) p. 3.

[21] G. M. Wysin, F. G. Mertens, A. R. Völkel, and A. R. Bishop in *Nonlinear Coherent Structures in Physics and Biology*, eds. K. H. Spatschek and F. G. Mertens, Plenum, New York (1994) p. 177.

[22] F. G. Mertens, G. M. Wysin, A. R. Völkel, A. R. Bishop, and H. J. Schnitzer, in *Nonlinear coherent structures in physics and biology,* ed. by K. H. Spatschek and F. G. Mertens (1994) p. 191.

[23] A. R. Völkel, G. M. Wysin, F. G. Mertens, A. R. Bishop, and H. J. Schnitzer, Phys. Rev. B **50**, 12711 (1994).

[24] F. G. Mertens, H. J. Schnitzer, and A. R. Bishop, Phys. Rev. B **56**, 2510 (1997).

[25] J. Tjon and J. Wright, Phys. Rev. B **15**, 3470 (1977).

[26] N. Papanicolaou and T. N. Tomaras, Nucl. Phys. B **360**, 425 (1991).

[27] J. D Jackson, *Classical Electrodynamics*, Wiley, New York (1975).

[28] A. R. Völkel, F. G. Mertens, G. M. Wysin, A. R. Bishop, and H. J. Schnitzer in *Nonlinear Coherent Structures in Physics and Biology*, eds. K. H. Spatschek and F. G. Mertens, Plenum, New York (1994) p. 199.

[29] H.-J. Schnitzer, *Zur Dynamik kollektiver Anregungen in Hamiltonschen Systemen,* Ph.D. thesis, University of Bayreuth (1996).

[30] N. Papanicolaou and W. J. Zakrzewski, Physica D **80**, 225 (1995).

[31] J. M. Duan, Phys. Rev. B **49**, 12381 (1994).

[32] L. K. Myklebust, Cand. Scient. thesis, Univ. of Oslo (1996), preprint (1998).

[33] G. N. Stratopoulos, T. N. Tomaras, Physica D **89**, 136 (1995), Phys. Rev. B **54**, 12493 (1996).

[34] G. M. Wysin, Phys. Rev. B **49**, 8780 (1994).

[35] G. M. Wysin and A. R. Völkel, Phys. Rev. B **52**, 7412 (1995); B **54**, 12921 (1996).

[36] B. A. Ivanov, A. K. Kolezhuk, and G. M. Wysin, Phys. Rev. Lett. **76**, 511 (1996).

[37] B. V. Costa, M. E. Gouvêa, A. S. T. Pires, Phys. Lett. A **156**, 179 (1992).

[38] A. R. Pereira, M. E. Gouvêa, A. S. T. Pires, Phys. Rev. B **54**, 6084 (1996).

[39] G. M. Wysin, M. E. Gouvêa, and A. S. T. Pires, Phys. Rev. B. **57**, 8274 (1998).

[40] B. A. Ivanov, H. J. Schnitzer, G. M. Wysin, and F. G. Mertens, Phys. Rev. B **58**, 8468 (1998).

[41] H. J. Schnitzer, F. G. Mertens, A. R. Bishop, preprint (1998).

[42] E. Tomboulis, Phys. Rev. D **12**, 1678 (1975).

[43] R. Boesch, P. Stancioff, and C. R. Willis, Phys. Rev. B **38**, 6713 (1988).

[44] C. R. Willis and R. Boesch, Phys. Rev. B **41**, 4570 (1990).

[45] T. Kamppeter, F. G. Mertens, E. Moro, A. Sánchez, and A. R. Bishop, Phys. Rev. B **59**, 11349 (1999).

[46] T. Kamppeter, F. G. Mertens, A. Sánchez, N. Grønbech-Jensen, and A. R. Bishop, in *Theory of Spin Lattices and Lattice Gauge Models*, eds. J. W. Clark and M. L. Ristig, Lecture Notes in Physics, Springer (1997) p. 147.

[47] T. Kamppeter, F. G. Mertens, A. Sánchez, A. R. Bishop, F. Domí nguez-Adame, and N. Grønbech-Jensen, Eur. Phys. J. B **7**, 607 (1999).

[48] D. A. Garanin, Phys. Rev. B **55**, 3050 (1997).

[49] S. Iida, J. Phys. Chem. Solids **24**, 625 (1963).

[50] N. van Kampen, *Stochastic processes in physics and chemistry*, North Holland, Amsterdam (1980).

[51] C. W. Gardiner, *Handbook of stochastic methods*, Springer, Berlin-Heidelberg (1990).

[52] A. Cuccoli, V. Tognetti, and R. Vaia, Phys. Rev. B **52**, 10221 (1995).

[53] H. G. Evertz and D. P. Landau, Phys. Rev. B **54**, 12302 (1996).

[54] Y. Gaididei, T. Kamppeter, F. G. Mertens, and A. R. Bishop, Phys. Rev. B **59**, 7010 (1999).

[55] J. S. Langer, Ann. Phys. N. Y., **54**, 258 (1969).

[56] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).

[57] D. L. Huber, Phys. Lett. **68A**, 125 (1978).

[58] D. L. Huber, Phys. Lett. **76A**, 406 (1980).

[59] F. G. Mertens, A. R. Bishop, G. M. Wysin, C. Kawabata, Phys. Rev. Lett. **59**, 117 (1987); Phys. Rev. B **39**, 591 (1989).

[60] A. R. Völkel, A. R. Bishop, F. G. Mertens, and G. M. Wysin, J. Phys. Cond. Matter **4**, 9411 (1992).

[61] L. P. Regnault, J. P. Boucher, J. Rossat-Mignod, J. Bouillot, R. Pynn, J. Y. Henry, J. P. Renard, Physica B+C **136B**, 329 (1986); L. P. Regnault and J. Rossat-Mignod, in *Magnetic Properties of Layered Transition Metal Compounds*, ed. by L. J. de Jongh and R. D. Willet, Kluwer Academic (1990).

[62] M. T. Hutchings, P. Day, E. Janke, and R. Pynn, J. Magn. Magn. Mater., **54-57**, 673, (1986).

[63] J. E. R. Costa and B. V. Costa, Phys. Rev. B **54**, 994 (1996).

[64] M. E. Gouvêa and G. M. Wysin, Phys. Rev. B **56**, 14192 (1997).

[65] S. T. Bramwell, M. T. Hutchings J. Norman, R. Pynn, and P. Day, J. de Phys. **49**, C8-1435 (1988).

[66] M. T. Hutchings, J. Als-Nielson, P. A. Lindgard, and P. J. Walker, J. Phys. C: Solid State Phys. **14**, 5327 (1981).

[67] D. A. Dimitrov and G. M. Wysin, Phys. Rev. B **53**, 8539 (1996).

[68] J. E. R. Costa, B. V. Costa, and D. P. Landau, Phys. Rev. B **57**, 11510 (1998).

[69] D. A. Dimitrov and G. M. Wysin, J. Phys. C **10**, 7453 (1998).

[70] B. V. Costa, D. P. Landau, J. E. R. Costa, and K. Chen in *Computer Simulation Studies in Condensed Matter Physics* VIII, ed. by D. P. Landau, K. K. Mon, and H. B. Schuettler, Springer Verlag, Berlin (1995).

8
Spatial Optical Solitons

Yuri S. Kivshar

ABSTRACT A brief overview of the recent advances in the theoretical and experimental study of self-focusing and self-trapping of light is given. Physical mechanisms of self-trapping and different types of self-trapped beams, *spatial optical solitons*, and their stability are discussed including solitons of non-Kerr media, self-trapped beams and their spiralling in photorefractive crystals, multi-hump solitons and solitonic gluons, discrete solitons in wave guide arrays, etc. A brief summary of the earlier and more recent experimental observations of spatial solitons, transverse instabilities, and soliton interactions is included as well.

1 Introduction

Recent years have shown increased interest in the study of self-guided (or self-trapped) optical beams that propagate in slab wave guides or bulk nonlinear media without supporting wave guide structures. Such optical beams are commonly referred to by physicists as *spatial optical solitons*. Simple physics explains the existence of spatial solitons in a generalized self-focusing nonlinear medium. First, we recall the physics of optical wave guides. Optical beams have an innate tendency to spread (diffract) as they propagate in a homogeneous medium. However, the beam's *diffraction* can be compensated for by beam *refraction* if the refractive index is increased in the region of the beam. An optical wave guide is an important mean to provide a balance between diffraction and refraction if the medium is uniform in the direction of propagation. The corresponding propagation of the light is confined in the transverse direction of the wave guide, and it is described by the so-called linear *guided modes*, spatially localized eigenmodes of the electric field in the wave guide.

As was discovered long time ago [1], the similar effect, i.e. suppression of diffraction through a local change of the refractive index, can be produced solely by nonlinearity. As has been already established in many experiments, some materials can display considerable optical nonlinearities when their properties are modified by the light propagation. In particular, if the nonlinearity leads to a change of the refractive index of the medium in such a way that it generates an effective positive lens to the beam, the beam can

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 173−194, 2000.

become *self-trapped* and propagates unchanged without any external waveguiding structure [1]. Such stationary self-guided beams are called these days *spatial optical solitons*, they exist with profiles of certain form allowing a local compensation of the beam diffraction by the nonlinearity-induced change in the material refractive index.

Until recently, the theory of spatial optical solitons has been based on the nonlinear Schrödinger (NLS) equation with a cubic nonlinearity, which is *exactly integrable* by means of the inverse scattering (IST) technique [2]. Generally speaking, the integrability means that any localized input beam will be decomposed into *stable solitary waves* (or *solitons*) and *radiation*, and also that interaction of solitons is elastic. From the physical point of view, the integrable NLS equation describes (1+1)-dimensional (i.e. one transverse and one longitudinal dimensions) beams in a Kerr nonlinear medium in the framework of the so-called paraxial approximation. The cubic NLS equation is known to be a good model for *temporal optical solitons* propagating large distances along *existing* wave guides, optical fibers. In application to *spatial optical solitons*, the cubic NLS equation is not an adequate model. First, for spatial optical solitons much higher input powers are required to compensate for diffraction, meaning that the refractive index experiences large deviations from a linear (Kerr) dependance. Second, as was recognized long time ago (see, e.g., Ref. [3]), radially symmetric stationary localized solutions of the (2+1)-dimensional NLS equation are unstable and may display collapse [4]. To deal with realistic optical models, saturation had been suggested as a way to stabilize such a catastrophic self-focusing and produce stable solitary waves of higher dimensions. Accounting for this effect immediately leads to *nonintegrable models of generalized nonlinearities*, not possessing the properties of integrability and not allowing elastic soliton collisions. Another mechanism of non-Kerr nonlinearities and enhanced nonlinear properties of optical materials is a resonant, phase-matched interaction between the modes of different frequencies. In this latter case, *multi-component solitary waves* are created, and the mutual beam coupling can modify drastically the properties of single beams, as it occurs in the case of the so-called quadratic solitons of cascaded nonlinearities.

In spite of the fact that, generally speaking, there exist no universal analytical tools for analyzing solitary waves and their interactions in non-integrable models, recent advances of the theory suggest that many of the properties of optical solitons in non-Kerr media are similar, and they can be approached with the help of rather general physical concepts. Also, from this perspective we understand that *there is no simple mapping* between temporal and spatial optical solitons. Spatial solitons are a much richer and more complicated phenomenon, and this has been already demonstrated by a number of elegant experiments in this field.

In particular, it has been recently demonstrated experimentally, that self-guided beams can be observed in materials with strong photorefractive and photovoltaic effect [5], in vapours with a strong saturation of the intensity-

dependent refractive index [6, 7], and also as a result of the mutual self-focusing due to the phase-matched three-wave mixing in quadratic (or $\chi^{(2)}$) nonlinear crystals [8]. In all these cases, propagation of self-guided waves is observed in *non-Kerr materials* which are described by the models more general than the cubic NLS equation.

2 Spatial vs. temporal solitons

Because the phenomenon of the long-distance propagation of temporal optical solitons in optical fibers is known to a much broader community of researchers in optics and nonlinear physics, first we emphasize the difference between spatial and temporal solitons. Indeed, stationary beam propagation in planar wave guides has been considered somewhat similar to the pulse propagation in fibers. This is a direct manifestation of the so-called *spatio-temporal analogy* in wave propagation [9], meaning that the propagation coordinate z is treated as the evolution variable and the spatial beam profile along the transverse direction, for the case of wave guides, is similar to the temporal pulse profile, for the case of fibers. This analogy has been employed for many years, and it is based on a simple notion that both beam evolution and pulse propagation can be described by the cubic NLS equation. However, contrary to this widely accepted opinion, we point out *a crucial difference between these two phenomena*. Indeed, in the case of the nonstationary pulse propagation in fibers, the operation wavelength is usually selected near the zero of the group-velocity dispersion. This means that the absolute value of the fiber dispersion is small enough to be compensated by a weak nonlinearity such as that produced by the (very weak) Kerr effect in optical fibers which leads to a relative nonlinearity-induced change in the refractive index of the order of 10^{-10}. Therefore, nonlinearity in such systems is *always weak* and it should be well modeled by the same form of the cubic NLS equation, which is known to be integrable by means of the IST technique. However, for very short (fs) pulses the cubic NLS equation describing the long-distance propagation of pulses should be corrected to include some additional effects such as higher-order dispersion, Raman scattering, etc. All these corrections can be taken into account by a perturbation theory [11]. Thus, in fibers nonlinear effects are weak and they become important only when dispersion is small (near the zero-dispersion point) affecting the pulse propagation over large distances (of order of hundred of meters or even kilometers).

In contrary to pulse propagation in optical fibers, the physics underlying stationary beam propagation in planar wave guides and bulk media is different. In this case the nonlinear change in the refractive index should compensate for the beam spreading caused by diffraction *which is not a small effect*. That is why to observe spatial solitons, much larger nonlin-

earities are usually required, and very often such nonlinearities are not of the Kerr type (e.g. they saturate at higher intensities). This leads to the models of generalized nonlinearities with the properties of solitary waves different from those described by the integrable cubic NLS equation. For example, unlike the solitons of the cubic NLS equation, solitary waves of generalized nonlinearities may be unstable, they also show some interesting features, such as fusion due to collision, inelastic interactions and spiralling in a bulk, wobbling, amplitude oscillation, etc. Propagation distances usually involved in the phenomenon of beam self-focusing and spatial soliton propagation are of the order of millimeters or centimeters. As a result, the physics of spatial solitary waves is rich, and it should be understood in the framework of the theory of nonintegrable nonlinear models.

3 Basic equations

To describe optical spatial solitons in the framework of the *simplest scalar model of nonresonant nonlinearities*, we consider the propagation of a monochromatic scalar electric field E in a bulk optical medium with an intensity-dependent refractive index, $n = n_0 + n_{\rm nl}(I)$, where n_0 is the linear refractive index, and $n_{\rm nl}(I)$ describes the variation in the index due to the field with the intensity $I = |E|^2$. The function $n_{\rm nl}(I)$ is assumed to be dependent on the light intensity only, and it may be introduced phenomenologically.

Solutions of the governing Maxwell's equation can be presented in the form

$$E(\mathbf{R}_\perp, Z; t) = \mathcal{E}(\mathbf{R}_\perp, Z) e^{i\beta_0 Z - i\omega t} + \text{c.c.}, \tag{8.1}$$

where c.c. denotes complex conjugate, ω is the source frequency, and $\beta_0 = k_0 n_0 = 2\pi n_0/\lambda$ is the plane-wave propagation constant for the uniform background medium, in terms of the source wavelength $\lambda = 2\pi c/\omega$, c being the free-space speed of light. Usually, the spatial solitons are discussed for two geometries. For the beam propagation in a bulk, we assume a (2+1)-dimensional model, so that the Z-axis is parallel to the direction of propagation, and the X- and Y-axis are two transverse directions. For the beam propagation in a planar wave guide, the effective field is found by integrating Maxwell's equations over the transverse structure defined by the wave guide confinement, and therefore the model becomes (1+1) dimensional.

The function $\mathcal{E}(\mathbf{R}_\perp, Z)$ describes the wave envelope which in the absence of nonlinear and diffraction effects $\mathcal{E}$ would be a constant. If we substitute Eq. (8.1) into the two-dimensional, scalar wave equation, we obtain the generalized nonlinear equation,

$$2ik_0 n_0 \frac{\partial \mathcal{E}}{\partial Z} + \left(\frac{\partial^2 \mathcal{E}}{\partial X^2} + \frac{\partial^2 \mathcal{E}}{\partial Y^2} \right) + 2n_0 k_0^2 n_{\rm nl}(I)\mathcal{E} = 0. \tag{8.2}$$

In dimensionless variables, Eq. (8.2) becomes the well-known *generalized NLS equation*, where local nonlinearity is introduced by the function $n_{\rm nl}(I)$.

For the case of the Kerr (or cubic) nonlinearity we have $n_{\rm nl}(I) = n_2 I$, n_2 being the coefficient of the Kerr effect of an optical material. Now, introducing the dimensionless variables, i.e. measuring the field amplitude in the units of $k_0\sqrt{n_0|n_2|}$ and the propagation distance in the units of $k_0 n_0$, we obtain the (2+1)-dimensional NLS equation in the standard form,

$$i\frac{\partial\Psi}{\partial z} + \frac{1}{2}\left(\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2}\right) \pm |\Psi|^2\Psi = 0, \tag{8.3}$$

where the complex Ψ stands for a dimensionless envelope, and the sign $(\pm)$ is defined by the type on nonlinearity, *self-defocusing* ('minus', for $n_2 < 0$) or *self-focusing* ('plus', for $n_2 > 0$).

For propagation in a slab wave guide, the field structure in one of the directions, say Y, is defined by the linear guided mode of the wave guide. Then, the solution of the governing Maxwell's equation has the structure

$$E(\mathbf{R}_\perp, Z; t) = \mathcal{E}(X, Z)A_n(Y)e^{i\beta_n^{(0)}z - i\omega t} + \ldots, \tag{8.4}$$

where the function $A_n(Y)$ describes the corresponding fundamental mode of the slab wave guide. Similarly, substituting this ansatz into Maxwell's equations and averaging over Y, we come again to the renormalized equation of the form (8.3) with the Y-derivative *omitted*, which in the dimensionless form becomes the standard cubic NLS equation

$$i\frac{\partial\Psi}{\partial z} + \frac{1}{2}\frac{\partial^2\Psi}{\partial x^2} \pm |\Psi|^2\Psi = 0. \tag{8.5}$$

Equation (8.5) coincides formally with the equation for the pulse propagation in dispersive nonlinear optical fibers, and it is known to be integrable by means of the inverse scattering transform (IST) technique [2].

For the case of nonlinearities more general then the cubic one, we should consider the generalized NLS equation,

$$i\frac{\partial\Psi}{\partial z} + \frac{\partial^2\Psi}{\partial x^2} + f(|\Psi|^2)\Psi = 0, \tag{8.6}$$

where the function $f(I)$ describes a nonlinearity-induced change of the refractive index, usually $f(I) \propto I$ for small I.

The generalized NLS equation (8.2) [or Eq. (8.6)] has been considered in many papers for analyzing the beam self-focusing and properties of spatial bright and dark solitons. All types of non-Kerr nonlinearities in optics can be divided, generally speaking, into *three main classes*: (i) *competing nonlinearities*, e.g. focusing (defocusing) cubic and defocusing (focusing) quintic nonlinearity, (ii) *saturable nonlinearities*, and (iii) *transiting nonlinearities*. Many references can be found in the recent review paper [12].

Usually, the nonlinear refractive index of an optical material deviates from the linear (Kerr) dependence for larger light intensities. Nonideality of the nonlinear optical response is known for semiconductor (e.g., AlGaAs, CdS, $CdS_{1-x}Se_x$) wave guides and semiconductor-doped glasses. In the case of small intensities, this effect can be modeled by *competing*, cubic-quintic nonlinearities, $n_{\rm nl}(I) = n_2 I + n_3 I^2$. This model describes a competition between self-focusing ($n_2 > 0$), at smaller intensities, and self-defocusing ($n_3 < 0$), at larger intensities. Similar models are usually employed to describe the stabilization of wave collapse in the (2+1)-dimensional NLS equation.

Models with *saturable nonlinearities* are the most typical ones in nonlinear optics. The effective generalized NLS equation with saturable nonlinearity is also the basic model to describe the recently discovered (1+1)-dimensional photovoltaic spatial solitons in photovoltaic-photorefractive materials such as $LiNbO_3$. Unlike the phenomenological models usually used to describe saturation of nonlinearity, for the case of photovoltaic solitons this model can be justified rigorously.

There exist several different models for saturating nonlinearities. In particular, the phenomenological model $n_{\rm nl}(I) = n_\infty[1-(1+I/I_{\rm sat})^{-1}]$ is used more frequently, and it corresponds to the well-known expression derived from the two-level model.

Finally, *bistable solitons* introduced by Kaplan [13] usually require a special type of the intensity-dependent refractive index which changes from one type to another one, e.g. it varies from one kind of the Kerr nonlinearity, for small intensities, to another kind with different value of n_2, for larger intensities. Unfortunately, examples of nonlinear optical materials with such dependencies are not yet known, but the bistable solitons possess attractive properties useful for their possible futuristic applications in all-optical logic and switching devices.

At last, we would like to mention the model of logarithmic nonlinearity, $n^2(I) = n_0^2 + \epsilon \ln(I/I_0)$, that allows close-form exact expressions not only for stationary Gaussian beams (or *Gaussons*, as they were introduced in Ref. [14]), but also for periodic and quasi-periodic regimes of the beam evolution [15]. The main features of this model are the following: (i) the stationary solutions do not depend on the maximum intensity (quasi-linearization) and (ii) radiation from the periodic solitons is absent (the linearized problem has purely discrete spectrum). Such exotic properties persist in any dimension.

4 Stability of solitary waves

Spatial optical solitons are of both fundamental and technological importance if they are stable under propagation. Existence of stationary solutions of the different models of non-Kerr nonlinearities does not guarantee their

stability. Therefore, stability is *a key issue in the theory of self-guided optical beams*.

For temporal solitons in optical fibers, nonlinear effects are usually weak and the model based on the cubic NLS equation and its deformations is valid in most of the cases [10]. Therefore, being described by integrable or nearly integrable models, solitons are always stable, or their dynamics can be affected by (generally small) external perturbations which can be treated in the framework of the soliton perturbation theory [11].

As has been discussed above, much higher powers are usually required for spatial solitons in wave guides or a bulk, so that real optical materials demonstrate essentially non-Kerr change of the nonlinear refractive index with the increase of the light intensity. Generally speaking, the nonlinear refractive index always deviates from Kerr for larger input powers, e.g. it saturates. Therefore, models with a more general intensity-dependent refractive index are employed to analyze spatial solitons and, as we discuss below, solitary waves in such non-Kerr materials can become unstable. Importantly, in many cases the stability criteria for solitary waves can be formulated in a rather general form using the system invariants.

4.1 One-parameter solitary waves

Stability of *bright solitons* of the NLS equation with a generalized nonlinearity has been extensively investigated for many years, and the criterion for the soliton stability has been derived by different methods (see, e.g., Refs. [16, 17, 18]). Stability of bright solitons in the generalized NLS equation of any dimension is given by the simple integral criterion [16]

$$\frac{dP}{d\beta} = \frac{d}{d\beta}\left(\int_V |E(\mathbf{r}, z)|^2 d\mathbf{r}\right) > 0, \tag{8.7}$$

where P is the total beam power and β is the soliton propagation constant. The validity of the Vakhitov-Kolokolov criterion (8.7) can be proven by analysing the eigenvalue problem that appears after linearizing the NLS equation (8.6) near the solitary wave solutions. The condition (8.7) is associated with the existence of one only negative eigenvalue of that problem. If this latter condition is not fulfilled, the stability criterion may be not directly formulated in terms of the beam power P, as we have in the case of nonlinear guided waves. Indeed, it has been already established that stability of self-guided waves in nonlinear wave guide structures can be given, in some cases, by the same criterion (see, e.g., Refs. [19]) but, in general, it is more complicated and depends on a particular mode structure and the type of nonlinearity.

Criterion of the soliton stability (8.7) is usually valid for bright solitons (and nonlinear guided waves) which constitute *a one-parameter family of spatially localized solutions*, i.e. their shape is solely defined by the beam

propagation constant β. The similar criterion is valid even in the case of two-component solitons governed by the only power invariant. For example, it has been shown that the similar criterion applies for two-wave solitons in $\chi^{(2)}$ materials [20].

Linear stability analysis does not allow one to predict the subsequent evolution of unstable solitons. Recently, an asymptotic analytical method for describing the dynamics of unstable solitons (e.g., their diffraction-induced decay, collapse, or switching from unstable to a novel stable state) has been developed [21, 22].

In contrast to bright solitons, the stability criterion for *dark solitons* of the generalized NLS equation has not been understood until recently and, even more, this issue created a lot of misunderstanding in the past. From a historical point of view, the first effort to analyze the stability of dark solitons was stimulated by numerical simulations done by Barashenkov and Kholmurodov [23] who observed instability of the so-called 'bubbles', localized waves of rarefaction of the Bose gas condensate described by the cubic-quintic NLS equation. Later, Bogdan *et al.* [24] explained this phenomenon through the multivalued dependence of the bubble's energy vs. the renormalized momentum.

However, it was believed for a long time that kink-type dark solitons (in particular, black solitons) are always stable. Instability of black solitons was observed for the first time by Kivshar and Krolikowski [25] in numerical simulations of the NLS equation with a saturable nonlinearity.

Stability criterion for dark solitons is defined through the renormalized soliton momentum,

$$\frac{dM_r}{dv} = \frac{d}{dv}\left\{\frac{i}{2}\int_{-\infty}^{+\infty}\left(u\frac{\partial u^*}{\partial x} - u^*\frac{\partial u}{\partial x}\right)\left(1 - \frac{u_0^2}{|u|^2}\right)dx\right\} > 0, \qquad (8.8)$$

and this result was shown to be consistent with numerical simulations and the variational principle. A rigorous proof of this stability criterion was presented only recently by Barashenkov [26], with the help of the Lyapunov function, and Pelinovsky *et al.* [22], by using the asymptotic expansion near the instability threshold. The first approach does not allow one to describe the instability itself but it is more general to prove the global stability if it exists, whereas the second method is valid in the vicinity of the instability threshold being also sufficient to determine the instability domains.

All basic models of optical nonlinearities, i.e. *saturable*, *transiting*, and *competing* nonlinearities display instabilities of dark solitons in some regions of the parameters. The instability of dark solitons is qualitatively different from that of bright solitons, it is *a drift instability* [27].

4.2 Two-parameter solitary waves

The Vakhitov-Kolokolov criterion of the soliton stability can be generalised to the case of solitary waves described by more than one parameter. This

has been demonstrated for the first time for the case of nondegenerate three-wave mixing in a diffractive medium [28]. The instability threshold for two-parameter solitons (marginal stability condition) is given by the following criterion

$$\mathcal{J}(\beta_1, \beta_2) = \frac{\partial(\mathcal{F}_1, \mathcal{F}_2)}{\partial(\beta_1, \beta_2)} \equiv \frac{\partial \mathcal{F}_1}{\partial \beta_1}\frac{\partial \mathcal{F}_2}{\partial \beta_2} - \frac{\partial \mathcal{F}_1}{\partial \beta_2}\frac{\partial \mathcal{F}_2}{\partial \beta_1} = 0, \tag{8.9}$$

where $\mathcal{F}_j$ $(j = 1, 2)$ are two invariants of the system (for the problem of three-wave mixing, these are the total and complimentary powers) describing two-parameter solitary waves, and β_j are two independent parameters of the stationary localized solutions. The stability criterion itself, i.e. the sign of the function $\mathcal{J}$, depends on the model under consideration.

The result (8.9) has been first derived by the asymptotic expansion technique [28] and then verified by the analysis of the global structure of the system invariants [29]. It is a direct consequence of the topology of the invariant surface $H(\mathcal{F}_1, \mathcal{F}_2)$, and it seems to be valid for different types of vectorial and coupled solitons described by two independent parameters introduced by two nontrivial invariants of the model. For example, in the case of *coupled bright-dark solitons*, $\mathcal{F}_1$ is the power of the bright component P and β_1 is the propagation constant of the bright component, whereas $\mathcal{F}_2$ is the total momentum M and β_2 is the soliton velocity V. For incoherently interacting bright solitons, the invariants $\mathcal{F}_j$ are two powers corresponding to two scalar components. The same result holds for the stability of vector solitons in the models with the absence of the Galilean invariance, such as two-wave parametric solitons with the walk-off effect. In this latter case, the second parameter is the soliton velocity V and the second invariant is the soliton momentum M; the same criterion has been recently demonstrated for the so-called *walking solitons* [30].

In general, the marginal stability threshold (8.9) is valid provided the conditions similar to those for the validity of the Vakhitov-Kolokolov criterion (8.7) are satisfied, namely the instability is associated with a kind of *translational* bifurcation of localized eigenmodes of an associated linear eigenvalue problem when the value λ^2, where λ is an eigenvalue, remains real but it changes its sign passing zero. For more complicated models, the invariant criterion may be not valid and other types of instabilities, e.g. *oscillatory instability*, may occur, as has been recently demonstrated for gap solitons in a nonintegrable deformation of the Thirring model [31]. Another example is given by a system where there exists, for the same values of the system parameters, a number of different soliton families corresponding to bifurcations of invariant surfaces. An example of this kind has been recently found for the problem of nondegenerated four-wave mixing where stable multi-color solitons correspond to the lowest invariant surface [32].

5 Experiments on self-focusing

Theoretical predictions of self-focusing of light in an optical medium with nonlinear refractive index [1, 33, 34] were followed by experimental evidence of this phenomenon in different optical materials, e.g. glasses, Raman-active liquids, gas vapors, etc. In particular, Pilipetskii and Rustamov [35] reported the generation of one-, two-, and three filaments due to self-focusing of a laser beam in different organic liquids. Later, Garmire *et al.* [36] reported a direct observation of the evolution of beam trapping in CS_2 in the simplest cylindrical mode. They found that the threshold, trapping length, nonlinearity-induced increase in the refractive index in the trapped region, and beam profile are consistent with theoretical predictions, and the steady-state input beam of circular symmetry asymptotically collapses to a bright filament as small as 50 μm. As a matter of fact, this was one of the first experimental observations of the phenomenon which we now call *spatial optical soliton*.

Because ruby-laser beams used in the experiments have intensities far above threshold for self-trapping in CS_2 (25 ± 5kW), Garmire *et al.* [36] also observed the formation of rings around the self-focused spots and the development of many filaments from an apparently homogeneous beam about 1 mm in diameter and considerably above the threshold power. The former effect can be associated with the existence of a set of higher-order circularly symmetric steady-state modes [37, 38], whereas the latter effect is a direct manifestation of the transverse beam instability, the spatial modulational instability of a broad beam.

Detailed studies of spatial break-up of broad optical beams due to self-focusing was reported almost seven years later by Campillo *et al.* [39, 40] (see also [41]) who used a 50-cm cell of CS_2 to study self-focusing. In particular, Campillo *et al.* [39] observed that radially symmetric ring patterns created by circular apertures break up into focal spots having azimuthal symmetry and regular spacing. This kind of effect can be associated with the transverse modulational instability of quasi-plane bright rings created by the input beam, and the number of the bright spots and critical powers are in a good qualitative agreement with the simple theory of transverse instabilities, as was discussed later [40].

A number of similar experiments were performed later for different types of nonlinear media, including artificial Kerr media made from liquid suspensions of submicrometer particles (see, e.g., Ref. [42] where the smallest-diameter self-trapped filaments ($\sim 2\mu$m) were observed.

More recent experiments on the beam self-focusing involved the so-called *vortex rings*, i.e. bright rings with a nonzero angular momentum created by passing the laser beam through a diffracting phase mask and then propagating it in a nonlinear medium (a 20-cm cell with rubidium vapor) [6, 7] and also for a quadratically nonlinear medium (KTP crystal) [43]. An angular momentum introduced in the input beam, strongly affects the dynamics of

bright spots (in fact, spatial solitons) created via the transverse instability of the rings, so that the solitons can attract and repel each other, or even fuse together.

Formation of a variety of different patterns of spots (bright spatial solitons) was investigated by Grantham *et al.* [44] in a sodium vapor. They varied the input beam power from 30 mW to 460 mW and observed spatial bifurcation sequences due to spatial instabilities seeded by intentionally introduced aberrations. They used the structure of the instability gain curve for an input-wavefront-encoding feedback to accelerate particular unstable wave vectors, and observed complicated spatial bifurcations as a function of intensity or detuning, with *"... complexity and beauty rivaling that of a kaleidoscope"*.

The analysis of self-focusing based on the spatial (2+1) dimensional NLS equation and associated with the spatial instabilities, bifurcations, and formation of spatial solitons is valid for both cw beams and long pulses. In contrast, short pulses undergoing self-focusing do not collapse to wavelength dimensions. A number of experimental results by Strickland and Corkum [45] demonstrated the resistence of short pulses ($\sim$ 50 fs) to self-focusing. In spite of the fact that these process can be modelled by the hyperbolic NLS equation with normal group-velocity dispersion, experimental results indicate that spectral dispersion and other non-slowly-varying are also important to explain different behaviour of short pulses.

A detailed experimental investigation of the self-focusing dynamics of a femtosecond pulse in a normally dispersive (glass) medium was recently reported by Ranka *et al.* [46] who observed one of the main effects predicted by the theory based on the hyperbolic NLS equation, i.e. the splitting of a short pulse (85-90 fs) into two pulses for the power above the threshold value, $P > P_{\rm cr} \approx 3MW$, and even additional splittings, for higher powers ($P \approx 4.8MW$). More recently, Diddams *et al.* [47] reported the similar effects for the propagation of intense fs pulses in fused silica. Frequency-resolved optical gating was used to characterise the pulse splitting into subpulses which were found to be not generally symmetric, in accordance with the theoretical predictions based on the three-dimensional NLS equation that includes the Raman effect, linear and nonlinear shock terms, and third-order dispersion [48].

Theoretical prediction and a number of experimental observations of beam self-trapping in photorefractive media allow to observe spatial solitons in crystals at relatively low input powers. The first observation of two-dimensional spatial solitons was reported by Shih *et al.* [49] (see also Shih *et al.* [50]) who used an electric field of 5.8 kV/cm applied to a crystal of strontium barium niobate (SBN) to create an effective self-focusing nonlinearity and trap an optical beam into a filament as small as 9.6 μm at micro-Watt power levels.

Experimental observation of break-up of a quasi-plane bright spatial soliton into a sequence of higher dimensional solitons due to the transverse

('neck'-type) instability was observed by Mamaev *et al.* [51].

Most of the experimental demonstrations of spatial and temporal dark solitons have been recently discussed in the paper by Kivshar and Luther-Davies [12]. For the case of spatial dark solitons, the experiments reported the creation of dark soliton stripes in a (2+1)-dimensional geometry. As discussed above, such stripes should be unstable due to transverse modulational instability which leads to stripe breakup and the eventual creation of optical vortex solitons. However, it turns out that this instability was avoided in the early experiments by the use of finite-sized background beams and weak nonlinearity. By increasing nonlinearity, the transverse instability should be observed even with finite sized beams. The first experiments to verify the existence of this transverse instability, and through it the creation of optical vortex solitons, have been performed by Tikhonenko *et al.* [52] using a continuous wave, Ti:sapphire laser and a nonlinear medium comprised of atomic rubidium vapour. Very similar observations, with less evidence of the stripe decay into a sequence of vortex solitons, were performed almost simultaneously by Mamaev *et al.* [53, 54] for spatial dark solitons in a photorefractive medium.

6 Soliton spiralling

Analysis of the soliton interaction in a bulk medium is one of the important issues for all-optical soliton-based nonlinear switching. In the wave guide geometry, the interaction of (1+1)-dimensional solitons was studied extensively both theoretically and experimentally. For a bulk medium, only recent experimental observations of stable (2+1)-dimensional solitons in photorefractive materials [49] and quadratic nonlinear media [8] made it possible to initiate an experimental study of fully 3D interactions of optical solitons. In particular, one of the interesting regimes of nonplanar soliton interactions is the soliton spiralling, first suggested for coherent soliton interaction in Ref. [55]. That suggestion, based on a phenomenology of ray optics, has been not confirmed by a detailed theoretical analysis of the nonplanar coherent interaction and scattering of stable solitary waves, developed for the case of the soliton interactions in a bulk quadratic medium [56] because the spiralling regime is unstable to small perturbations. Nevertheless, already in 1997 an experimental group from Princeton reported the observation of stable soliton spiralling in a bulk photorefractive medium for two incoherently interactying optical beams [57]. These experimental results were in a sharp contrast with all theoretical works. For a long time the stable soliton spiralling was believed to be impossible, and the experimental results to be controversial. Just recently, the combined efforts of two groups shed some light on the physical mechanism leading to a stable soliton spiralling [58]. As was found in Ref. [58], the new physical effect –

the so-called *induced coherence*– is a physical phenomenon that allows one to understand if the soliton spiralling is possible at all as a stable dynamical regime of the soliton interaction.

The main result of Ref. [58] can be understood from simple physics. Indeed, when two identical beams are launched together with a very small separation between them, they create an effective optical coupler. Due to the properties of an incoherent interaction in a photorefrcative optical medium, the nonlinearity of the interacting beams depends only on the total beam intensity. This means that two arms of the effective wave guide coupler are identical, and the energy should oscillate periodically between the solitons introducing an effective coherence into the initially incoherent soliton interaction. Due to such an induced coherence, the soliton interaction energy consists of two parts, coherent and incoherent. The periodic energy exchange between the solitons leads to rapid oscillations of the soliton parameters and effective suppression of the soliton phase-sensitive repulsion, supporting stable soliton spiralling as shown in Fig. 1.

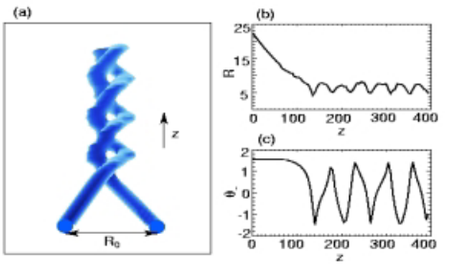

FIGURE 1. Stable soliton spiralling observed in numerical simulations: (a) 3D view; (b) oscillations of the soliton relative distance R, (c) large-amplitude oscilations of the relative power between the components characterized by the effective angle θ_- (see details in Ref. [58]).

The model used to obtain analytical results in Ref. [58] is isotropic. However, some nonlinear media are anisotropic, which means that the nonlinearity is not the same in all directions in the plane transverse to the propagation direction. For (1+1)-dimensional solitons, the anisotropy is manifested

in the fact that solitons can exist in specific crystalline orientations only and not in others. For example, (1+1)-dimensional quasi-steady-sate and screening photorefractive solitons can form with their trapping direction (i.e., their "narrow" direction) parallel to the applied field direction, but cannot form when their trapping direction is perpendicular to the applied field [59]. For (2+1)-dimensional photorefractive solitons, the situation is different: if the anisotropy is large, it can lead to solitons with *non-circular cross-section*. For photorefractive screening solitons the anisotropy is very small when the nonlinearity is saturated, so the resultant solitons are almost circular. However, far away from saturation (when the peak intensity of the solitons is much smaller than the saturation intensity), the solitons are elliptical. When the solitons are anisotropic, their interaction is expected to exhibit anisotropic features as well. For example, it turns out that (2+1)-dimenional incoherently-interacting photorefractive solitons exhibit anomalous repulsion [60], which is absent if the medium was isotropic. This happens only for an incoherent collision and when the colliding solitons are in a particular plane. Similarly, a full 3D interaction-collision in regimes where the anisotropy is large, leads to spiralling trajectories that are even more complicated than those of the isotropic case, and in some cases to fusion of the (otherwise spiraling indefinitely) solitons [61, 62]. In particular, when the input solitons do not possess any angular momentum, that is if the input beams have parallel trajectories (yet do not lie in the same plane), the solitons initially orbit a little about each other. They then settle down in a local minimum in the "induced potential" and, in many cases, eventually fuse [62]. We emphasize that this effect is a net result of the anisotropy, and does not exist for solitons in a nonlinear isotropic medium, in which solitons with zero initial angular momentum can never spiral. In other words, in isotropic nonlinear media the spiraling-orbiting motion conserves angular momentum, so in absence of initial angular momentum spiraling cannot exist.

7 Multi-hump solitons and solitonic gluons

A simple mechanism of the beam self-focusing can be applied to the general case when an optical beam consists of several components. In this case, a stationary state in the form of a spatial soliton is composed of two (or mode) modes, and its structure can be more complicated. We can view this composite soliton as an effective wave guide created by light in a nonlinear medium with one- (or more-) wave guide modes excited in it [63]. This picture holds provided all guided modes are orthogonal to each other, and the parametric processes such as four-wave mixing, do not occur. This is possible also for a photorefractive medium where an optical response is slow and the wave coupling is determined by the total beam intensity.

Indeed, the first observation of such multi-mode optical solitons was indeed reported for photorefractive crystals [64].

The most interesting feature of such multi-mode optical solitons is that their self-trapped stationary structure can be quite complicated, and the total soliton intensity profile may display several peaks. Such solitons are usually referred to as *multi-hump solitons*. In the simplest case, also corresponding to the nonlinearities of a photorefractive medium, spatial multi-hump solitons can be generated by incoherent interaction of two optical beams. The corresponding model was first introduced by Christodoulides *et al.* [65], and it is described by a system of two coupled (dimensionless) NLS equations

$$\begin{aligned} i\frac{\partial u}{\partial z} + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + \frac{u(|u|^2+|w|^2)}{1+s(|u|^2+|w|^2)} - u = 0, \\ i\frac{\partial w}{\partial z} + \frac{1}{2}\frac{\partial^2 w}{\partial x^2} + \frac{w(|u|^2+|w|^2)}{1+s(|u|^2+|w|^2)} - \lambda w = 0, \end{aligned} \tag{8.10}$$

where x and z are the transverse and propagation coordinates, respectively. The parameter λ stands for a ratio of the nonlinear propagation constants, and s is an effective *saturation parameter*. For $s \to 0$, the system (8.10) reduces to the integrable *Manakov equations* [66].

Stationary two-component solitons of the model (8.10) are described by $z-$independent solutions $u(x)$ and $w(x)$ with vanishing asymptotics as $|x| \to \infty$, and they were first analysed in Refs. [67]. Two-component solutions appear via bifurcations of one-component soliton $u_0(x)$, a solution of the first equation of the system (8.10) at $w = 0$, and they can be characterised by the dependence of the total soliton power,

$$P(\lambda, s) = \int_{-\infty}^{+\infty} (|u|^2 + |w|^2) dx,$$

on the parameter λ. The saturation parameter s determines the total number of guided modes and the cut-off values for each mode, $\lambda_n(s)$, i.e. the bifurcation points. Figure 2 shows the soliton bifurcation diagram for $s = 0.8$, where the horizontal line stands for the fundamental one-component soliton $u_0(x)$ that does not depend on s, and other curves show two families of composite solitons. Inserts demonstrate how the soliton profiles modify, with varying the soliton parameter λ, from one-hump to two- and three-hump solutions.

The important issue is stability of such multi-hump solitons, and for the model (8.10), the stability was analysed by Ostrovskaya *et al.* [68]. In contrast to what was believed before for other types of multi-hump multi-component solitons, Ostrovskaya *et al.* found that solitons of both the families of the composite solitons of the model (8.10) can be stable. Moreover, the two-hump solitons (upper row of the insert in Fig. 2) can also be stable, whereas the solitons of the second branch lose their stability

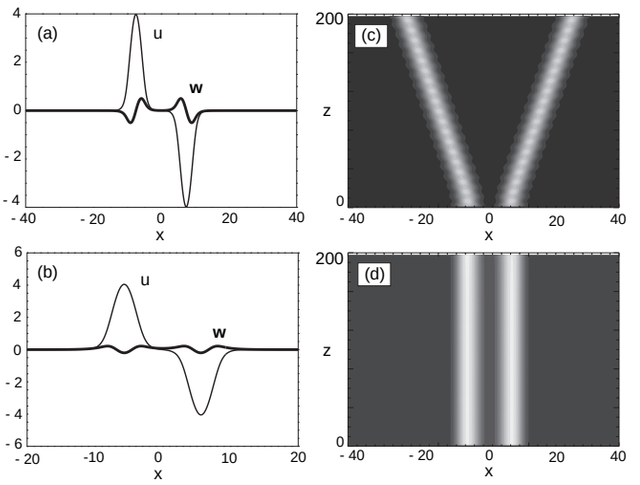

FIGURE 2. Soliton bifurcation diagram for the model (8.10). Inserts show the transverse profiles of the fields $u(x)$ (thin) and $w(x)$ (dashed), and the total intensity (thick) for the marked profiles [68].

well before they develop the humps. This gives us the first example of stable multi-hump solitons. Experimental results by Mitchell *et al.* [64] reported the first observation of stationary structures resembling multi-hump solitary waves that is in accord with the theory.

Existence of multi-hump solitons and their properties can be understood from a completely different approach valid in the limit of well-separated humps. Indeed, a multi-hump soliton can be considered as a soliton complex where the total force balance between the individual solitons and their components (as first-, second-, etc. guided modes) is satisfied. This idea gives birth to an interesting concept of the so-called *solitonic gluons*, non-soliton modes guided by solitons that allow to bind (or 'glue') solitons together forming multi-soliton states provided some phase conditions are satisfied (see an example in Fig. 3). The concept of solitonic gluons has been developed for different types of multi-component solitons [67, 69, 70]. The first experimental observation of the effect of solitonic gluons has been reported in Ref. [70] where the suppression of repulsion of a pair of dark solitons has been observed in the case of the second ('gluon') component added. All these results demonstrate a number of novel and important features in the interaction of multi-component solitary waves.

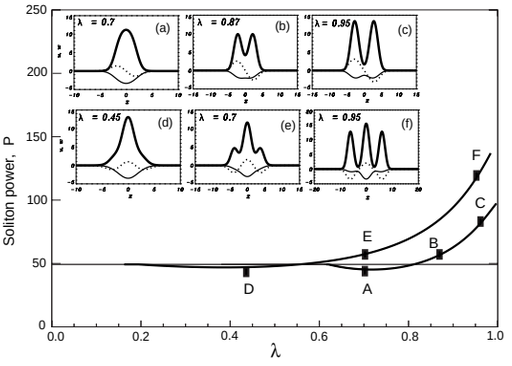

FIGURE 3. Two-soliton bound states in the model (8.10) with the solitonic gluons in the form of the first (a) and second (b) guided modes. (c,d) Interaction of two solitons from (a) for $s = 0.8$ and $\lambda = 0.63$, without and with the gluon component, respectively [70].

8 Discrete spatial optical solitons

In application to optical wave guides, the concept of a discrete soliton was first introduced by Christodoulides and Joseph [71] (see also [72] for an array of defocusing wave guides), who extended the ideas of nonlinear localised modes in discrete molecular chains to the case of an infinite array of identical, weakly coupled nonlinear wave guides. In such an array, when low intensity light is injected into one or a few neighbouring wave guides, it will couple to more and more wave guides and broaden its spatial distribution. For large intensities, the light distribution becomes self-trapped with a fixed spatial profile, and it involves only a few neighbouring wave guides; these are *discrete spatial solitons*. The theory of localised modes in wave guide arrays has been elaborated in detail, and different types of such modes, their steering and stability have been analysed [73]. The first observation of discrete spatial solitons in an array of 41 wave guides has been recently reported by Eisenberg, *et al.* [74].

The model for discrete spatial solitons can be introduced by considering a partial electric field amplitude E_n in the $n-$th wave guide, that can be also treated as *an averaged electric field* guided by the $n'th$ core of the wave guide array. Then, assuming the optical Kerr nonlinearity, we can write the difference equation,

$$i\frac{dE_n}{dz} + \beta E_n + C(E_{n+1} + E_{n-1}) + \gamma |E_n|^2 E_n = 0, \qquad (8.11)$$

where β is the linear propagation constant, C is the coupling coefficient, $\gamma = \omega_0 n_2 / c A_{\text{eff}}$, ω is the carrier wave frequency, n_2 is the nonlinear refractive

index, and $A_{\rm eff}$ is the effective area of the wave guide mode guided by the individual core with the number n (all wave guides are assumed to be identical).

Equation (8.11) is the well-known *discrete NLS equation* that occurs in many different physical problems (see, e.g., the review papers [75, 76] and references therein). At low input powers, we can neglect the nonlinear term in Eq. (8.11) and then this equation describes a simple effect when the light propagates from an initially excited $n'th$ wave guide into the two directions spreading the energy between the wave guide array. For large input powers, the light can become localised around the excited wave guide. For slowly varying distribution of the electric field in a wave guide array, an approximate solution can be obtained from the continuum limit approximation,

$$E_n(z) = E_0 \, \mathrm{sech}(\alpha n) \, \exp\left\{i(2C + \beta)z\right\}. \tag{8.12}$$

In spite of the fact that the solution (8.12) looks similar to a spatial soliton described by the continuous cubic NLS equation, the properties of such a discrete soliton are different, and some specific features in the soliton dynamics can be observed. For example, due to an inherent discreteness of the array, the soliton (8.12) cannot move or be steered freely; its motion is affected by the so-called *Peierls-Nabarro periodic potential* [77] that appears as a result of the fact that the transverse invariance of the discrete system is broken.

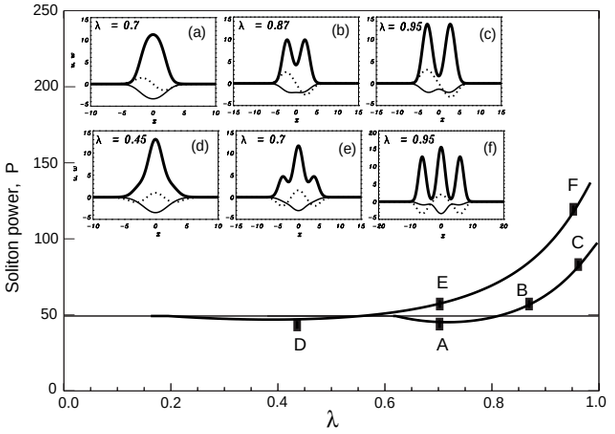

FIGURE 4. Images of the output facet of the 41 wave guides for different powers. (a) Peak power 70 W; linear regime. (b) Peak power 320 W; intermediate regime. (c) Peak power 500 W; *a discrete soliton* is formed [74].

In experiments, Eisenberg *et al.* [74] studied arrays of 41 ridge wave guides etched onto an AlGaAs substrate. The wave guides of 4 μm wide were etched on top of a slab wave guide. Samples with different separation D, that affects the coupling coefficient C in Eq. (8.11), were used ($D = 4, 5,$ and 7μm). The light source was an optical parametric oscillator (pumped by a Ti:sapphire laser) tuned to a wavelength of 1.53 μm, below the half-bandgap of the AlGaAs material, to minimise nonlinear absorption. The input beam was reshaped into an oval shape to match closely a guided mode of a single wave guide of the array. Some results of the output images from a 6 mm long wave guide array ($D = 4\mu$m) are presented in Fig. 4. At low power, the light spreads among nearly all the 41 wave guides, see Fig. 4(a). Increasing the power narrows the light distribution [see Fig. 4(b)] until a discrete spatial soliton is formed [see Fig. 4(c)]. Similar results were obtained for the wave guide separations of 5 and 7 μm.

Acknowledgments: I wish to thank Peter Christiansen and Mads Peter Sørensen for their great support and patience that helped me to complete this work. This short summary of the recent advances in the field of spatial optical solitons does not cover a broad diversity of the different topics in this rapidly developing field, and the problems discussed here are dictated by the author's research interests. My research work addressed in this Chapter has been carried out in collaboration with my colleagues and PhD students from the Optical Sciences Centre and Laser Physics Centre of the Australian National University, and also in collaboration with many overseas colleagues, including M. Segev and his group, D. Pelinovsky, and D. Skryabin. I am deeply indebted to all of them.

9 References

[1] R.Y. Chiao, E. Garmire, and C.H. Townes, Phys. Rev. Lett. **13**, 479 (1964).

[2] V.E. Zakharov and A.B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].

[3] P.L. Kelley, Phys. Rev. Lett. **15**, 1005 (1965).

[4] For a comprehensive overview of the wave collapse phenomenon, see J.J. Rasmussen and K. Rypdal, Phys. Scripta **33**, 481 (1986).

[5] M. Segev, G.C. Valley, B. Crosignani, P. DiPorto, and A. Yariv, Phys. Rev. Lett. **73**, 3211 (1994); M. Shih, P. Leach, M. Segev, M.H. Garrett, G. Salamo, and G.C. Valley, Opt. Lett. **21**, 324 (1996); M.D. Iturbe-Castillo, P.A. Marquez Aguilar, J.J. Sanchez-Mondragon, S. Stepanov, and V. Vysloukh, Appl. Phys. Lett. **64**, 408 (1994).

[6] V. Tikhonenko, J. Christou, and B. Luther-Davies, J. Opt. Soc. Am. B **12**, 2046 (1995).

[7] V. Tikhonenko, J. Christou, and B. Luther-Davies, Phys. Rev. Lett. **76**, 2698 (1996).

[8] W.E. Torruellas, Z. Wang, D.J. Hagan, E.W. VanStryland, G.I. Stegeman, L. Torner, and C.R. Menyuk, Phys. Rev. Lett. **74**, 5036 (1995).

[9] S.A. Akhmanov, A.P. Sukhorukov, and R.V. Khokhlov, Usp. Fiz. Nauk **93**, 19 (1967) [Sov. Phys. Uspekhi **10**, 609 (1968)].

[10] See, e.g., A. Hasegawa and Y. Kodama, *"Solitons in Optical Communications"* (Oxford University Press, Oxford, 1995).

[11] For a comprehensive overview of the soliton perturbation theory, see Yu.S. Kivshar and B.A. Malomed, Rev. Mod. Phys. **61**, 763 (1989).

[12] Yu.S. Kivshar and B. Luther-Davies, Phys. Rep. **298**, 81 (1998).

[13] A.E. Kaplan, Phys. Rev. Lett. **78**, 1291 (1985).

[14] I. Bialynicki-Birula and J. Mycielski, Phys. Scripta **20**, 539 (1979).

[15] A.W. Snyder and D.J. Mitchell, Opt. Lett. **22**, 16 (1997).

[16] M. G. Vakhitov and A. A. Kolokolov, Radiophys. Quantum Electron **16**, 783 (1973) [Izv. Vyssh. Uch. Zav. Radiofizika **16**, 1020 (1973)].

[17] E. A. Kuznetsov, A. M. Rubenchik, and V. E. Zakharov, Phys. Rep. **142**, 103 (1986).

[18] F.V. Kusmartsev, Phys. Rep. **183**, 2 (1989).

[19] C.K.R.T. Jones and J. Moloney, Phys. Lett. A **117**, 175 (1986); D. Hart and E.M. Wright, Opt. Lett. **17**, 121 (1992).

[20] D.E. Pelinovsky, A.V. Buryak, and Yu.S. Kivshar, Phys. Rev. Lett. **75**, 591 (1995).

[21] D.E. Pelinovsky, V.V. Afanasjev, and Yu.S. Kivshar, Phys. Rev. E **53**, 1940 (1996).

[22] D.E. Pelinovsky, Yu.S. Kivshar, and V.V. Afanasjev, Phys. Rev. E **54**, 2015 (1996).

[23] I.V. Barashenkov and Kh.T. Kholmurodov, JINR Preprint P17-86-698, Dubna (in Russian).

[24] M.M. Bogdan, A.S. Kovalev, and A.M. Kosevich, Fiz. Nizk. Temp. **15**, 511 (1989) [Sov. J. Low Temp. Phys. **15**, 288 (1989)].

[25] Yu.S. Kivshar and W. Krolikowski, Opt. Lett. **20**, 1527 (1995).

[26] I.V. Barashenkov, Phys. Rev. Lett. **77**, 1193 (1996).

[27] Yu.S. Kivshar and V.V. Afanasjev, Opt. Lett. **21**, 1135 (1996).

[28] A.V. Buryak, Yu.S. Kivshar, and S. Trillo, Phys. Rev. Lett. **77**, 5210 (1996).

[29] A.V. Buryak and Yu.S. Kivshar, Phys. Rev. Lett. **78**, 3286 (1997).

[30] L. Torner, D. Mihalache, D. Mazilu, and N.N. Akhmediev, Opt. Commun. **138**, 105 (1997).

[31] I.V. Barashenkov, D.E. Pelinovsky, and E.V. Zemlyanaya, Phys. Rev. Lett. **80**, 5117 (1998).

[32] P.B. Lundquist, D.R. Andersen, and Yu.S. Kivshar, Phys. Rev. E **57**, 3551 (1998).

[33] G.A. Askar'yan, Zh. Eksp. Teor. Fiz. **42**, 1567 (1962) [Sov. Phys. JETP **15**, 1088 (1962)].

[34] V.I. Talanov, Izv. VUZov-Radiofizika **7**, 564 (1964) (in Russian).

[35] N.F. Pilipetskii and A.R. Rustamov, Pisma Zh. Eksp. Teor. Fiz. **2**, 88 (1965)

[36] E. Garmire, R.Y. Chiao, and C.H. Townes, Phys. Rev. Lett. **16**, 347 (1966).

[37] Z.K. Yankauskas, Izv. VUZov Radiofiz. **9**, 412 (1966) [Sov. Radiophys. **9**, 261 (1966)].

[38] H.A. Haus, Appl. Phys. Lett. **8**, 128 (1966).

[39] A.J. Campillo, S.L. Shapiro, and B.R. Snydam, Appl. Phys. Lett. **23**, 628 (1973).

[40] A.J. Campillo, S.L. Shapiro, and B.R. Snydam, Appl. Phys. Lett. **24**, 178 (1974).

[41] J.E. Bjorkholm and A. Ashkin, Phys. Rev. Lett. **32**, 129 (1974).

[42] A. Ashkin, J.M. Dziedzic, and P.W. Smith, Opt. Lett. **7**, 276 (1982).

[43] D.V. Petrov, L. Torner, J. Martorell, R. Vilaseca, J.P. Torres, and C. Cojocaru, Opt. Lett. **23**, 1444.(1998).

[44] J.W. Grantham, H.M. Gibbs, G. Khitrova, J.F. Valley, and Xu Jiajin, Phys. Rev. Lett. **66**, 1422 (1991).

[45] D. Strickland and P.B. Corkum, J. Opt. Soc. Am. B **11**, 492 (1994).

[46] J.K. Ranka, R.W. Schirmer, and A.L. Gaeta, Phys. Rev. Lett. **77**, 3783 (1996).

[47] S.A. Diddams, H.K. Eaton, A.A. Zozulya, and T.S. Clement, Opt. Lett. **23**, 379 (1998).

[48] A.A. Zozulya, S.A. Diddams, and T.S. Clement, Phys. Rev. A **58**, 3303 (1998).

[49] M.F. Shih, M. Segev, G.C. Valley, G. Salamo, B. Crosignani, and P. Di Porto, Electron. Lett. **31**, 826 (1995).

[50] M.F. Shih, P. Leach, M. Segev, M.H. Garrett, G. Salamo, and G.C. Valley, Opt. Lett. **21**, 324 (1996).

[51] A.V. Mamaev, M. Saffman, and A.A. Zozulya, Europhys. Lett. **35**, 25 (1996).

[52] V. Tikhonenko, J. Christou, B. Luther-Davies, and Yu.S. Kivshar, Opt. Lett. **21**, 1129 (1996).

[53] A.V. Mamaev, M. Saffman, and A.A. Zozulya A.A., Phys. Rev. Lett. **76**, 2262 (1996).

[54] A.V. Mamaev, M. Saffman, D.Z. Anderson, and A.A. Zozulya, Phys. Rev. A **54**, 870 (1996).

[55] L. Poladian, A.W. Snyder, and D.J. Mitchell, Opt. Commun. **85**, 59 (1991).

[56] V.V. Steblina, Yu.S. Kivshar, and A.V. Buryak, Opt. Lett. **23**, 156 (1998).

[57] M. Shih, M. Segev, and G. Salamo, Phys. Rev. Lett. **78**, 2551 (1997).

[58] A.V. Buryak, Yu.S. Kivshar, M. Shih, and M. Segev, Phys. Rev. Lett. **82**, 81 (1999).

[59] G. Duree, G. Salamo, M. Segev, A. Yariv, B. Crosignani, P. DiPorto, and E. Sharp, Opt. Lett. **19**, 1195 (1994).

[60] W. Krolikowski, M. Saffman, B. Luther-Davies, and C. Denz, Phys. Rev. Lett. **80**, 3240 (1998).

[61] A. Stepken, M. R. Belic, F. Kaiser, W. Krolikowski, and B. Luther-Davies, Phys. Rev. Lett. **82**, 540 (1999).

[62] M. R. Belic, A. Stepken, and F. Kaiser, Phys. Rev. Lett. **82**, 544 (1999).

[63] A.W. Snyder, S.J. Hewlett, and D.J. Mitchell, Phys. Rev. Lett. **72**, 1012 (1994).

[64] M. Mitchell, M. Segev, and D.N. Christodoulides, Phys. Rev. Lett. **80**, 4657 (1998).

[65] D.N. Christodoulides, S.R. Singh, M.I. Calvalho, and M. Segev, Appl. Phys. Lett. **68**, 1763 (1996).

[66] S.V. Manakov, Sov. Phys. JETP **38**, 248 (1974).

[67] E.A. Ostrovskaya and Yu.S. Kivshar, J. Opt. B: Quant. Semiclass. Opt. **1**, 81 (1999).

[68] E.A. Ostrovskaya, Yu.S. Kivshar, D.V. Skryabin, and W.J. Firth, *"Stability of multi-hump optical solitons"*, submitted to Phys. Rev. Lett. (1999).

[69] E.A. Ostrovskaya and Yu.S. Kivshar, Opt. Lett. **23**, 1268 (1998).

[70] E.A. Ostrovskaya, Yu.S. Kivshar, Z. Chen, and M. Segev, *"Interaction of vector solitons and solitonic gluons"*, Opt. Lett. **24**, March 1 (1999).

[71] D.N. Christodoulides and R.I. Joseph, Opt. Lett. **13**, 794 (1988).

[72] Yu.S. Kivshar, Opt. Lett. **18**, 1147 (1993).

[73] See, e.g., W. Królikowski and Yu.S. Kivshar, J. Opt. Soc. Am. B **13**, 876 (1996); A.B. Aceves *et al.*, Phys. Rev. E **53**, 1172 (1996); M. Johansson and Yu.S. Kivshar, Phys. Rev. Lett. **82**, 85 (1999).

[74] H.S. Eisenberg, Y. Silberberg, R. Morandotti, A.R. Boyd, and J.S. Aitchison, Phys. Rev. Lett. **81**, 3383 (1998).

[75] S. Flach and C.R. Willis, Phys. Rep. **295**, 182 (1998).

[76] O.M. Braun and Yu.S. Kivshar, Phys. Rep. **306**, 1 (1998).

[77] See, e.g., Yu.S. Kivshar and D.K. Campbell, Phys. Rev. E **48**, 3077 (1993).

9
Nonlinear Fiber Optics

Govind P. Agrawal

1 Introduction

Nonlinear fiber optics concerns with the nonlinear optical phenomena occurring inside optical fibers. Although the field of nonlinear optics traces its beginning to 1961, when a ruby laser was first used to generate the second-harmonic radiation inside a crystal [1], the use of optical fibers as a nonlinear medium became feasible only after 1970 when fiber losses were reduced to below 20 dB/km [2]. Stimulated Raman and Brillouin scatterings in single-mode fibers were studied as early as 1972 [3] and were soon followed by the study of other nonlinear effects such as self- and cross-phase modulation and four-wave mixing [4]. By 1989, the field of nonlinear fiber optics has advanced enough that a whole book was devoted to it [5]. This book or its second edition has been translated into Chinese, Japanese, and Russian languages, attesting to the worldwide activity in the field of nonlinear fiber optics.

Nonlinear fiber optics has continued to grow during the decade of 1990s, perhaps even more dramatically than anticipated. This growth is motivated by several recent advances in lightwave technology, the most important being the advent of high-capacity fiber-optic communication systems [6]. In such systems, the transmitted signal is amplified periodically by using optical amplifiers to compensate for residual fiber losses. As a result, the nonlinear effects accumulate over long distances, and the effective interaction length can exceed thousands of kilometers! Among other fiber devices in which nonlinear effects are becoming increasingly important are mode-locked fiber lasers, distributed fiber amplifiers, and fiber Bragg gratings. It is impossible to review the entire field of nonlinear fiber optics in a chapter of this size. The focus therefore is on optical solitons and optical switching, the two topics that are driven by advances in the field of fiber-optic communications and are likely to remain important as we enter the 21st century.

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 195−211, 2000.

2 Fiber characteristics

Before describing the nonlinear effects in optical fibers, it is worthwhile to ponder why optical fibers are useful for nonlinear optics. This section describes the properties of optical fibers that are unique to them and their relevance to the study of nonlinear optical phenomena.

2.1 Single-mode fibers

An optical fiber looks like a thin strand of glass and consists of a central core surrounded by a cladding whose refractive index is slightly lower than that of the core. Both the core and the cladding are made of fused silica, a glassy material with an ultra-low loss (about 0.2 dB/km) in the near-infrared region near 1.5 μm. The refractive-index difference between the core and the cladding is realized by the selective use of dopants during the fabrication process. Dopants such as GeO_2 and P_2O_5 increase the refractive index of pure silica and are suitable for the core, while materials such as boron and fluorine are used for the cladding because they decrease the refractive index of silica. Even a relatively small refractive-index difference between the core and the cladding (typically less than 1%) can guide the light along the fiber length through the well-known phenomenon of total internal reflection.

The guiding properties of an optical fiber are characterized by a dimensionless parameter defined as

$$V = a(\omega/c)(n_1^2 - n_2^2)^{1/2}, \tag{9.1}$$

where a is the core radius, ω is the frequency of light, and n_1 and n_2 are the refractive indices of the core and the cladding, respectively. The parameter V determines the number of mode supported by the fiber. Optical fibers with $V < 2.405$ support only a single mode and are called single-mode fibers. Such fibers have a microscopic core ($a < 5$ μm) and are used almost exclusively for a variety of applications including optical communications.

2.2 Fiber nonlinearities

The response of any dielectric to light becomes nonlinear for intense electromagnetic fields. In the transparent region of optical fibers, the lowest-order nonlinear effects originate from the third-order susceptibility $\chi^{(3)}$, which is responsible for phenomena such as third-harmonic generation, four-wave mixing (FWM), and nonlinear refraction. Among these, nonlinear refraction, a phenomenon referring to the intensity dependence of the refractive index, plays the most important role. The effective refractive index of the fiber mode has a general form

$$\tilde{n}(\omega, I) = n(\omega) + n_2 I, \tag{9.2}$$

where $n(\omega)$ is the linear part of the mode index at the frequency ω, I is the optical intensity, and n_2 is the nonlinear parameter related to $\chi^{(3)}$. Several physical mechanisms contribute to n_2, the dominant contribution coming from the anharmonic motion of valence electrons. Because of a fast response of such electrons, the frequency dependence of n_2 can often be ignored.

The intensity dependence of the refractive index leads to a large number of interesting nonlinear effects; the two most widely studied are self-phase modulation (SPM) and cross-phase modulation (XPM). SPM refers to the self-induced phase shift experienced by an optical field during its propagation inside optical fibers. Its magnitude can be obtained by noting that the phase of an optical field changes during transmission through the fiber by

$$\phi = (n + n_2 I)k_0 L, \tag{9.3}$$

where $k_0 = \omega/c = 2\pi/\lambda$, λ is the wavelength, and L is the fiber length. The nonlinear phase shift resulting from SPM is $\phi_{NL} = n_2 k_0 L I$.

Silica glass is a relatively weak nonlinear medium with a measured value of $n_2 \approx 2.7 \times 10^{-20}$ m^2/W. For silica fibers this value can vary in the range $n_2 = 2.2$–3.0×10^{-20} m^2/W depending on the density of dopants and on whether the fiber preserves polarization of light [7]. However, even though n_2 is relatively small compared with most other nonlinear media, the nonlinear phase shift ϕ_{NL} can become large since the intensity I is enhanced in optical fibers by orders of magnitude because of a small mode diameter (typically < 10 μm). At the same time, relatively low losses in fibers can maintain this intensity over long lengths (~ 10 km). If fiber losses are compensated periodically by using optical amplifiers, the interaction length L can exceed thousands of kilometers. It is this combination of a high intensity and a long interaction length that makes optical fibers so attractive for nonlinear optics.

2.3 Group-velocity dispersion

As seen in Eq. (9.2), the refractive index in fibers also depends on the optical frequency ω. This chromatic dispersion plays an important role in nonlinear fiber optics and leads to formation of optical solitons under certain conditions. To understand its significance, consider a single-mode fiber of length L. A specific spectral component at the frequency ω would arrive at the output end of the fiber after a time delay $T = L/v_g$, where v_g is the group velocity defined as $v_g = (d\beta/d\omega)^{-1}$ and $\beta = n(\omega)\omega/c$ is the propagation constant.

The frequency dependence of the group velocity leads to pulse broadening simply because different spectral components of the pulse do not arrive simultaneously at the fiber output. If $\Delta\omega$ is the spectral width of the pulse,

the extent of pulse broadening is governed by

$$\Delta T = \frac{dT}{d\omega}\Delta\omega = \frac{d}{d\omega}\left(\frac{L}{v_g}\right)\Delta\omega = L\frac{d^2\beta}{d\omega^2}\Delta\omega \equiv L\beta_2\Delta\omega. \tag{9.4}$$

This phenomenon is referred to as the group-velocity dispersion (GVD), and the parameter $\beta_2 = d^2\beta/d\omega^2$ is known as the GVD parameter.

In standard silica fibers, β_2 changes sign from positive to negative as wavelength of light increases beyond 1.3 μm. The region in which β_2 is negative is referred to as the anomalous-GVD regime. The magnitude of β_2 can be controlled by shifting the wavelength at which β_2 changes sign. Dispersion-shifted fibers used for optical communications are designed to have $\beta_2 = 0$ near 1.5 μm. It is possible to design fibers such that β_2 is relatively small over a wide wavelength range extending from 1.3 to 1.6 μm. Such fibers are called *dispersion-flattened fibers*. Recently, attempts have been made to produce fibers whose GVD decreases along the fiber length through axial variations in the core radius. Such fibers are called *dispersion-decreasing fibers* and are likely to become important in near future.

3 Pulse propagation in fibers

3.1 Nonlinear Schrödinger equation

Most nonlinear effects in optical fibers are observed by using short optical pulses because the dispersive effects are enhanced for such pulses. Propagation of optical pulses through fibers can be studied by solving Maxwell's equations. In the slowly varying envelope approximation, these equations lead to the following nonlinear Schrödinger equation (NSE) [5]

$$\frac{\partial A}{\partial z} + \frac{i}{2}\beta_2\frac{\partial^2 A}{\partial t^2} = i\gamma|A|^2A - \frac{\alpha}{2}A, \tag{9.5}$$

where $A(z,t)$ is the slowly varying envelope associated with the optical pulse, α accounts for fiber losses, β_2 governs the GVD effects, and γ is the nonlinear parameter defined as

$$\gamma = n_2\omega/(cA_{\text{eff}}). \tag{9.6}$$

Here A_{eff} is the effective core area of the fiber. This equation is appropriate for pulses wider than 5 ps. For an accurate description of shorter pulses, several higher-order dispersive and nonlinear terms must be added to the NSE [5].

3.2 Modulation instability

The nonlinear phenomenon of modulation instability is perhaps the simplest to understand since it follows readily from the NSE and emphasizes

the role played by the GVD. Consider propagation of a continuous-wave (CW) beam of input power P_0 inside an optical fiber. If the fiber loss is ignored by setting $\alpha = 0$, Eq. (9.5) is easily solved to yield $A_s(z) = \sqrt{P_0}\exp(i\phi_{NL})$, where $\phi_{NL} = \gamma P_0 z$ is the SPM-induced nonlinear phase shift. Equation (9.5) thus shows that a CW beam should propagate through the fiber unchanged except for acquiring a power-dependent phase shift.

Before reaching this conclusion, however, one must ask whether the steady-state solution $A_s(z)$ is stable against small perturbations. To answer this question, we perturb the steady state slightly such that $A(z,t) = [\sqrt{P_0} + a(z,t)]\exp(i\phi_{NL})$ and linearize Eq. (9.5) in $a(z,t)$ by assuming that the perturbation is relatively small. The resulting equation,

$$\frac{\partial a}{\partial z} + \frac{i}{2}\beta_2\frac{\partial^2 a}{\partial t^2} = i\gamma P_0(a + a^*), \tag{9.7}$$

is readily solved by assuming a general solution of the form

$$a(z,t) = a_1\cos(Kz - \Omega t) + ia_2\sin(Kz - \Omega t), \tag{9.8}$$

where Ω is the perturbation frequency and K is the wave number associated with it. By using Eqs. (9.7) and (9.8), K is found to be given by [5]

$$K(\Omega) = \pm\frac{1}{2}|\beta_2|\Omega|[\Omega^2 + (4\gamma P_0/\beta_2)]^{1/2}. \tag{9.9}$$

The dispersion relation (9.9) shows that stability of the CW solution depends critically on whether light experiences normal or anomalous GVD inside the fiber. In the case of normal GVD ($\beta_2 > 0$), the wave number K is real for all Ω, and the CW beam is stable against small perturbations. By contrast, in the case of anomalous GVD ($\beta_2 < 0$), K becomes imaginary for a range of frequencies, and the perturbation $a(z,t)$ grows exponentially with z. Thus, propagation of CW beams through fibers is inherently unstable when $\beta_2 < 0$. This instability is referred to as the modulation instability since it leads to spontaneous modulation of the CW beam at a certain frequency whose value depends on the beam power. Similar instabilities occur in many other nonlinear systems and are often called self-pulsing instabilities.

4 Optical solitons

The occurrence of modulation instability in the anomalous-GVD regime of optical fibers is ultimately related to the solutions of the NSE known as *solitons*. In this section we first discuss the bright and dark solitons and then focus on the use of bright solitons for fiber-optic communication systems.

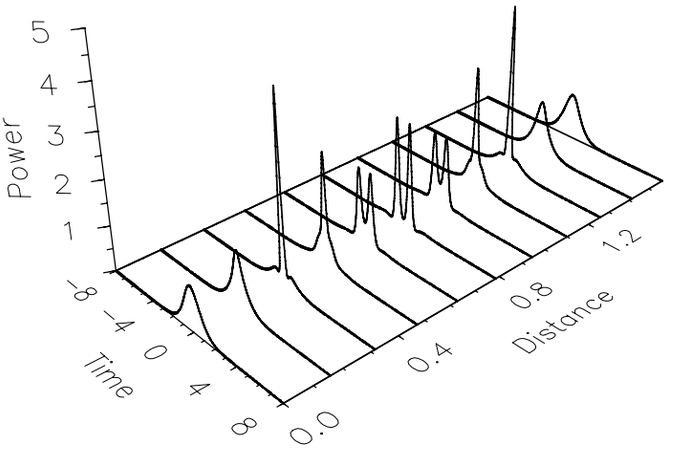

FIGURE 1. Evolution of a third-order soliton over one soliton period.

4.1 Bright solitons

The NSE belongs to a special class of nonlinear differential equations that can be solved exactly by using a mathematical technique known as the inverse scattering method [8]. It is useful to write Eq. (9.5) in a normalized form by introducing the so-called soliton units as

$$\tau = t/T_0, \qquad \xi = z/L_D, \qquad u = \sqrt{\gamma L_D}\, A, \tag{9.10}$$

where $L_D = T_0^2/|\beta_2|$ is the dispersion length and T_0 is related to the pulse width. If we neglect fiber losses ($\alpha = 0$), Eq. (9.5) takes its canonical form

$$i\frac{\partial u}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 u}{\partial \tau^2} + |u|^2 u = 0, \tag{9.11}$$

where + or − sign is chosen depending on whether the GVD is anomalous or normal. The NSE supports solitons for both normal and anomalous GVD, known as the dark and bright solitons, respectively.

Consider first this case of bright solitons by choosing the + sign in Eq. (9.11). The analytical solutions of this equation can be summarized as follows. When an input pulse having an initial amplitude

$$u(0, \tau) = N\mathrm{sech}(\tau) \tag{9.12}$$

is launched into the fiber, its shape remains unchanged during propagation when $N = 1$ but follows a periodic pattern for integer values of $N > 1$ such that the input shape is recovered at $\xi = m\pi/2$, where m is an integer. The optical pulse corresponding to $N = 1$ is called the *fundamental soliton*. Pulses corresponding to other integer values of N are called higher-order solitons. The parameter N represents the order of the soliton. As an

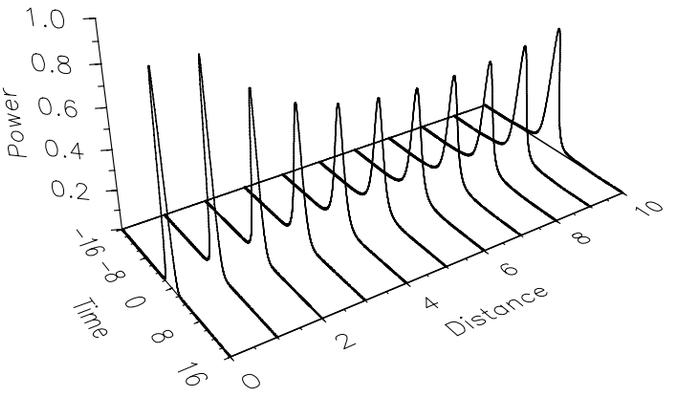

FIGURE 2. Evolution of a Gaussian pulse with $N = 1$ over the range $\xi = 0$–10.

example, Fig. 1 shows evolution of a third-order soliton ($N = 3$) over one soliton period ($z_0 = (\pi/2)L_D$). This soliton exhibits frequency chirp, defined as the time derivative of the soliton phase. Only the fundamental soliton remains chirp-free during propagation while maintaining its shape. Indeed, the solution of the NLS for $N = 1$ can be written as

$$u(\xi, \tau) = \operatorname{sech}(\tau) \exp(i\xi/2). \tag{9.13}$$

It shows that the input pulse acquires a phase shift $\xi/2$ as it propagates inside the fiber, but its amplitude remains unchanged. It is this property of solitons that makes them an ideal candidate for optical communications. In essence, the effects of fiber dispersion are exactly compensated by the fiber nonlinearity when the input pulse has a "sech" shape and its width and peak power are related such that $N = 1$.

An important property of optical solitons is that they are remarkably stable against perturbations. Thus, even though the fundamental soliton requires a specific shape and a certain peak power, it can be generated even when the pulse shape and the peak power deviate from the ideal conditions. Figure 2 shows evolution of a Gaussian input pulse for which $N = 1$ but $u(0, \tau) = \exp(-\tau^2/2)$. As seen there, the pulse adjusts its shape and width to become a fundamental soliton and attains a "sech" profile for $\xi \gg 1$. A similar behavior is observed when N deviates from 1. In fact, a fundamental soliton can be excited for values of N in the range 0.5–1.5.

4.2 Dark solitons

The NSE can be solved by the inverse scattering method even in the case of normal dispersion. The intensity profile of the resulting solutions exhibits a dip in a uniform background, and it is the dip that remains unchanged during propagation inside the fiber. For this reason, such solutions of the

NSE are called *dark solitons*. This section describes briefly the properties of dark solitons [9]–[13]. We refer the reader to a 1998 review for details [13].

The NSE describing dark solitons is obtained from Eq. (9.11) by choosing the − sign for the second term. Its general solution can be written as [10]

$$u_d(\xi,\tau) = (\eta \tanh\zeta - i\kappa)\, \exp(iu_0^2\xi), \tag{9.14}$$

where

$$\zeta = \eta(\tau - \kappa\xi), \quad \eta = u_0 \cos\phi, \quad \kappa = u_0 \sin\phi, \tag{9.15}$$

u_0 is the amplitude of the CW background, ϕ is an internal phase angle in the range $0 < \phi < \pi/2$), and η and κ are the amplitude and velocity of the dark soliton, respectively.

An important difference between the bright and dark solitons is that the velocity κ of a dark soliton depends on its amplitude η through the internal phase angle ϕ. For $\phi = 0$, Eq. (9.14) reduces to $|u_d(\xi,\tau)| = u_0 \tanh(u_0\tau)$, a form that shows that the soliton power drops to zero at the center of the dip. Such a soliton is referred as the *black* soliton. When $\phi \neq 0$, the intensity does not drop to zero at the dip center; such solitons are called *gray* solitons. Another interesting feature of dark solitons is related to their phase. In contrast with bright solitons which have a constant phase, the phase of a dark soliton changes across its width, i.e., dark solitons are generally chirped.

Several techniques can be used to generate dark solitons, including electric modulation in one arm of a Mach–Zehnder interferometer [9], nonlinear conversion of a beat signal in a dispersion-decreasing fiber [11], and conversion of a nonreturn-to-zero (NRZ) signal into a return-to-zero (RZ) signal and then into dark solitons by using a balanced Mach–Zehnder interferometer [12]. In a 1995 experiment [12], a 10-Gb/s signal was transmitted over 1200 km by using dark solitons. Further improvements are likely to occur with the development of sources capable of generating a dark-soliton bit stream with little amplitude and width fluctuations.

4.3 Loss-managed solitons

As discussed earlier, solitons use fiber nonlinearity to maintain their width in the presence of fiber dispersion. However, this property holds only if the fiber loss is negligible. It is not difficult to see that a decrease in the soliton energy because of fiber loss would lead to soliton broadening simply because the reduced peak power weakens the SPM effect necessary to counteract the GVD. To overcome the effect of fiber loss, solitons need to be amplified periodically to recover their original width, peak power, and energy. A fiber amplifier is commonly used for this purpose.

An important design parameter for long-haul communication systems is the spacing L_A between amplifiers. For non-soliton lightwave systems, L_A is typically in the range 60–100 km. For soliton communication systems, L_A

is restricted to smaller values. The reason is that optical amplifiers boost the soliton energy over a relatively short distance ($\sim$ 10 m). The amplified soliton adjusts its width dynamically in the fiber section following the optical amplifier. However, it also sheds a part of its energy as dispersive waves during this adjustment phase. The dispersive part can accumulate to significant levels over a large number of amplification stages and must be avoided. One way to reduce the dispersive part is to reduce the amplifier spacing L_A such that the soliton is not perturbed much over this length. Numerical and analytical results show that this is the case when L_A is a small fraction of the dispersion length [14]. In such a system, solitons can be amplified hundreds of time while preserving their shape. Since soliton evolution is governed by the average soliton energy over one amplifier spacing, this mode of operation is referred to as the *average-soliton regime* and the corresponding solitons are called guiding-center solitons [14].

The periodic amplification of solitons can be accounted for by adding the gain and loss terms to Eq. (9.11), resulting in

$$i\frac{\partial u}{\partial \xi} + \frac{1}{2}\frac{\partial^2 u}{\partial \tau^2} + |u|^2 u = -\frac{i}{2}\Gamma u + i(\sqrt{G}-1)\sum_{m=1}^{N_A}\delta(\xi - m\xi_A)u, \qquad (9.16)$$

where N_A is the total number of amplifiers, $\xi_A = L_A/L_D$, $\Gamma = \alpha L_D$, and $G = \exp(\alpha L_A)$ is the amplifier gain needed to compensate for the fiber loss. The delta function accounts for the lumped nature of amplification at locations $\xi = m\xi_A$. The factor $\sqrt{G}-1$ represents the change in the soliton amplitude during amplification. In the limit $\xi_A \ll 1$, this equation can be reduced to the standard NSE for the guiding-center soliton whose evolution is governed by Eq. (9.13) provided the input peak power is given by

$$P_{\text{in}} = \frac{G\ln G}{G-1}\,P_0, \qquad (9.17)$$

where P_0 is the peak power required when $\alpha = 0$. As an example, $G = 10$ and $P_{\text{in}} \approx 2.56\,P_0$ for 50-km amplifier spacing and a fiber loss of 0.2 dB/km. The enhanced input power balances the impact of fiber losses in an average sense. Figure 3 shows soliton evolution for this case over a distance of 10,000 km. When soliton width corresponds to a dispersion length of 200 km, soliton is preserved quite well even after 200 amplification stages since the condition $\xi_A \ll 1$ is reasonably well satisfied. However, if the dispersion length is reduced to 25 km, the soliton is destroyed because it no longer propagates in the average-soliton regime.

4.4 Dispersion-managed solitons

The dispersion-management technique consists of using multiple sections of constant-dispersion fibers whose lengths and GVDs are judiciously chosen.

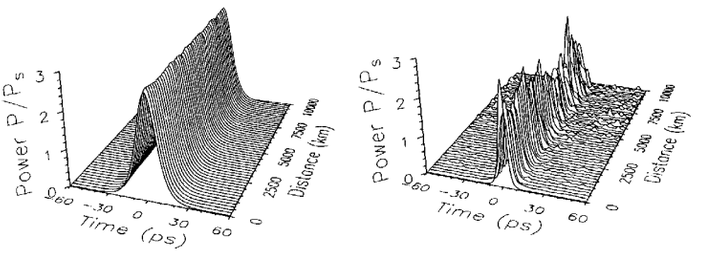

FIGURE 3. Evolution of a loss-managed soliton over a distance of 10,000 km with $L_A = 50\ km$ for two cases of $L_D = 200$ km (left) and $L_D = 25$ km (right). Soliton is destroyed when amplifier spacing exceeds the dispersion length.

This technique has attracted considerable attention since 1995 because of several advantages offered by it. It is especially advantageous in the context of wavelength-division multiplexing (WDM) where FWM can lead to considerable interchannel crosstalk if the GVD is reduced by operating close to zero-dispersion wavelength. Dispersion management solves this problem since it can lower the average GVD of the entire link while keeping the GVD of each section large enough that the FWM and higher-order dispersive effects remain negligible. Variants of this scheme are also referred to as partial soliton communication [15] and dispersion allocation [16].

The propagation of solitons through dispersion-managed fiber links can be studied by using Eq (9.11) after including fiber losses and variations of GVD along the fiber length. The resulting equation becomes

$$i\frac{\partial u}{\partial \xi} + \frac{1}{2}p(\xi)\frac{\partial^2 u}{\partial \tau^2} + |u|^2 u = -\frac{i}{2}\Gamma u, \tag{9.18}$$

where $p(\xi) = |\beta_2(\xi)/\beta_2(0)|$ is the normalized GVD at ξ. The distance ξ is normalized using the dispersion length $L_D = T_0^2/|\beta_2(0)|$.

Because of the ξ dependence of the second term, Eq. (9.18) is no longer a standard NSE. However, it can be transformed into a perturbed NSE by using

$$v = p^{-1/2}u, \qquad \xi' = \int_0^{\xi} p(\xi)\, d\xi. \tag{9.19}$$

These transformations renormalize the soliton amplitude and the distance scale to the local GVD. In terms of v and ξ', Eq. (9.18) becomes

$$i\frac{\partial v}{\partial \xi'} + \frac{1}{2}\frac{\partial^2 v}{\partial \tau^2} + |v|^2 v = -i\left(\frac{\Gamma}{2p} + \frac{1}{2p}\frac{dp}{d\xi'}\right)v. \tag{9.20}$$

If the GVD profile is chosen such that $dp/d\xi' = -\Gamma$, or $p(\xi) = \exp(-\Gamma\xi)$, the terms on the right side of Eq. (9.20) vanish, and the fiber loss has no effect on soliton width or shape [17]. Thus, solitons can maintain the

balance between GVD and SPM even in a lossy fiber if the GVD decreases exponentially as $|\beta_2(z)| = |\beta_2(0)|\exp(-\alpha z)$ along the fiber length. Such fibers are called dispersion-decreasing fibers and have been used to verify the above prediction in several recent experiments [18]–[20].

Since fibers with a continuously varying dispersion are not yet commercially available, a practical approach consists of splicing together several constant-dispersion fibers with different β_2 values. A common dispersion map alternates fiber segments with opposite GVD values in an attempt to reduce the average dispersion along the communication link. The question then arises whether solitons can survive rapid variations in β_2. This issue has been studied extensively since 1996 [21]–[27], and a new class of solitons, referred to as dispersion-managed solitons, has been discovered. This discovery has jolted the field of nonlinear fiber optics since it is forcing us to modify some long-held notions about solitons.

Dispersion-managed solitons are not true solitons in the usual sense since the underlying NSE, Eq. (9.20), is not integrable by the inverse scattering method. Because of large variations in the local GVD, the dispersive and nonlinear effects cannot be balance locally, resulting in large variations in shape, width, and frequency chirp of the transmitted pulse. Nevertheless, they can be balanced in a global sense since the optical pulse evolves in a periodic manner and require considerably higher input powers [21]. It was initially thought that such a periodic evolution can occur only if the average GVD is anomalous [23]. However, it was discovered in 1997 that dispersion-managed solitons can exist even when the average GVD over a dispersion map is normal [24]. The research on dispersion-managed solitons is continuing, and further advances are likely to occur in this field.

5 Nonlinear optical switching

Optical (or photonic) switching refers to a phenomenon in which transmission of an optical field through a device is switched among two or more possible states by optical means. In the case of nonlinear optical switching the device transmission is intensity dependent so that the optical beam itself induces switching depending on its intensity. Both SPM and XPM occurring in an optical fiber are well suited for this application and their use results in an all-fiber device capable of switching on a femtosecond time scale.

5.1 SPM-based optical switching

The SPM-based optical switching is realized in practice by using an interferometer. Any interferometer can be used for this purpose including Fabry–Perot, Michelson, and Mach–Zehnder interferometers. A Sagnac in-

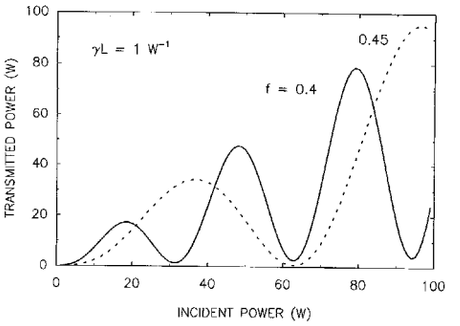

FIGURE 4. Transmitted power as a function of incident power for a nonlinear fiber-loop mirror.

terferometer is commonly used since it can be constructed by using a fiber coupler whose two output ports are connected to a fiber segment, forming a loop. The input pulse enters from one port and is split into two counterpropagating pulses at the fiber coupler, which interfere at the coupler after acquiring a relative phase shift during one round trip through the loop. The input pulse is transmitted through the other port or is reflected back to the input port depending on the relative phase shift. Such a device acts as a perfect mirror at low powers (the linear regime) when the fiber coupler splits the pulse equally (a 50:50 coupler) and is referred to as a fiber-loop mirror. However, if the coupler splits the pulse unequally, the same device acts as an all-optical switch because of the SPM-induced nonlinear phase shift and is referred to as the nonlinear optical loop mirror (NOLM) [28]. Such a device has attracted considerable attention and has found applications not only for optical switching but also for mode locking and wavelength demultiplexing.

Switching characteristics of a NOLM depend on the splitting ratio of the fiber coupler. If a fraction f of the input power P_0 travels in the clockwise direction, then the transmissivity for a loop of length L is obtained by calculating the phase shifts acquired during a round trip by the counterpropagating optical waves, and then recombining them interferometrically at the coupler. The result is [28]

$$T_m = 1 - 2f(1-f)\{1 + \cos[(1-2f)\gamma P_0 L]\}. \tag{9.21}$$

For $f = 0.5$ the loop reflectivity is 100% for all powers. However, if the splitting fraction f is different than 0.5, then the NOLM can act as a switch.

Figure 4 shows the transmitted power as a function of P_0 for two values of f. At low powers, little light is transmitted if f is close to 0.5 since $T_m \approx 1 - 4f(1-f)$. At high powers, the nonlinear phase shift leads to complete transmission whenever $|1-2f|\gamma P_0 L = (2m-1)\pi$, where m is an

integer. As seen in Fig. 3, the NOLM switches from low to high transmission periodically as input power increases. In practice, only the first transmission peak ($m = 1$) is used for switching because it requires the least power. In fact, switching power is rather high even for $m = 1$. For this reason, experiments on nonlinear optical switching typically use ultrashort optical pulses to obtain high peak powers. However, only the central part of the pulse is generally intense enough to experience switching, leading to pulse distortion. This problem can be solved by using optical solitons as input pulses since solitons switch as a whole entity because of their particle-like nature [29]–[31].

The switching threshold of an NOLM can be reduced considerably by incorporating an optical amplifier within the loop [32]. If the amplifier is located close to the fiber coupler, its presence introduces an asymmetry such that even a 50:50 coupler ($f = 0.5$) can be employed. This feature can be understood by noting that one wave is amplified at the entrance to the loop while the counterpropagating wave experiences amplification just before exiting the loop. Since the intensities of the two waves differ by a large amount throughout the loop, the differential phase shift can be quite large. In fact, assuming that the clockwise wave is amplified first by a factor G, Eq. (9.21)) becomes

$$T_m = 1 - 2f(1-f)\{1 + \cos[(1 - f - Gf)\gamma P_0 L]\}. \tag{9.22}$$

For $f = 0.5$ the switching power for $m = 1$ becomes $P_0 = 2\pi/[(G-1)\gamma L]$. Since the amplification factor G is typically 30 dB, the switching power is reduced by a factor of 1000. Such a device, referred to as the nonlinear amplifying loop mirror [32], provides switching with gain and can switch at power levels below 1 mW. Indeed, switching at a power of less than 250 μW has been demonstrated for a 17-m fiber loop [33].

The NOLM has found many applications. It can be used for pulse shaping because of its intensity-dependent transmission. For example, if a short optical pulse contains a broad pedestal, the pedestal can be removed by passing it through such a device. Its use for passive mode locking permits generation of femtosecond pulses in figure-eight fiber lasers [34]. Another important application of NOLM is for demultiplexing of individual channels in a WDM lightwave system. Since demultiplexing requires injection of a control pulse together with the signal, the XPM, rather than SPM, is the main nonlinear effect behind the operation of such devices. We turn to it in the next subsection.

5.2 XPM-based optical switching

The physics behind XPM-induced switching can be understood by considering a generic interferometer designed such that a weak signal pulse, divided equally between its two arms, experiences identical phase shifts in

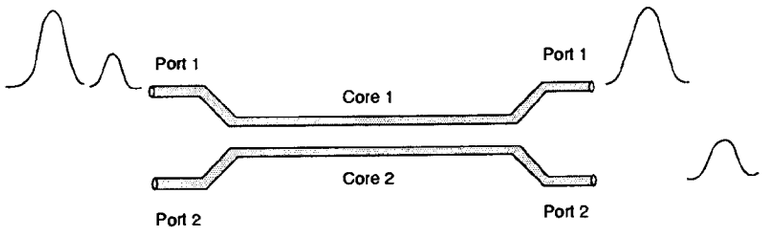

FIGURE 5. Schematic illustration of intensity-dependent switching in a dual-core fiber coupler.

each arm and is transmitted through constructive interference. If a pump pulse at a different wavelength is injected into one of the arms of the interferometer, it would change the signal phase through XPM in that arm. If the XPM-induced phase shift is large enough (close to π), the signal pulse will not be transmitted because of the destructive interference occurring at the output. Thus, an intense pump pulse can switch the signal pulse through the XPM-induced phase shift.

XPM-induced optical switching by using a NOLM was demonstrated in 1990 [35]. A dichroic fiber coupler with 50:50 splitting ratio at 1.53 μm and 100:0 splitting ratio at 1.3 μm was used to allow for the dual-wavelength operation. A 1.53-μm color-center laser provided low-power ($\sim$ 5 mW) CW signal. As discussed before, the counterpropagating signal beams experienced identical phase shifts, and the 500-m-long fiber loop acted as a perfect mirror in the absence of a pump beam. When 130-ps pump pulses, obtained from a 1.3-μm Nd:YAG laser, were injected in the clockwise direction, the XPM interaction between the pump and the signal introduced a phase difference between the counterpropagating signal beams. Most of the signal power was transmitted when the peak power of the pump pulse was large enough to introduce a π phase shift.

Several other devices utilize XPM for optical switching, the most studied device perhaps being a nonlinear directional coupler [36]–[38]. Nonlinear directional couplers are four-port devices in which an optical beam can be switched from one port to another by changing the input power. Such couplers can be made by using dual-core fibers whose two cores are close enough that the evanescent-wave coupling between the optical modes associated with each core transfers power from one core to another. SPM and XPM modify this coupling and affect the switching behavior. Figure 5 shows schematically how an optical pulse can be directed toward different ports depending on its peak power. Ultrafast all-optical switching in dual-core fibers was observed in 1987. Since then, considerable attention has focused on the performance of fiber directional couplers. Since the use of

solitons permits switching of the entire pulse, soliton switching in dual-core fibers has attracted the most attention.

XPM-induced optical switching can also be realized by using birefringent effects in single-core fibers. In fact, its use has led to the realization of all-optical, cascadable, ultrafast, logic gates [39]. In this case, switching occurs between the orthogonally polarized modes of the same fiber core. In fact, optical switching is described mathematically in both cases by the same set of coupled NSEs [5]

$$i\frac{\partial u}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 u}{\partial \tau^2} + (|u|^2 + B|v|^2)u + \kappa v = 0, \tag{9.23}$$

$$i\frac{\partial v}{\partial \xi} \pm \frac{1}{2}\frac{\partial^2 v}{\partial \tau^2} + (|v|^2 + B|u|^2)v + \kappa u = 0, \tag{9.24}$$

where $B = 2$ for a twin-core directional coupler but reduces to $2/3$ in single-core birefringent fibers. The above equations have been extensively studied [38] not only in the context of optical switching but also other nonlinear effects such as XPM-induced modulation stability and pulse compression. These equations also have solutions in the form of pairs of bright and dark solitons supported mutually through the XPM interaction. We refer the reader to recent books for further details on these topics [5], [38].

6 Concluding remarks

At the dawn of the 21st century, the field of nonlinear fiber optics is vibrant with a lot of research activity. This chapter has touched only the tip of an iceberg by focusing on two themes related to solitons and optical switching. Stimulated Raman scattering is attracting a lot of attention for making broadband optical amplifiers. Similarly, FWM has been used for channel demultiplexing in a WDM lightwave system. FWM is also harmful for WDM systems since it leads to interchannel crosstalk, and the dispersion-management technique is often used to suppress it. Another topic that is attractive considerable attention is the study of nonlinear effects in fiber gratings [40]. Fiber gratings exhibit modulation instability and support a new type of solitons referred to as Bragg solitons. From the wide range of nonlinear effects in optical fibers currently being studied, it is expected that nonlinear fiber optics will remain a topic of interest well into the 21st century.

7 References

[1] P. Franken, A. E. Hill, C. W. Peters, and G. Weinrich, *Phys. Rev. Lett.* **7**, 118 (1961).

[2] F. P. Kapron, D. B. Keck, and R. D. Maurer, *Appl. Phys. Lett.* **17**, 423 (1970).

[3] R. H. Stolen, E. P. Ippen, and A. R. Tynes, *Appl. Phys. Lett.* **20**, 62 (1972); E. P. Ippen and R. H. Stolen, *Appl. Phys. Lett.* **21**, 539 (1972).

[4] R. H. Stolen, *Proc. IEEE* **68**, 1232 (1980).

[5] G. P. Agrawal, *Nonlinear Fiber Optics* (Academic Press, San Diego, CA, 1989); 2nd ed., 1995; Chinese translation, 1992; Russian translation, 1996; Japanese translation, 1997.

[6] G. P. Agrawal, *Fiber-Optic Communication Systems*, 2nd ed. (Wiley, New York, 1997).

[7] G. P. Agrawal, in *Properties of Glass and Rare-Earth Doped Glasses for Optical Fibers*, D. Hewak, Ed. (IEE, Stevenage, U.K., 1998).

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

[9] W. Zhao and E. Bourkoff, *J. Opt. Soc. Am. B* **9**, 1134 (1992).

[10] Y. S. Kivshar and X. Yang, *Phys. Rev. E* **49**, 1657 (1994).

[11] D. J. Richardson, R. P. Chamberlin, L. Dong, and D. N. Payne, *Electron. Lett.* **30**, 1326 (1994).

[12] M. Nakazawa and K. Suzuki, *Electron. Lett.* **31**, 1076 (1995); *Electron. Lett.* **31**, 1084 (1995).

[13] Y. S. Kivshar and B. Luther-Davies, *Phys. Rep.* **298**, 81 (1998).

[14] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, Clarendon Press, Oxford, 1995.

[15] H. Kubota and M. Nakazawa, *Opt. Commun.* **87**, 15 (1992).

[16] M. Nakazawa and H. Kubota, *Electron. Lett.* **31**, 216 (1995).

[17] K. Tajima, *Opt. Lett.* **12**, 54 (1987).

[18] D. J. Richardson, R. P. Chamberlin, L. Dong, and D. N. Payne, *Electron. Lett.* **31**, 1681 (1995).

[19] A. J. Stentz, R. Boyd, and A. F. Evans, *Opt. Lett.* **20**, 1770 (1995).

[20] D. J. Richardson, L. Dong, R. P. Chamberlin, A. D. Ellis, T. Widdowson, and W. A. Pender, *Electron. Lett.* **32**, 373 (1996).

[21] N. J. Smith, F. M. Knox, N. J. Doran, K. J. Blow, I. Bennion, *Electron. Lett.* **32**, 54 (1996).

[22] I. R. Gabitov, E. G. Shapiro, and S. K. Turitsyn, *Phys. Rev. E* **55**, 3624 (1997); *Opt. Commun.* **134**, 317 (1997).

[23] A. Hasegawa, Y. Kodama, and A. Maruta, *Opt. Fiber Technol.* **3**, 197 (1997).

[24] J. H. B. Nijhof, N. J. Doran, W. Forysiak, and F. M. Knox, *Electron. Lett.* **33**, 1726 (1997).

[25] T. Georges, F. Favre, and D. L. Guen, *IEICE Tran. Electron.* **E81-C**, 226 (1998).

[26] G. M. Carter and J. M. Jacob, *IEEE Photon. Technol. Lett.* **10**, 546 (1998).

[27] K. N. Kutz and P. K. A. Wai *Electron. Lett.* **34**, 522 (1998); *IEEE Photon. Technol. Lett.* **10**, 702 (1998).

[28] N. J. Doran and D. Wood, *Opt. Lett.* **13**, 56 (1988).

[29] M. C. Farries and D. N. Payne, *Appl. Phys. Lett.* **55**, 25 (1989).

[30] K. J. Blow, N. J. Doran, and B. K. Nayar, *Opt. Lett.* **14**, 754 (1989).

[31] M. N. Islam, E. R. Sunderman, R. H. Stolen, W. Pleibel, and J. R. Simpson, *Opt. Lett.* **14**, 811 (1989).

[32] M. E. Fermann, F. Haberl, M. Hofer, and H. Hochstrasser, *Opt. Lett.* **15**, 752 (1990).

[33] A. W. O'Neill and R. P. Webb, *Electron. Lett.* **26**, 2008 (1990).

[34] I. N. Duling III, *Compact Sources of Ultrashort Pulses* (Cambrige University Press, New York, 1995).

[35] K. J. Blow, N. J. Doran, B. K. Nayar, and B. P. Nelson, *Opt. Lett.* **15**, 248 (1990).

[36] G. I. Stegeman and A. Miller, in *Photonics in Switching*, Vol. 1, J. E. Midwinter, Ed. (Academic Press, San Diego, CA, 1993)

[37] I. M. Uzunov, R. Muschall, M. Gölles, Y. S. Kivshar, B. A. Malomed, and F. Lederer, *Phys. Rev. E* **51**, 2527 (1995).

[38] N. N. Akhmediev and A. Ankiewicz, *Solitons: Nonlinear Pulses and Beams* (Chapman and Hall, New York, 1997).

[39] M. N. Islam, *Ultrafast Fiber Switching Devices and Systems* (Cambridge University Press, New York, 1992).

[40] B. J. Eggleton, C. M. de Sterke, and R. E. Slusher, *J. Opt. Soc. Am. B* **14**, 2980 (1997).

10

Self-Focusing and Collapse of Light Beams in Nonlinear Dispersive Media

Luc Bergé
Jens Juul Rasmussen

ABSTRACT The collapse of self-focusing beams in nonlinear dispersive media described by the nonlinear Schrödinger (NLS) equation is reviewed. Conditions for blow-up of solutions to the NLS equation with a cubic nonlinearity and isotropic dispersion properties are recalled, together with the self-similar analyses employed for modelling wave collapses. Emphasis is then laid on the influence of anisotropic (negative) dispersion and on the deviations from the spatio-temporal envelope approximations, which are shown to strongly alter the blow-up dynamics.

1 Introduction

The self-focusing and collapse of wave-packets in nonlinear dispersive media as, e.g., plasmas and nonlinear optical materials, is in general described by the nonlinear Schrödinger (NLS) equation for a scalar wave envelope $E(x, y, z, t)$. In normalized form this equation reads

$$i\frac{\partial E}{\partial t} + \Delta_\perp E + s\frac{\partial^2 E}{\partial z^2} + |E|^2 E = 0. \tag{10.1}$$

The different signs of the coefficient s allow for treating media with isotropic dispersion ($s > 0$) as well as with anisotropic dispersion ($s < 0$). In Eq.(10.1) we have used standard notations and normalized variables. In the context of nonlinear optics, the variable t often refers to the longitudinal length along the beam propagation axis, whereas z occurring in the space derivatives of the envelope corresponds to a retarded time variable $t' = t - z/\omega'$ ($\omega' \equiv \partial\omega/\partial k$). These derivatives then reflect the "temporal" wave dispersion measured through the group velocity dispersion (GVD) coefficient s, related to the dispersion factor $\partial^2 k/\partial\omega^2|_{\omega_0}$. The coefficient s can be either positive in the case of a so-called *anomalous* dispersion, or negative in the opposite case of a so-called *normal* dispersion [1]. More generally, equation (10.1) applies to the description of nonlinear wave propagation in media with an anisotropic dispersion, i.e., with a dispersion law

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 213−228, 2000.

$\omega(\vec{k})$ for which the second-order derivatives with respect to $k_\perp$ and k_z differ in sign. This is often the case for waves in magnetized plasmas, such as lower-hybrid waves [2]. However, adopting the terminology used in nonlinear optics we henceforth refer to $s = +1$ and $s = -1$ as the anomalous and normal dispersion regimes, respectively. Before proceeding, let us recall that equation (10.1) conserves the Hamiltonian

$$H = \int \{|\vec{\nabla}_\perp E|^2 + s|E_z|^2 - \frac{1}{2}|E|^4\} d\vec{r}_\perp dz \equiv I_1 + sI_2 - I_3 \tag{10.2}$$

and the total "mass" or "power" $N \equiv \int |E|^2 d\vec{r}_\perp dz$. For $s = 0$ and when disregarding the z axis, it describes the 2D self-focusing/collapse of wave-packets at a finite time $t = t_0$. A necessary condition for self-focusing is $N > N_c \simeq 11.68$, where the critical mass N_c is the mass calculated on the stationary ground state of NLS [3]. This ground state is the positive, radially-symmetric solution ϕ_0 of

$$-\lambda\phi_0 + \Delta_\perp \phi_0 + \phi_0^3 = 0, \tag{10.3}$$

derived from (10.1) after setting $E(x, y, z, t) = \phi_0(x, y)e^{i\lambda t}$ and computed with $\lambda = 1$. A 2D self-focusing, leading to an ultimate collapse, is ensured for $H < 0$.

In this contribution, we summarize standard results concerning wave collapses in cubic media with anomalous GVD ($s = +1$), where we mainly focus on the self-similar collapse dynamics. A detailed discussion of the basic collapse phenomenology can be found in the recent reviews [4, 5, 6] and in the contribution to this volume by Kuznetsov and Zakharov [7], which contains a general account of collapse dynamics in different physical systems. Here, we describe recent results on the collapse dynamics in the normal GVD regime ($s = -1$), providing evidence for the absence of 3D collapse together with the formation of splitting events preventing a 2D transverse self-similar collapse. The influence of deviations to the envelope approximations, which are usually applied for deriving Eq.(10.1), is also discussed.

2 General properties of self-focusing with anomalous group velocity dispersion

2.1 Basic properties

With anomalous GVD, the dispersion coefficient s can be set equal to unity, so that we simply deal with the so-called scaled NLS model

$$i\partial_t E + \vec{\nabla}^2 E + |E|^2 E = 0 \tag{10.4}$$

involving the Laplacian $\vec{\nabla}^2 = \Delta_\perp + \partial_z^2$. We first briefly recall basic definitions concerning the local and global existence of time-dependent solutions to Eq.(10.4), regarding its initial-value (Cauchy) problem. The initial wave function $E(\vec{x}, t = 0) \equiv E_0(\vec{x})$ is formulated in a Sobolev space usually selected among the possible Hilbert spaces $H^p = \{v : (1-\Delta)^{p/2} v \in L^2(\mathbf{R}^D)\}$, where D denotes the total space dimension number $D = D_\perp + 1$ attached to the vector $\vec{x} = (\vec{r}_\perp, z)$. By convention, the Sobolev space H^1 is currently chosen with norm

$$\|E\|_{H^1} = (\|E\|_2^2 + \|\vec{\nabla} E\|_2^2)^{1/2}, \tag{10.5}$$

where both E and $\vec{\nabla} E$ belong to $L^2(\mathbf{R}^D)$. Here and in the sequel, the use of the standard L^p norms

$$\|f\|_p \equiv (\int |f|^p d\vec{r}_\perp dz)^{1/p} \tag{10.6}$$

will sometimes be made for notational convenience. For the NLS equation, the L^2 norm of the solution $E(\vec{x}, t)$ is thus conserved with $N \equiv \|E\|_2^2 = \|E_0\|_2^2$. Keeping this property in mind, we define a blow-up by the existence of a finite time $t_0 < +\infty$, such that $\|E\|_{H^1} \to +\infty$ for $t \to t_0$, which implies that the gradient norm $\|\vec{\nabla} E\|_2$ diverges as $t \to t_0$. From this, it can be shown that $\max|E(\vec{x}, t)|$ also diverges as $t \to t_0$. To illustrate a blow-up proof, we employ the arguments emphasized in [8, 9], which are based on the vanishing of the so-called virial integral

$$I(t) \equiv \int (r_\perp^2 + z^2)|E|^2 d\vec{r}_\perp dz = \|\vec{x} E\|_2^2 \to 0 \ \text{ as } \ t \to t_0^* < +\infty. \tag{10.7}$$

The evolution of $I(t)$ is governed by

$$\partial_t^2 I(t) = 4\{2H + (1 - \frac{D}{2}) \int |E|^4 d\vec{r}_\perp dz\} \tag{10.8}$$

and the "virial" integral $I(t)$ is a measure of the square-width of the wave-packet. Valid for localized time-dependent solutions with a center of mass located on the origin and possessing zero velocity, this relation enables us to determine whether a given initial waveform will collapse into a point-singularity in a finite time, such that the wave amplitude blows up at this point. Let us assume *a priori* that $I(t)$ vanishes exactly at $t = t_0^*$. By integrating the L^2 norm of E by parts and using the Cauchy-Schwarz inequality, we get

$$\|E\|_2^2 \leq \frac{2}{D} \|\vec{\nabla} E\|_2 \|\vec{x} E\|_2. \tag{10.9}$$

Considering non-trivial solutions with a finite norm $\|E_0\|_2 \neq 0$, we then deduce from Eqs.(10.7) and (10.9) that, as long as the solution exists for all $t < t_0^*$, the gradient norm $\|\vec{\nabla} E\|_2$ tends to infinity for $t \to t_0^*$ and blow-up is demonstrated. Note that the instant t_0^* can differ from the moment

t_0 at which the solution diverges. Indeed, the blow-up time t_0 has been revealed to be smaller than the time t_0^* when $I(t)$ strictly reaches zero, because only a limited amount of mass is captured within the blowing-up spike as $t \to t_0$. For the 2D NLS, this amount is just the critical mass $N_c \equiv \|\phi_0\|_2^2$ defined by the ground state solution, whereas, for a given wave having initially $N >> N_c$, the remaining part of the mass is evacuated through radiations. Hence, t_0^* must be viewed as a maximum existence time in case of blow-up. From Eq.(10.8), it is seen right away that $H < 0$ is a sufficient condition for collapse of 3D beams with anomalous GVD. Also in the 3D case the stationary radially-symmetric ground state solution plays an essential role in providing a sharper criterion for collapse [10]. This criterion is $H < H_{sol} = N_c^2/N$, where H_{sol} refers to the Hamiltonian integral computed on the soliton solution, that corresponds to a given initial mass N. As before, N_c refers to the mass of the ground state solution to Eq.(10.1), i.e., the solution of (10.3) for $\lambda = 1$, where $\Delta_\perp$ is replaced by the 3D Laplacian.

2.2 Self-similar wave collapses

A current solution modelling the self-focusing part of a collapsing field is the so-called "self-similar" solution. Written in its simplest form, the amplitude evolves as

$$|E(\vec{x},t)| = \frac{1}{J(t)}|Q(\frac{\vec{x}}{a(t)})|, \tag{10.10}$$

where the functions $J(t)$ and $a(t)$ both decrease to zero, as t goes to the singularity time t_0. The amplitude function $|Q|$ is called "self-similar", because it keeps the same spatial distribution at every time preceding t_0. For the sake of simplicity, we limit our investigation to radially-symmetric scalar solutions $E(r \equiv |\vec{x}|, t)$ to the NLS equation (10.4). Blowing-up solutions for NLS can then be sought under the generic form

$$E(r,t) = \frac{1}{[g(\tau)]^\alpha} Q(\xi,\tau) \exp(i\lambda\tau) \tag{10.11}$$

where $\xi \equiv r/a(t) \equiv r/g(\tau)$ is the spatial coordinate rescaled with respect to the radius $a(t)$, which is re-expressed as a τ-dependent function $g(\tau)$. The new time $\tau(t) \equiv \int_0^t du/a^2(u)$ varies continuously with t. The parameter λ denotes the positive eigenvalue attached to the NLS equation [see Eq.(10.3)], and it assures localized eigenstates Q vanishing at infinity. In Eq.(10.11), the exponent α is positive, which allows us to describe the divergence of the wave when $a(t) \to 0$. By definition, the evolution of $E(r,t)$ is *exactly* self-similar, if, in the limit $g(\tau) \to 0$, the following requirement holds

$$Q(\xi,\tau) \to Q_0(\xi;\lambda) \text{ as } t \to t_0 \text{ [or } \tau(t) \to \tau(t_0) \equiv \tau_0]. \tag{10.12}$$

$Q_0(\xi;\lambda)$ is an exactly self-similar function in the sense that, apart from its functional dependence on $\lambda > 0$, it does not explicitly depend on the new time variable τ, but on the rescaled variable ξ only. Under this assumption, we can express the invariants N and H associated with the NLS equation in terms of Q [Eq.(10.11)]

$$N\{E\} = [g(\tau)]^{D-2\alpha} N\{Q\}, \;\; N\{Q\} \equiv \|Q\|_2^2 = 2^{D-1}\pi \int_0^{+\infty} |Q|^2 \xi^{D-1} d\xi, \tag{10.13}$$

$$H\{E\} = \{\frac{\|\partial_\xi Q\|_2^2}{[g(\tau)]^2} - \frac{1}{2}\frac{\|Q\|_4^4}{[g(\tau)]^{2\alpha}}\} \times [g(\tau)]^{D-2\alpha}. \tag{10.14}$$

In accordance with the terminology introduced in Ref.[11], we use the term "strong" self-similar solutions for the singular states (10.11) that blow up with an exactly self-similar shape preserving the mass integral $N\{E\}$ with $N\{Q\} = N\{Q_0\}$. In this case, the exponent α in the amplitude factor of solution (10.11) is equal to $\alpha = D/2$, and it particularly suits a critical ($D = 2$) collapse for which this choice is compatible with both conservations of N and H. Strong solutions are expected to capture the mass N around the origin $\xi = 0$ where the waveform diverges. On the contrary, so-called "weak" solutions correspond to self-similar states which cannot preserve these integrals of motion, in such a way that they cannot be valid within the entire space in ξ. This concept applies to a supercritical ($D > 2$) collapse for which the constancy of H near the collapse time t_0 implies that the exponent α in Eq.(10.11) must be equal to $\alpha = 1$. For "weak" solutions, the function Q cannot be exactly self-similar in the whole space domain, since the integral $N\{Q\}$ has to restore the mass conservation with a dominant contribution located around some spatial bound $\xi_{max}(\tau)$, such that $N\{Q\} \simeq 2^{D-1}\pi \int_0^{\xi_{max}(\tau)} |Q_0|^2 \xi^{D-1} d\xi \sim [g(\tau)]^{2-D}$ diverges as $\tau \to \tau_0$. These solutions describe so-called weak collapses, in the sense that the mass is shifted to some points $\sim \xi_{max}(\tau)$ different from the origin, while the singular waveform still collapses on the center $\xi = 0$ [11]. Formally, the divergence of $N\{Q\}$ occurring in the self-similar limit $i\partial_\tau Q \to 0$ can be demonstrated from the equation for Q. The functions Q are moreover zero-energy states which preserve the invariance of $H\{E\}$.

Fixing henceforth $\alpha = 1$, we describe the self-similarity of collapsing solutions by means of a general analysis, which enables us to determine their precise blow-up rates $a(t)$. We introduce the substitution (10.11) into Eq.(10.4) and use the additional transformation

$$Q(\xi,\tau) = \phi(\xi,\tau)\exp(-i\beta\xi^2/4), \tag{10.15}$$

where β is a time-dependent function

$$\beta(t) \equiv -\dot{a}a = \partial_\tau \ln(\frac{1}{g(\tau)}) \tag{10.16}$$

(dot notation means a differentiation with respect to the time variable t). By doing so, we obtain the self-similarly transformed equation of NLS

$$i\partial_\tau\phi + \Delta_\xi\phi + |\phi|^2\phi + \epsilon(\tau)[\xi^2 - \xi_T^2(\tau)]\phi = 0, \tag{10.17}$$

with $\Delta_\xi \equiv \xi^{1-D}\partial_\xi\xi^{D-1}\partial_\xi$. Here, the quantities $\epsilon(\tau)$ and $\xi_T^2(\tau)$ are time-dependent functions

$$\epsilon(\tau) \equiv -\frac{1}{4}a^3\ddot{a} = \frac{\beta^2 + \beta_\tau}{4} \tag{10.18}$$

$$\xi_T^2(\tau) \equiv \frac{\lambda + i\beta(D-2)/2}{\epsilon}. \tag{10.19}$$

According to the definition of self-similarity, the new function ϕ is said to be exactly self-similar when it satisfies $\partial_\tau\phi = 0$ in the neighborhood of the collapse moment t_0, which means: $\partial_\tau\phi \to 0$ as $\tau(t) \to \tau(t_0) = \tau_0$. To determine the scaling law $a(t)$, solutions to equation (10.17) must be treated by means of a perturbation theory in the vicinity of the collapse singularity t_0. This perturbation theory assures that, as $t \to t_0$, the solution ϕ tends to a self-similar state satisfying the limit $\partial_\tau\phi \to 0$, while the time-dependent functions ϵ and $\xi_T^2(\epsilon)$ are expected to converge adiabatically towards steady-state values. In this limit, by assuming a priori $\tau(t) \to \tau_0 = +\infty$, the time-dependent functions ϵ and β can be viewed as adiabatic functions of time satisfying $\epsilon \simeq \beta^2/4$ with $|\partial_\tau\beta| << \beta^2$. It is then convenient to divide the solution ϕ into two distinct contributions, namely a core ϕ_c and a tail ϕ_T

$$\phi(\xi,\tau) = \phi_c(\xi,\tau) + \phi_T(\xi,\tau), \tag{10.20}$$

where the time dependence of ϕ is only brought by the adiabatic functions ϵ and β provided that the limit $\partial_\tau\phi \to 0$ holds. ϕ_c corresponds to the central part of ϕ located at the origin, around which the localizing nonlinearities in $|\phi|^2\phi$ are efficient, and it extends in the spatial domain: $\xi < \xi_T$. In this region, ϕ_c is close to the exact self-similar state $\phi_0(\xi;\lambda)$ and it can be expressed perturbatively as a Taylor expansion in power series of, e.g., $(\epsilon - \epsilon_0)$ with $\epsilon \to \epsilon_0$ as $\tau \to +\infty$

$$\phi_c(\xi,\epsilon) \simeq \phi_0(\xi;\lambda) + (\epsilon - \epsilon_0)\frac{\partial\phi}{\partial\epsilon}|_{\epsilon_0} + ... \tag{10.21}$$

On the other hand, ϕ_T extends in the domain $\xi > \xi_T$ associated with the asymptotic regions of space, where the cubic nonlinearity can be disregarded. ϕ_T is then a solution of a parabolic cylinder equation, which must be matched to ϕ_c in the vicinity of the complex turning point $\xi = \xi_T$. Around this point, the appropriate boundary conditions determine the right functional dependences of ϕ_T

$$\phi_T(\xi,\epsilon) \simeq C(\epsilon)\frac{e^{i\sqrt{\epsilon}\xi^2/2}}{\xi^{D/2+i\lambda/(2\sqrt{\epsilon})}}\Big[\frac{\xi}{|\xi_T|}\Big]^{(D-2)\frac{\beta}{4\sqrt{\epsilon}}} \tag{10.22}$$

$$|C(\epsilon)| \propto \epsilon^{-1/4} e^{G(\epsilon)/2} \tag{10.23}$$

$$G(\epsilon) = -\frac{\pi\lambda}{2\sqrt{\epsilon}} + \frac{\beta}{4\sqrt{\epsilon}}(D-2)(1+2\ln 2) - \frac{\lambda}{2\sqrt{\epsilon}}\arctan[\frac{\beta(D-2)}{2\lambda}]. \tag{10.24}$$

Now we face another problem: due to the weak algebraic decay of ϕ_T (10.22), the mass integral $N\{\phi\}$ exhibits spatial divergence. This is removed by considering that self-similar solutions necessarily have a finite spatial extension for all times preceding t_0. The cut-off radius bounding this extension domain has been determined in [12, 13] and reads

$$\xi_{max}(\tau) \equiv \frac{A|\xi_T|}{g(\tau)}, \quad A = const >> 1. \tag{10.25}$$

The final step of the analysis consists in identifying accurately the contraction rate $g(\tau) = a(t)$, by using the perturbed state (10.20). To this aim, we use the mass continuity equation for the solution ϕ, namely

$$\int_0^\xi \partial_\tau |\phi(\xi',\tau)|^2 {\xi'}^{D-1} d\xi' = -2|\phi(\xi,\tau)|^2 \xi^{D-1} \partial_\xi \arg\{\phi(\xi,\tau)\} + \beta(D-2)\int_0^\xi |\phi(\xi',\tau)|^2 {\xi'}^{D-1} d\xi', \tag{10.26}$$

which follows from multiplying Eq.(10.17) by ϕ^* and integrating the resulting imaginary part from zero to ξ. Multiplied by the numerical factor $\mathcal{R} = 2^{D-1}\pi$, Eq.(10.26) actually describes the mass exchanges between the core and the tail contribution of ϕ by choosing ξ in the range $\xi >> \xi_T$. By inserting the solution (10.20) into Eq.(10.26) and using the Taylor expansion (10.21), we readily get

$$\mathcal{J}\partial_\tau \epsilon = -2\mathcal{R} e^{G(\epsilon)} [\frac{\xi}{|\xi_T|}]^{(D-2)\frac{\beta}{2\sqrt{\epsilon}}}_{\xi\to+\infty} + \beta(D-2)N\{\phi\} \tag{10.27}$$

$$\epsilon \simeq \beta^2/4, \quad \mathcal{J} \equiv 2\mathrm{Re}\int \{\phi_0^* \frac{\partial\phi}{\partial\epsilon}|_{\epsilon_0}\} d\vec{\xi} > 0. \tag{10.28}$$

We now specify the right-hand side of Eq.(10.27) and determine the blow-up scales of beams collapsing in isotropic media for space dimensions 2 and 3.

(i) *At the critical dimension* $D = 2$, equation (10.27) simplifies into

$$\partial_\tau \epsilon \simeq -\frac{2\mathcal{R}}{\mathcal{J}} \exp(-\frac{\pi\lambda}{2\sqrt{\epsilon}}), \tag{10.29}$$

from which the approximated behavior: $2\sqrt{\epsilon} \simeq \beta = \pi\lambda/\ln(\tau)$ follows. This yields the contraction scale

$$a(t) = a_0 \sqrt{\frac{2\pi\lambda(t_0-t)}{\ln(\ln[\frac{1}{t_0-t}])}} \tag{10.30}$$

after solving $\beta(t) = -a\dot{a}$ and employing $\tau(t) \simeq \ln[1/(t_0 - t)]$. This result agrees with Fraiman [14] and Malkin [15] when $\lambda = 1$, and it was accurately verified in direct numerical solutions of the NLS equation [16, 17]. Critical collapses are strong collapses that capture a finite amount of mass $N\{E\} = N\{\phi\}$, relaxing ultimately to the mass of the ground state solution N_c on the center. This has been confirmed by numerical solutions of the NLS equation with a weak nonlinear dissipation term acting on the collapsing spike [18, 19]. In this context, it was observed that the mass decreased in steps of $0.2 - 0.3 \times N_c$ during each collapse event.

(ii) *At the supercritical dimension* $D > 2$, Eq.(10.27) must be rewritten by using $\epsilon \to \beta^2/4$, $N\{E\} = [g(\tau)]^{D-2} N\{\phi\}$, and by taking into account that a self-similar solution cannot exceed the cut-off radius (10.25). This leads to the relation

$$\mathcal{J}\partial_\tau \epsilon \simeq [g(\tau)]^{2-D}[2\sqrt{\epsilon}N\{E\}(D-2) - 2\mathcal{R}e^{G(\epsilon)}], \tag{10.31}$$

from which a stationary fixed point $\epsilon \to \epsilon_0 > 0$ arises, corresponding to an exact self-similarity reached for $\epsilon_\tau = 0$. A self-similar state is then rapidly attained and it is characterized by the limit $\epsilon \to \epsilon_0 > 0$ ($\beta \to \beta_0 > 0$), which yields the contraction law

$$a(t) = a_0\sqrt{t_0 - t}. \tag{10.32}$$

Weak collapses driven by the scaling law (10.32) were numerically observed for a large class of initial data in Refs.[17, 18, 19, 20]. As a conclusion, 3D beams in cubic media with anomalous GVD achieve their self-focusing evolution with a weak collapse, which is characterized by the contraction scale (10.32) and by a mass integral $N\{E\} = a(t)N\{\phi\}$ becoming zero on the center. In accordance with this it was observed in numerical solutions of the 3D NLS with nonlinear dissipation that the total mass decreases monotonically [18], in contrast to the step-wise decrease observed in strong 2D collapses. Further discussions on weak, strong and superstrong collapse scenarios can be found in Ref. [7].

3 Self-focusing with normal group velocity dispersion

For media with normal dispersion described by Eq.(10.1) with $s = -1$, antecedent studies [21, 22, 23, 24, 25] indicate that the evolution of a wave pulse is governed by the interplay between two main tendencies, namely a 2D compression in the transverse plane and a stretching along the z axis. This competition leads to the splitting of a single pulse into small-scale bunches, accompanied by a simultaneous compression of the resulting structures. The latter can, in turn, be broken up into smaller-scale cells [25].

To describe this splitting phenomenon we suppose *a priori* that the solution E exists at least locally in time. From Eq.(10.1) it is easy to construct the longitudinal and transverse mean square radius of the pulse, associated respectively with the virial-type equations

$$\partial_t^2 I_z(t) \equiv \partial_t^2 \|zE\|_2^2 = 8\|\partial_z E\|_2^2 + 2\|E\|_4^4 = 8I_2 + 4I_3 \tag{10.33}$$

$$\partial_t^2 I_\perp(t) \equiv \partial_t^2 \|r_\perp E\|_2^2 = 8\|\vec{\nabla}_\perp E\|_2^2 - 4\|E\|_4^4 = 8H + 8I_2. \tag{10.34}$$

After multiplying Eq.(10.33) by $I_z(t)$ and applying Eq.(10.9) along z, we derive the inequality

$$\partial_t^2 I_z^2/2 \geq I_z \ddot{I}_z \geq 2N^2, \tag{10.35}$$

which enables us to conclude that for any initial data, $I_z(t)$ can never vanish in finite time. Instead it globally increases as $t \to \infty$, indicating thereby an asymptotic stretching of the waveform along z. Also, $I_z(t)$ appears to be bounded from below: let us assume a compression regime along the z axis and multiply both sides of the inequality (10.35) by $\dot{I}_z < 0$, we then immediately get the following bound for the longitudinal wave extension

$$I_z(t) \geq I_z^{min} = I_z(0) \exp\left[-\frac{(\dot{I}_z(0))^2}{4N^2}\right] > 0. \tag{10.36}$$

Regarding now the evolution of $I_\perp(t)$, we can show that a transverse collapse defined by $I_\perp \to 0$ is impossible in the total compression regime for which both of the requirements $\dot{I}_z < 0$ and $\dot{I}_\perp < 0$ hold in the vicinity of a given finite instant t_0^*. Indeed, let us multiply Eqs.(10.33) and (10.34) by $\dot{I}_\perp < 0$ and $\dot{I}_z < 0$, respectively, within the time interval $[T_0, t_0^*[$ where the wave is assumed to self-contract. We then add the resulting equations to obtain the estimate

$$\frac{d}{dt}[\dot{I}_\perp \dot{I}_z] = 8\dot{I}_z(H + \|\partial_z E\|_2^2) + 4\dot{I}_\perp(-H + \|\vec{\nabla}_\perp E\|_2^2 + \|\partial_z E\|_2^2) \tag{10.37}$$

$$\leq \frac{d}{dt}(-4HI_\perp + 8HI_z + 2N^2 \ln I_z + 4N^2 \ln I_\perp),$$

leading to

$$\mathcal{E}(t) = \dot{I}_z \dot{I}_\perp + 4H(I_\perp - 2I_z) - 4N^2 \ln I_\perp - 2N^2 \ln I_z \leq \mathcal{E}(T_0) < +\infty. \tag{10.38}$$

Because $I_z(t)$ is bounded as $I_z^{min} \leq I_z(t) \leq I_z(T_0)$, the inequality (10.38) proves that $I_\perp(t)$ cannot vanish under the above assumptions. Thus, $I_\perp(t)$ always remains bounded from below in this forced compression regime. From a mathematical viewpoint, the non-vanishing of $I_\perp(t)$ is not a rigorous proof for the absence of a blow-up-type singularity, since, for instance, such singularities can occur *before* the complete vanishing of the virial integral as is the case for the NLS equations with $s = +1$. Nevertheless, the impossibility for $I_\perp(t)$ to reach zero shows, from a physical point of view,

that no local concentration of the peak intensity, in the usual NLS sense, is possible. In this respect, it is amazing that even 2D self-focusing cannot take place in the transverse plane with $I_\perp \to 0$ under the conditions that would be believed to favour it by forcing a compression of the wave radius along every direction with $\partial_t I_z < 0$ and $\partial_t I_\perp < 0$.

Therefore, the only possibility for a collapse to occur seems to be for the case $\partial_t I_z > 0$ and $\partial_t I_\perp < 0$, for which virial arguments have limited value for excluding the possibility of collapse. Alternatively, the splitting process that appears close to the self-focusing stage (i.e., in the limit $t \to t_0$) can be described by employing a quasi-self-similar analysis [26, 27] analogous to the one outlined in Sec. 2. This shows that a transverse collapse cannot take place within the self-similar approximation. Here we give a brief account of the main ingredients of this analysis. We consider the case where the longitudinal scale size $Z(t)$ is assumed to increase, while the transverse scale size $R(t)$ is still ensured to decrease in the vicinity of the instant t_0 (Z and R have obvious connections with I_z and $I_\perp$). The quasi-self-similar analysis is based on the following solution

$$E(t,\vec{r}_\perp,z) = \frac{\sqrt{N}}{R(t)\sqrt{Z(t)}}\tilde{\phi}(\tau(t),\xi,\zeta)\exp\left(i\lambda\tau(t) + i\frac{R_t R}{4}\xi^2 - i\frac{Z_t Z}{4}\zeta^2\right) \tag{10.39}$$

with the rescaled variables

$$\tau(t) \equiv \int_0^t \frac{du}{[R(u)]^2}\,, \quad \xi \equiv \frac{r_\perp}{R(t)}\,, \quad \zeta \equiv \frac{z}{Z(t)}\,.$$

The positive parameter λ corresponds to the eigenvalue attached to Eq.(10.1), and we assume a pulse self-focusing in the transverse plane with $R(t \to t_0) \to 0$, while $Z(t)$ reaches a finite value $Z(t_0) > 0$. Inserting Eq.(10.39) into Eq.(10.1) with $s = -1$ leads to a self-similarly transformed equation for $\tilde{\phi}$, involving the time-dependent functions

$$\epsilon_R = [a_R^2 + \partial_\tau(a_R)]/4 = -R^3 R_{tt}/4\,, \quad \text{and} \quad \gamma\epsilon_Z = -R^2 Z Z_{tt}/4, \tag{10.40}$$

with $a_R \equiv -R_t R$ and $\gamma(t) \equiv R^2(t)/Z^2(t)$. As $t \to t_0$, it can be shown that both of these functions tend towards values of comparable magnitude, $|\epsilon|$, in the self-similar limit $\partial_\tau\tilde{\phi} \to 0$ reached as $\tau(t) \to \tau(t_0) = +\infty$. After simple rescalings and using $\gamma(t) \to 0$, the eigenfunction $\tilde{\phi}$ is found to obey the quasi-2D equation

$$i\partial_\tau\tilde{\phi} + \partial_\xi^2\tilde{\phi} + (1/\xi)\partial_\xi\tilde{\phi} + [\epsilon(\xi^2+\zeta^2) - \lambda]\tilde{\phi} + |\tilde{\phi}|^2\tilde{\phi} = 0, \tag{10.41}$$

where the variable ζ plays the role of a *localizing* parameter, in the sense that in the self-similar limit $\epsilon \to \text{const} > 0$, the solution $\tilde{\phi}$ can only be localized along the ζ-axis for $\tilde{\lambda} \equiv \lambda - \epsilon\zeta^2 > 0$. Next, the mass continuity relation associated with Eq.(10.41) describes the power exchanges in the

transverse plane between the inner core of the solution defined in the range $\xi < \sqrt{\tilde{\lambda}/\epsilon}$ and its tail extending for $\xi > \sqrt{\tilde{\lambda}/\epsilon}$. This allows us to derive a dynamical equation for ϵ

$$K\partial_\tau \epsilon \quad = \quad -2\exp[-\pi(\lambda - \epsilon\zeta^2)/(2\sqrt{\epsilon})] \tag{10.42}$$

with $K = \int_0^{+\infty} \xi^3 |\tilde{\phi}(\xi, \epsilon = 0)|^2 d\xi$. When the longitudinal effects are ignored, i.e., for $\zeta = 0$, Eq.(10.42) restores the contraction rate (10.30) of a 2D critical collapse. However, taking the longitudinal dimension into account ($\zeta > 0$), we deduce from (10.42) that $R(t)$ varies explicitly with ζ. In the self-similar limit $\epsilon_R \to a_R^2/4$ with $\partial_\tau |a_R| << a_R^2$, it is found that a_R is positive in the *localizing* domain $\zeta < \zeta^* \equiv \sqrt{\lambda/\epsilon} \simeq 2\sqrt{\lambda}/|a_R|$, yielding a contraction rate close to (10.30), but negative in the *delocalizing* domain $\zeta > \zeta^*$ for which $R(t)$ is dispersing. The pulse thus self-contracts until the rescaled distance ζ^* only and $R(t)$ must reach a minimum value in the vicinity of $\zeta \sim \zeta^*$, where the beam intensity $|E|^2 = N|\tilde{\phi}|^2/(Z(t)R^2(t))$ attains a maximum. Hence, if a transverse self-focusing promoted a collapse, it would take place as $\zeta \to \zeta^*$, i.e., displaced from the center of the pulse $\zeta = 0$. However, in this limit, Eq.(10.42) would lead to $\epsilon_R \sim -2\tau < 0$ as $\tau \to +\infty$, forcing ultimately $R(t)$ to disperse, not to vanish. We therefore conclude that a self-similar transverse collapse should not be realized in normally dispersive media.

As the pulse tends to form a maximum as $t \to t_0$ at the focus $z \to z^*(t_0) = Z(t_0)\zeta^*$, two symmetric spikes of the field, located on $z = \pm z^*(t_0)$ respectively, must develop. This explains the occurrence of one splitting event. For incident beams containing a transverse power exceeding widely N_c, the beam may continue to split up. Multi-splitting then follows from repeating the above analysis on each cell produced by an antecedent splitting event and located on the new "origin", e.g., $z = z^*$. This process stops when the resulting cells have a transverse power below the self-focusing threshold N_c. Therefore, we can empirically estimate the number of cells produced near the point of maximal transverse compression by $\mathcal{N} = 2N_\perp(t_0)/N_c \simeq N/(N_c Z(t_0))$, conjecturing that before splitting again, a given cell first self-focuses with a transverse power *at least* above critical. For anisotropic beams for which $Z(t_0)$ remains of the order of $Z(0)$ this number can directly be approximated in terms of the initial data: $\mathcal{N} = N/(N_c Z(0))$. Otherwise this latter estimate only provides an absolute upper limit for the number of cells, since the longitudinal extension of the beam, $Z(t)$, increases with time. A somewhat better estimate is then obtained by substituting $Z(0)$ by $Z(t_0)$, which is the length of the pulse at the time of maximum contraction [26]. Reasoning only in terms of the initial data, the *maximum* number of cells before their final spreading is evaluated by $\mathcal{N}_{\max} = 2N_\perp/N_c \simeq N/[N_c Z(0)]$. This estimate applied to Zharova *et al.*'s simulation [25] with $N_c \simeq 11.7$ fits the number of cells ($\mathcal{N} \simeq 8$) they observed. It is also consistent with the number of split

pulses revealed in the experiments performed with 90-fs laser pulses by Ranka et al [28] and by Diddams et al [29]: in these experiments, the nonlinear media were served by a BK7 glass window and fused-silica fibers, respectively. At least 4 cells were detected from pulses with peak power $P/P_c \sim N_\perp/N_c \simeq 4.22$ in [28] and 3-4 split pulses were observed from Gaussian pulses with $P/P_c \sim N_\perp/N_c \simeq 2.3$ in [29]. On the whole, these experimental data corroborate the above theoretical results. However, apart from the preliminary numerical simulations presented in Refs.[25] and [27] evidencing multi-splitting, very few numerical works have been performed so far that clearly confirm the formation of multi-splitting events for beams with a very high initial transverse power. Thus, numerical investigations in this field still need to be improved.

Finally, we should mention that for the case of two spatial dimensions with anisotropic dispersion, i.e., Eq.(10.1) with $D_\perp = 1$ and $s = -1$, it was strictly proven from virial arguments that collapse cannot take place [30].

4 Discussion of the general properties, outlook

In the previous sections, we have thoroughly detailed the self-focusing solutions of the NLS equation which classically proceeds from a combination of the Maxwell equations in nonlinear optics. Deriving NLS requires to assume certain simplifications. Here we shall briefly discuss the consequences of some of these simplifications in relation to the collapse dynamics. Let us for instance consider the scalar ansatz for modelling a wave envelope

$$\mathcal{E}(\vec{r},t) = \mathrm{Re}[E(\vec{r},t)e^{i(kz-\omega t)}] \tag{10.43}$$

where $\mathcal{E}$ is the scalar electric field of an electromagnetic wave propagating along z. This expression, once introduced in the combined Maxwell equations, gives a dispersion relation for which, in the presence of a nonlinear frequency shift $f(|E|^2)$, changes in frequency and wave number are equivalent to slow changes in the envelope of the wave with: $\omega \to \omega_0 + i\partial_t$, $k \to k(\omega_0) - i\partial_z$ around the central carrier wave number $k_0 \equiv k(\omega_0)$. Expanding this dispersion relation around k_0 yields [31]

$$\{-\vec{\nabla}_\perp^2 - \partial_z^2 - 2ik_0(\partial_z + \partial_t/v_g) + [(\frac{\partial k}{\partial \omega})^2 + k_0\frac{\partial^2 k}{\partial \omega^2}]_{\omega_0}\partial_t^2 - 2k_0\frac{\partial k}{\partial f}|_{f_0}\cdot f(|E|^2) \tag{10.44}$$
$$+\frac{\partial^2(k^2)}{i\partial\omega\partial f}|_{\omega_0,f_0}\partial_t\cdot f(|E|^2) + \frac{\partial^2(k^2)}{i\partial f\partial\omega}|_{f_0,\omega_0}f(|E|^2)\partial_t\cdot + ...\}E = 0$$

with $v_g = [\partial k/\partial\omega|_{\omega_0}]^{-1}$, and the leading balance in the paraxial approximation is then given by

$$k_0\partial_z E \sim \vec{\nabla}_\perp^2 E \sim (k_0/v_g)\partial_t E \sim 2k_0\frac{\partial k}{\partial f}|_{f_0}\cdot f(|E|^2)E \tag{10.45}$$

whence $\partial_z^2 E << k_0 \partial_z E$ and $(1/v_g)^2 \partial_t^2 E << k_0 \partial_z E$. Keeping the dominant first-order terms leads to the paraxial NLS equation for the wave field envelope expressed in the group velocity frame ($\eta = z/v_g - t$, $\xi = z$), while accounting for the group-velocity dispersion amounts to keeping the contribution $\partial^2 k/\partial\omega^2|_{\omega_0} \simeq s$ in the fourth term of Eq.(10.44). Up to the exchange in the (z, t) variables, this results in the starting equation Eq.(10.1) when considering a cubic nonlinearity $f(|E|^2) = |E|^2$. However, when nonlinear waves self-focus, the variations of their envelope with the propagation distance z and time t suddenly increase, so that the basic hypotheses of slowly-varying envelope approximations in space ($\partial_z^2 E << k_0 \partial_z E$) and in time ($\partial_t^2 E << \omega_0 \partial_t E$) may no longer hold. In particular, the paraxial approximation, for which the second derivative of the envelope with respect to z is supposed negligible, breaks down when the transversal spot size becomes of the order of the beam wavelength λ_0, i.e., when the transverse wave number of the beam envelope, $k_\perp$, increases such that $k_\perp^2/k_0^2 \to 1$. By investigating the influence of this second-order derivative on beam collapse, it was discovered that a 2D self-focusing, characterized by such an increase of $k_\perp$, would be arrested before the beam diameter becomes comparable to λ_0 [32, 33]. In addition, numerical investigations [34, 35] also showed that the catastrophic blow-up of high-power beams could be avoided near the point of self-focus by deviations from standard paraxiality introduced by the changes of the rapid phase contribution in Eq.(10.43) along the longitudinal direction. The beam power, that is an invariant of motion for the paraxial model, is no longer conserved, which finally allows the avoidance of collapse near the point of self-focusing, where the beam suddenly diffracts. This scenario can possibly be repeated along several focusing/defocusing cycles.

However, retaining the second-order derivative alone, as the only deviation to paraxiality, implies to deal with an ill-posed problem for which linear solutions diverge in the limit of large $k_\perp$ and generate explosive dispersion, whatever the nonlinearity may be. This cannot constitute a correct refinement of the paraxial approximation. To justify this statement we write the equation governing the evolution of the wave envelope E in the form

$$(\partial_z^2 + 2ik_0\partial_z + \vec{\nabla}_\perp^2)E = -2k_0 \frac{\partial k}{\partial f}|_{f_0} \cdot f(|E|^2)E \; . \tag{10.46}$$

To make sense, i.e., simply to admit solutions to a Cauchy problem, Eq.(10.46) must be a well-posed problem for which all the solutions to the linear part are regular functions, bounded for every transverse wave number $k_\perp$. In the opposite case, the linear solutions decompose over hyperbolic functions diverging exponentially: they cannot be bounded within standard spaces and yield non-integrable functions when iterating them along z. It can easily be seen that linear states associated with Eq.(10.46) are given by the Fourier multipliers $\exp\left[ik_0 z(\pm\sqrt{1 - k_\perp^2/k_0^2} - 1)\right]$, such that regular solutions do not exist for solving this Cauchy problem. As a particular

consequence, for any z arbitrarily close to zero, the linear states generate short-wave instabilities for $k_\perp > k_0$, leading to explosive dispersion [36]. Whereas the short-wave instability may be avoided in solutions to a linear wave equation by cancelling the range of large $k_\perp$ *a priori*, it certainly participates in nonlinear wave equations where the nonlinearity mixes all the short/long modes. In that case, Eq.(10.46) cannot constitute a correct refinement of the basic NLS model. Short-wave instability generating explosive dispersion develop on short transverse scales, which can explain the violent balance between the self-focusing potential and the singular longitudinal diffraction observed in Refs. [32, 33], when $k_\perp$ increases above the carrier wave number k_0. Such singular behaviors can be removed by adding even small additional dissipative terms in the form of either a time derivative accounting for the temporal dispersion of the waves or any other damping effect, which can regularize the short-wave instability [36].

In the opposite case when the second-order derivative in the evolution variable has an opposite sign to the Laplacian in Eq.(10.46), the Cauchy problem is generally well-posed and all the linear solutions appear to be regular, in the elementary sense that their Fourier transforms are bounded functions of the Fourier variable ($k_\perp$). This applies for instance when the temporal envelope approximation is dropped, as studied in [37] in the context of plasma turbulence.

In conclusion, we have reviewed the collapse properties of the NLS Eq.(10.1). In particular, we have described the wave evolution for the case of normal dispersion when s is negative in Eq.(10.1). We presented strong evidence for the absence of collapse in that case and emphasized that, instead, the wave pulse will split into two or more cells.

Acknowledgments: We thank Prs. D. Pesme, J. Ginibre, E.A. Kuznetsov, G. Laval, G. Matthieussent and J.C. Saut for stimulating discussions.

5 References

[1] A.C. Newell and J.V. Moloney, "Nonlinear Optics", (Addison-Wesley, Redwood City, California, 1991).

[2] J.R. Myra and C.S. Liu, "Self-modulation of ion Bernstein waves", Phys. Fluids 23 (1980) 2258.

[3] M.I. Weinstein, "Nonlinear Schrödinger equations and sharp interpolation estimates", Commun. Math. Phys. 87 (1983) 567.

[4] J. J. Rasmussen and K. Rypdal, "Blow-up in nonlinear Schroedinger equations-I: a general review", Physica Scripta 33 (1986) 481.

[5] L. Bergé, "Wave collapse in physics: Principles and applications to light and plasma waves", Physics Reports 303 (1998) 259.

[6] S.N. Vlasov and V.I. Talanov, "Wave self-focussing", Publications of the Institute of Applied Physics, Nizhny Novgorod, Russian Academy of Sciences (1997).

[7] E.A. Kuznetsov and V.E. Zakharov, "Nonlinear Coherent Phenomena in Continuous Media", This issue.

[8] R.T. Glassey, "On the blowing up of solutions to the Cauchy problem for nonlinear Schrödinger equations", J. Math. Phys. 18 (1977) 1794.

[9] S.N. Vlasov, V.A. Petrishchev and V.I. Talanov, "Averaged description of wave beams in linear and nonlinear media (the method of moments)", Izv. Vuz. Radiofiz. 14 (1971) 1353 [Radiophys. and Quantum Electronics 14 (1974) 1062].

[10] E.A. Kuznetsov, J. Juul Rasmussen, K. Rypdal and S.K. Turitsyn, "Sharper criteria for the wave collapse", Physica D 87 (1995) 273.

[11] V.E. Zakharov and E.A. Kuznetsov, "Quasi-classical theory of three-dimensional wave collapse", Zh. Eksp. Teor. Fiz. 91 (1986) 1310 [Sov. Phys. JETP 64 (1986) 773].

[12] L. Bergé and D. Pesme, "Bounded spatial extension of the self-similar collapsing solutions of the nonlinear Schrödinger equation", Physica Scripta 47 (1993) 323.

[13] L. Bergé and D. Pesme, "Non self-similar collapsing solutions of the nonlinear Schrödinger equation at the critical dimension", Phys. Rev. E 48 (1993) R684.

[14] G.M. Fraiman, "Asymptotic stability of manifold of self-similar solutions in self-focusing", Zh. Eksp. Teor. Fiz. 88 (1985) 390 [Sov. Phys. JETP 61 (1985) 228].

[15] V.M. Malkin, "Dynamics of wave collapse in the critical case", Phys. Lett. A 151 (1990) 285.

[16] M.J. Landman, G.C. Papanicolaou, C. Sulem, and P.L. Sulem, "Rate of blowup for solutions of the nonlinear Schrödinger equation at critical dimension", Phys. Rev. A 38 (1988) 3837.

[17] G. Akrivis, V. Dougalis, O. Karakashian and W. McKinney, "Galerkin-finite element methods for the nonlinear Schrödinger equation", Advances on computer mathematics and its applications, E. Lipitakis Ed., (World Scientific, 1993) p. 85-106.

[18] S.N. Vlasov, L.V. Piskunova and V.I. Talanov, "Three-dimensional wave collapse in the nonlinear Schrödinger equation model", Zh. Eksp. Teor. Fiz. 95 (1989) 1945 [Sov. Phys. JETP 68 (1989) 1125].

[19] N.E. Kosmatov, V.F. Shvets and V.E. Zakharov, "Computer simulation of wave collapses in the nonlinear Schrödinger equation", Physica D 52 (1991) 16.

[20] M.J. Landman, G.C. Papanicolaou, C. Sulem, P.L. Sulem and X.P. Wang, "Stability of isotropic singularities for the nonlinear Schrödinger equation", Physica D 47 (1991) 393.

[21] J. E. Rothenberg, "Pulse splitting during self-focusing in normally dispersive media", Opt. Lett. 17 (1992) 584.

[22] P. Chernev and V. Petrov, "Self-focusing of short light pulses in dispersive media", Opt. Commun. 87 (1992) 28 ; "Self-focusing of light pulses in the presence of normal group-velocity dispersion", Opt. Lett. 17 (1992) 172.

[23] G.G. Luther, A.C. Newell and J.V. Moloney, "The effects of normal dispersion on collapse events", Physica D 74 (1994) 59.

[24] G. Fibich, V. Malkin and G. Papanicolaou, "Beam self-focusing in the presence of small normal time dispersion", Phys. Rev. A 52 (1995) 4218.

[25] N.A. Zharova, A.G. Litvak, T.A. Petrova, A.M. Sergeev and A.D. Yunakovskii, "Multiple fractionation of wave structures in a nonlinear medium", Pis'ma Zh. Eksp. Teor. Fiz. 44 (1986) 12 [JETP Lett. 44 (1986) 13].

[26] L. Bergé and J. Juul Rasmussen, "Multi-splitting and collapse of self-focusing anisotropic beams in normal/anomalous dispersive media", Phys. Plasmas 3 (1996) 824.

[27] L. Bergé, J. Juul Rasmussen, E.A. Kuznetsov, E.G. Shapiro and S.K. Turitsyn, "Self-focusing of chirped optical pulses in media with normal dispersion", J. Opt. Soc. Am. B 13 (1996) 1879.

[28] J.K. Ranka, R.W. Schirmer and A.L. Gaeta, "Observation of pulse splitting in nonlinear dispersive media", Phys. Rev. Lett. 77 (1996) 3783.

[29] S.A. Diddams, H.K. Eaton, A.A. Zozulya and T.S. Clement, "Amplitude and phase measurements of femtosecond pulse splitting in nonlinear dispersive media", Opt. Lett. 23 (1998) 379.

[30] L. Bergé, E.A. Kuznetsov and J. Juul Rasmussen, "Defocusing regimes of nonlinear waves in media with negative dispersion", Phys. Rev. E 53 (1996) R1340.

[31] J.A. Powell, J.V. Moloney, A.C. Newell and R.A. Albanese, "Beam collapse as an explanation for anomalous ocular damage", J. Opt. Soc. Am. B 10 (1993) 1230.

[32] M.D. Feit and J.A. Fleck, Jr. "Beam nonparaxiality, filament formation, and beam breakup in the self-focusing of optical beams", J. Opt. Soc. Am. B 5 (1988) 633.

[33] G. Fibich, "Small beam nonparaxiality arrests self-focusing of optical beams", Phys. Rev. Lett. 76 (1996) 4356.

[34] N. Akhmediev, A. Ankiewicz and J.M. Soto-Crespo, "Does the nonlinear Schrödinger equation correctly describe beam propagation ?", Opt. Lett. 18 (1993) 411.

[35] J.M. Soto-Crespo and N. Akhmediev, "Description of the self-focusing and collapse effects by a modified nonlinear Schrödinger equation", Opt. Commun. 101 (1993) 223.

[36] D.D. Joseph and J.C. Saut, "Short-wave instabilities and ill-posed initial-value problems", Theoret. Comput. Fluid Dynamics 1 (1990) 191.

[37] L. Bergé and T. Colin, "A singular perturbation problem for an envelope equation in plasma physics", Physica D 84 (1995) 437.

11

Coherent Structures in Dissipative Nonlinear Optical Systems

J.V. Moloney

ABSTRACT Critical self-focusing can manifest itself in many different ways in nonlinear optical systems. In transparent materials, a continuous wave laser beam can collapse to a localized hot spot after propagating some finite distance. Here the nonlinear optical interaction is local and assumed to be instantaneous. In a wide aperture semiconductor laser, the self-focusing nonlinearity is sluggish, responding to the relatively slow motion of the total carrier density within the lasing medium. In both situations, large scale coherent structures are generated. In this article, we will review some novel manifestations of the self-focusing nonlinearity in which dissipation plays a prominent role. Our first example will address the issue of the propagation of an intense femtosecond-duration laser pulse which can undergo critical collapse in a transparent bulk material. Here weak plasma generation in the vicinity of the focal spot regularizes the collapse singularity. In the second example, we will illustrate how dynamic filamentation instabilities in wide aperture semiconductor lasers can be stabilized via an all-optical spatial/spectral filtering delay feedback scheme.

1 Introduction

The critical self-focusing nonlinearity is well established in nonlinear optics. Typically experimentalists go to great lengths to avoid this instability as its consequences are often catastrophic. For an excellent review of the early developments in the field, the reader should consult the articles by Shen [21] and Marburger [16]. In a transparent bulk dielectric medium, a continous wave (CW) laser beam or long laser pulse can undergo critical self-focusing when its power exceeds some critical value. The high fields generated at the focus lead to optical breakdown of the material. The optical breakdown mechanism depends on the length of the laser pulse. For CW down to microsecond duration laser pulses, the breakdown is dominated by thermal processes. Nanosecond pulses typically drive hypersonic acoustic waves via a nonlinear electrostriction mechanism and the material damage is usually accompanied by a loud bang. In the picosecond regime, the breakdown is typically due to plasma generation whereby seed elec-

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 229−246, 2000.

trons, existing due to impurities or generated by multiphoton ionization, are accelerated by the light fields in a multiplicative cascade fashion to form a relatively dense plasma in the focal region. The generated plasma can both absorb and defocus the laser field. This plasma shielding mechanism is utilized extensively in opthamology as a means of protecting the retina from damage during laser surgery. The story for femtosecond duration pulses is less clear. As the pulse duration decreases, the energy in the pulse (fluence) decreases. There is a definite correlation between energy fluence and optical damage, at least for nanosecond pulses through to CW beams. Shorter duration pulses display higher breakdown power thresholds and energy fluence is no longer a useful measure. For pulses with durations of the order of a few hundred femtoseconds or less, it becomes clear that the avalanche (cascade) breakdown mechanism should be less significant as the pulse is too short to provide photons on its trailing edge to sustain the cascade. It appears that the only breakdown mechanism available at this time scale is multiphoton ionization which is essentially instantaneous down to the order of a femtosecond interaction timescale. In this article we will provide evidence that the critical collapse singularity, which is so damaging for longer laser pulse interactions, provides a robust mechanism for forming and sustaining a novel dynamic nonlinear waveguide in air. This problem should be a source of exciting new developements both in physics and mathematics.

The self-focusing nonlinearity in wide aperture high power semiconductor lasers arises from an inertial response of the generated carriers to the internal laser fields. Although, the collapse singularity is not present here due to strong saturation of the nonlinear response, the self-focusing instability gives rise to focused hot spots, called intensity filaments, which can lead to facet damage and consequently, irreversible damage to the laser. This problem has plagued the laser engineering community for over two decades and so far has not been resolved satisfactorily. Interaction of light with a semiconductor material is an extremely complicated many-body problem involving the coupling of light to a multi-component plasma [6]. Understanding these interactions is an active area of ongoing research, and there remain many open questions. Significant progress has been made on this problem by the group of Stephan Koch in Marburg and our group at Arizona. For simplicity, we will ignore these complications in the present article and instead present a complex order parameter equation description which captures the essential experimentally observed features and offers the advantage of allowing substantial analytic progress. Basically, these lasers display complex dynamic filamentation instabilities from the moment of turn-on and these persist at all operating levels. The dynamic filaments are large scale spatial coherent structures which appear as intensity bursts at random across the laser device. These weakly turbulent spatiotemporal structures have an average period of the order of a few hundred picoseconds and can only be resolved in detail using streak camera techniques [23]. Typ-

ically laser experimentalists measure time-averaged outputs in the far-field (spatial Fourier transform) and these appear as a significant broadening beyond the diffraction-limited spot size expected for an illuminated aperture of the width of the laser. Basically, the laser output is randomly steering in space over an angular range of about 20 degrees with an average period of a few hundred picoseconds. This far-field broadening can be highly detrimental when such high power laser sources are envisioned as speed-of-light space-based communications systems. In the present article we will propose a control scheme which both stabilizes and steers the laser output.

Both problems are related mathematically as their description arises from an envelope approximation to the original vector Maxwell equations with an appropriate constitutive relation coupling the optical field to the material oscillation. Critical collapse occurs in the 2D Nonlinear Schrödinger equation,

$$i\frac{\partial\psi}{\partial z} = a\nabla^2\psi + |\psi|^2\psi, \tag{11.1}$$

when the total power $P = \int_{-\infty}^{+\infty} |\psi(r,t)|^2 dA$ exceeds a critical value. Here dA is an element of cross-sectional area. The space variable z refers to a moving reference frame and the critical collapse distance will be denoted by z_{NL}.

Here ψ is the complex envelope of the linearly polarized (scalar) electric field, a is a measure of the linear diffractive spreading of the transverse waist, ∇^2 is the 2D Laplacian operator where we assume cylindrical symmetry. The third term in the equation is the optical Kerr self-focusing nonlinearity. The "blow-up in finite time singularity" problem has been studied in great detail in the literature [8], and it is known that the collapsing filament assumes the following self-similar form,

$$\psi(\tau,\zeta,t) = g(\tau)\chi(\zeta)e^{-i\frac{\alpha}{4}\zeta^2+i\tau}, \tag{11.2}$$

where τ and ζ are scaled time-like and space-like variables and the pulse is parameterized by the variable t [15]. This nonlinear mode will play a prominent role in our later discussion.

The question that we will address here is what physical mechanism can regularize the collapse singularity when we extend the model to describe femtosecond laser pulse propagation in air. In essence, we add perturbation terms describing weak group velocity dispersion (GVD) (a Hamiltonian perturbation) and coupling to a plasma via a simple Drüde model (a dissipative perturbation). The moral of our story will be that the collapse singularity of the underlying 2D NLS equation becomes the robust nonlinear mechanism which is responsible for forming and sustaining a novel highly dynamic nonlinear waveguide in the atmosphere that may persist for tens of kilometers. The collapse singularity, instead of being an on-off explosive dynamical event, can recur a few times until the background

reservoir of energy remaining in the laser pulse is insufficient to sustain it. The perturbation terms essentially turn-on in some localized region of space (physically of the order of 80 microns in diameter) where the singularity is developed. The remainder of the laser pulse is unaffected and sees the 2D NLS as its governing equation. We remark here that the role of normal group velocity dispersion in regularizing the collapse singularity was first explained analytically in reference [15] and has subsequently been discussed by several authors [3, 4, 5, 10]. These works were preceded by some numerical simulations [7, 19] showing evidence that the pulse would split due to the local defocusing induced by normal GVD. Figure 1 displays a 3D surface rendering of the pulse after it has undergone a splitting. The ripples at both extremities are a manifestation of dispersive regularization of the shock waves formed due to the strong defocusing of the pulse in the time dimension. Indeed the on-axis split section of the pulse is reminiscent of the dark soliton solution to the 1D defocusing NLS equation. A recent experiment has confirmed this pulse splitting [20] and moreover, shown evidence for a finite cascade of splittings. Although we include the normal GVD term in our model below, we will not discuss it further as our results indicate that it is not the main player in the air propagation problem. A systematic numerical study of the critical collapse singularity and its regularization in the inert gas Argon has recently been carried out [18]. This work has shown that there is a transition from normal GVD dominated regularization to plasma dominated regularization as the gas pressure is varied. Moreover, this work establishes that more careful analytical and numerical studies are needed to resolve the extremely subtle balances between the different physical mechanisms operative in the vicinity of the critical focus. In the present article we confine our attention to the plasma-dominated regime.

The rather complicated semiconductor laser model can be reduced to a mathematically tractable form by making a few simplifying assumptions. We will restrict our study to a short cavity laser which is constrained to run in a single longitudinal mode. Transversely (in x and y) the laser is extended and can undergo dynamic filamentation instabilities. This level of approximation is a reasonable description of existing wide aperture Vertical Cavity Surface Emitting (VCSEL) lasers. In principle one can derive a complex order parameter description of such a semiconductor laser by analyzing the full physical laser model close to lasing threshold. This has been done in some detail for two-level and Raman lasers in references [13, 14, 12]. Basically the underlying mathematical description of such lasers is the complex Swift-Hohenberg (CSH) equation which has the generic form,

$$\frac{\partial \psi}{\partial t} = (\mu + i\nu)\psi + ia\nabla^2\psi + (\alpha + i\beta)(\Omega + \nabla^2)^2\psi + (1 + i\gamma)|\psi|^2\psi \quad (11.3)$$

This is an extension of the real Swift-Hohenberg equation derived originally to describe pattern formation in fluids. The present complex form is the

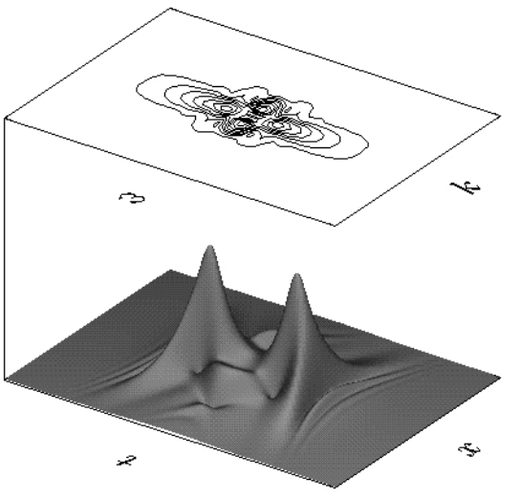

FIGURE 1. 3D rendering of the pulse splitting due to normal GVD. The top contour shows a modulated, strong spectral $(k-\omega)$ superbroadening.

generic description of pattern formation in a spatially extended system undergoing a Hopf bifurcation at $\mu = \mu_c = 0$. Here μ is the linear growth which is saturated by the real coefficient part of the nonlinear term. ν is a frequency and the imaginary part of the nonlinear term is the nonlinear modification to this term. The second term on the right hand side is the standard diffraction operator. The real part of the third term is a purely diffusive contribution which discriminates between transverse modes which may grow beyond $\mu = \mu_c$. The imaginary part of this term is a higher order correction to the diffraction operator. Obviously this equation contains the 2D NLS equation. A proper description of the semiconductor laser will require that the above equation be coupled to a mean flow describing the rather sluggish carrier density motion. An extended CSH model of this form will be shown to display the same qualitative weakly turbulent behavior as observed experimentally in broad area high power semiconductor lasers. Figure 2a shows a schematic of the technique used to measure the far-field output of a broad area semiconductor laser and, on the right, we show the experimentally measured nonlinear dispersion of a 100μ wide by 500μ long high power semiconductor laser undergoing strong dynamic filamentation while simultaneously oscillating in many longitudinal modes. By shortening the laser cavity from 500μ to about 20μ, we can limit the laser to running in a single longitudinal mode while retaining the parabolic shaped dispersion characteristic of transverse dynamic filamentation. We are interested in designing a scheme that can stabilize the laser output. As an added bonus of our control scheme, we obtain a means of steering the stabilized laser

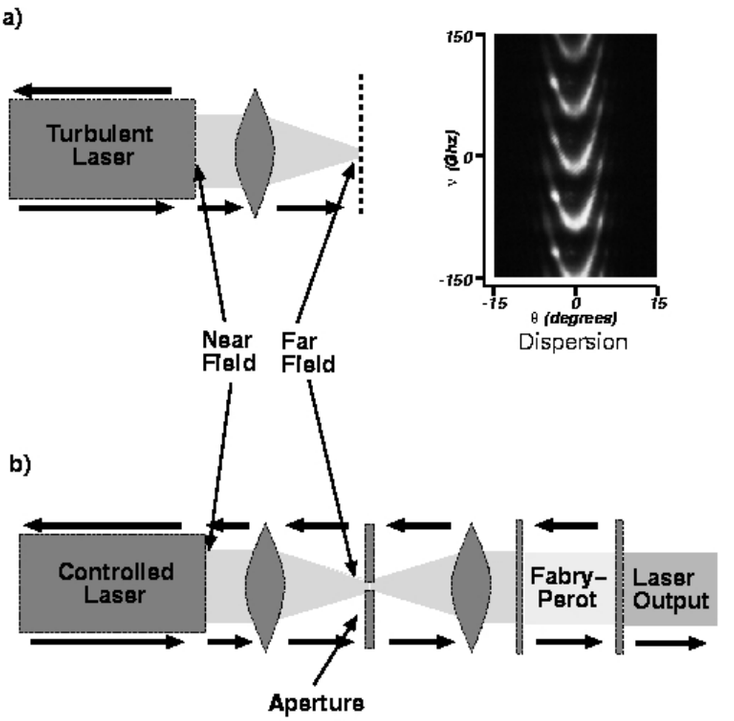

FIGURE 2. a) Left: schematic of a broad area semiconductor laser showing the near- and far-field locations. The picture on the right is an experimentally measured nonlinear dispersion of a real semiconductor laser (courtesy of D. Bossert, Air Force Research Laboratory). b) The proposed external feedback control. Placing an aperture in the far-field acts as a spatial filter.

2 Nonlinear waveguide channeling in air

Recent experiments using intense laser pulses with characteristic durations of the order of a hundred femtoseconds, have displayed a novel self-induced channeling phenomenon, accompanied by plasma generation [2, 24, 1]. There has been some debate in the literature as to the precise mechanism by which such a nonlinear channel could be formed. One suggestion is that there can be a perfect balance between the intrinsic self-focusing of air and the weak self-defocusing due to the generated plasma, leading to a passive stationary waveguide. Another suggestion is that the characteristic

length of the channel should correlate to the predictions of the so-called "moving focus model". Here different transverse cross-sections of the pulse undergo critical collapse at different distances due to the different powers contained in each transverse slice. We proposed an alternative mechanism based on numerical simulations which contains the moving focus picture as an ingredient. Basically, this latter picture was used earlier to predict analytically that normal GVD would arrest collapse [15]. Our model for the propagation of femtosecond pulses in air requires augmentation of the basic 2D NLS equation by a plasma absorption and defocusing mechanism and we include normal GVD for completeness. The extended envelope equation becomes [9, 17],

$$\begin{aligned} \frac{\partial \psi}{\partial z} &= \frac{i}{2k}\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}\right)\psi - \frac{ik''}{2}\frac{\partial^2\psi}{\partial t^2} - \frac{\sigma}{2}(1+i\omega\tau)\rho\psi - \frac{\beta^{(K)}}{2}|\psi|^{2K-2}\psi \\ &+ ik_0 n_2|\psi|^2\psi \qquad (11.4) \\ \frac{\partial \rho}{\partial t} &= \frac{1}{n_b^2}\frac{\sigma}{E_g}\rho|\psi|^2 + \frac{\beta^{(K)}|\psi|^{2K}}{K\hbar\omega} - a\rho^2 \qquad (11.5) \end{aligned}$$

where the terms on the right-hand-side describe transverse diffraction, group-velocity dispersion (GVD), absorption and defocusing due to the electron plasma, multi-photon absorption (MPA), and nonlinear self-focusing (SF), respectively. Here ω is the optical frequency, $|\psi|^2$ the intensity, $k = n_b k_0 = n_b\omega/c$, $k'' = \partial^2 k/\partial\omega^2$ charaterizes the GVD, positive for normal dispersion and negative for anomalous dispersion, ρ is the electron density, σ the cross-section for inverse bremsstrahlung, τ is the electron collision time, and $\beta^{(K)}$ is the K-photon absorption coefficient. The first term on the right-hand-side of the density equation describes growth of the electron plasma by cascade (avalanche) ionization, and the second term is the contribution of multiphoton absorption (MPA) which generates free electrons. The interaction with the light is so fast ($\approx 2\times 10^{-13}$ of a second), that we can ignore plasma recombination (third term) and plasma transport.

The key idea now is that the additional terms introduced over and above the basic 2D NLS terms are weak perturbations that take effect only in localized regions of space about where the singularity develops. Their effect is not felt anywhere else in the interaction zone. We imagine a tightly constrained laser pulse propagating in the z-direction with an initial critical power well exceeding that required for collapse. A wide laser pulse can collapse into multiple transverse interacting filaments which would violate the radial symmetry assumptions used here. A good estimate of the critical collapse (self-focusing) power is given by the formula $P_{crit} = \lambda_0^2/2\pi n_b n_2$. For propagation in air this critical power is of the order 1 to 2 GW. The multiphoton ionization is a 7 photon process so $K = 7$ in the above equations. For the physical parameters for air, the normal GVD term is not a player in arresting critical collapse, at least at normal atmospheric pressure. The only physics that can stop the collapse is the generation of a

plasma when the critical intensity for plasma generation in air is reached. The optical breakdown intensity measured for air is $I_{Br} \approx 10^{13} W/cm^2$. We assume an input laser pulse with a Gaussian spatial and temporal profile.

2.1 Dynamic spatial replenishment of femtosecond pulses propagating in air

Laser pulses exceeding about a picosecond in duration are likely to undergo optical breakdown by a mechanism that involves the generation of a plasma via the avalanche photo-ionization mechanism [17]. The first term on the right hand side of the Drüde model for the plasma density above captures this avalanche process. Seed electrons are needed to get the process going and these can exist due to impurities or may be produced by multiphoton ionization (the second term on the RHS of the Drüde model above). In any event, the initial electron density begins to grow dramatically as long as there exists an optical field to drive the process. As the pulse duration gets shorter, the relative importance of this plasma generation mechanism decreases and the essentially instantaneous multiphoton mechanism takes over. The plasma generated by a 20 picosecond duration laser pulse in water, is sufficiently dense to defocus and absorb the laser light thereby acting as a plasma shield [9]. The situation for a few hundred femtosecond laser pulse propagating in air is quite different as we shall now see.

Figure 3 shows a sequence of graphs of the maximum (in time) of the on-axis laser intensity and corresponding plasma density as a function of propagation distance z. Each curve corresponds to a different initial peak power in the pulse. The initial pulse waist was chosen to be $0.7mm$, the central pulse wavelength was $770nm$ giving a Rayleigh range (measure of linear diffraction) of $2m$. The rapid rise in the pulse intensity is simply the onset of self-similar collapse of the 2D NLS equation. The collapse distance z_{NL} is about $0.5m$. The subsequent flattening of the intensity is associated with the initial sharp spike in the plasma density. Thus weak plasma generation acts to arrest the collapse singularity. The flattened intensity profile begins to slowly decay (on-axis) and the plasma generation begins to fall off. We note the appearance of a sharp kink in the intensity profile at about $1.3m$ with an accompanying secondary plasma spike. Indeed this occurs a third time at around $2m$. We can conclude from this picture that light appears to self-channel well beyond the initial collapse distance and that plasma spiking is playing a central role.

Figure 4 presents a sequence of instantaneous images of the spatiotemporal evolution of the full pulse and accompanying plasma density at a set of locations in z. At $z = 0.43m$ we see the initial focusing of the laser pulse with the accompanying growth of the plasma density in time. Just after the arrest of the collapse singularity we observe at $z = 0.65m$, a focused intensity spike on the leading edge of the pulse, followed by a deep furrow

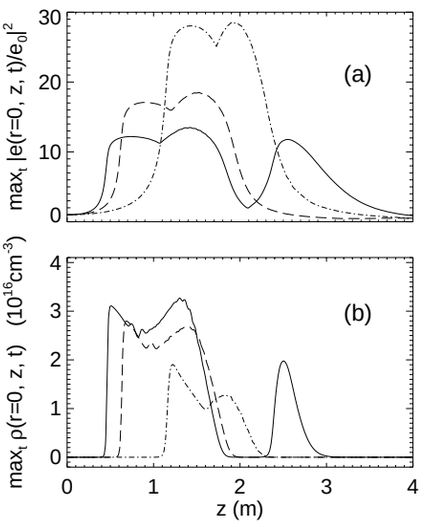

FIGURE 3. Maximum intensity (a) and plasma density (b) in z for their different initial powers. $P = 5.5P_{crit}$ (dash-dot), $P = 6.0P_{crit}$ (dashed), $P = 6.5P_{crit}$ (solid).

on-axis. This furrow arises primarily because of the defocusing action of the generated plasma. The plasma is too weak to significantly absorb the pulse energy. At $z = 0.83m$, the light behind the leading intensity spike is completely evacuated from the central region due to the defocusing action and the plasma density begins to decrease ($z = 1.16m$). There appears to be a single focused intensity spike remaining. As the plasma has disappeared, we now have a situation where the defocused back region has sufficient power in a transverse slice to attempt a secondary critical collapse. Here the 2D NLS is essentially the operative mathematical prescription as the perturbation terms (GVD and plasma density) are very small. We now observe a secondary intensity and plasma spike ($z = 1.49m$). Notice that the leading pulse is decaying and the secondary pulse has taken over. The first notch appearing in Figure 3 is indicative of the second spike taking over from the first as the highest intensity on-axis point. This highly dynamic explosive situation occurs once more before the pulse exhausts its energy for a further collapse.

This scenario of recurring collapse events regularized by weak plasma generation provides a beautiful illustration of the role of the self-similar collapse mode as the robust nonlinear structure which causes and sustains a highly dynamical waveguide in air. If one scales this problem up to an initial waist of around $6cm$ the laser pulse could, in principal, propagate over tens of kilometers. A recent experiment by a group at Jena in Ger-

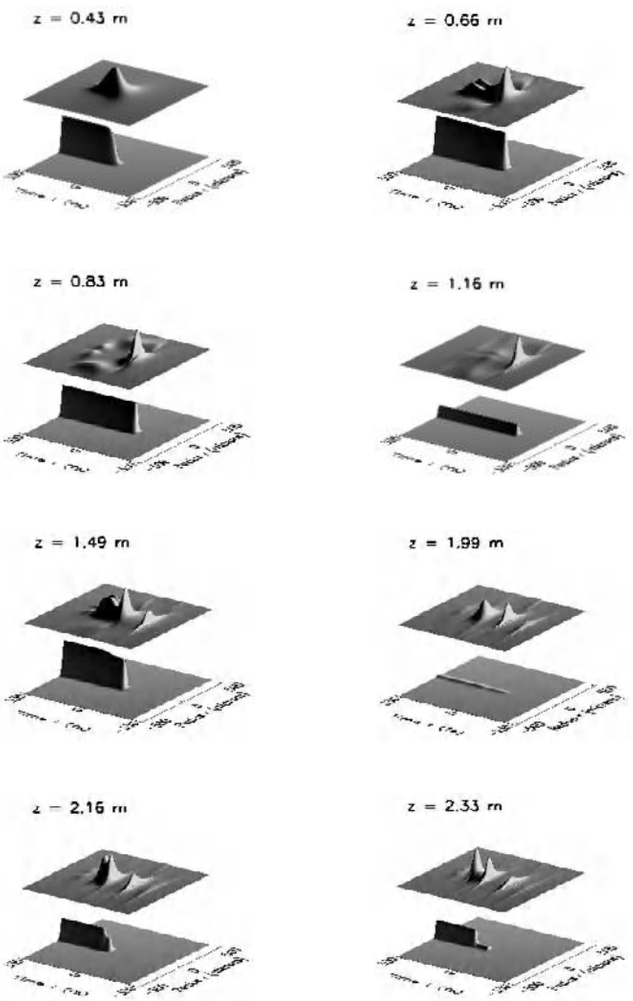

FIGURE 4. Sequence of snapshots from a movie showing 3D rendering of the recurring collapse events.

many, has provided the first evidence that high power very intense 100 femtosecond duration laser pulses can propagate on the order of 12 kilometers vertically into the upper atmosphere [22]. In this experiment the initial pulse had around 1000 critical powers and was observed to break up transversely into hundreds of interacting filaments. Accompanying the critical collapse is the generation of a white-light continuum. Potential applica-

tions for this remarkable phenomenon are lightning control [24], LIDAR and remote sensing applications [22].

3 Control of optical turbulence in semiconductor lasers

The weakly optically turbulent behavior exhibited by Figure 2 for a high power broad-area edge emitting semiconductor laser has frustrated the attempts by laser engineers to produce high-brightness compact light sources for a wide range of important applications. One can reduce the complexity of this system by making the laser short (at the expense of reduced output power) and opening it up in its two transverse (x-y) dimensions (to recover some of the power). In effect, we now have a surface-emitting rather than an edge emitting laser. By shortening the cavity, we restrict the laser to oscillate in a single longitudinal mode, while retaining the full complexity of the transverse filamentation instability. The latter can be modelled by a phenomenological complex Swift-Hohenberg equation coupled to a mean flow. We will see that this model captures the filamentation observed in the real device. The CSH model, which is a simple extension of that derived for a 2-level system [13, 14], is given by

$$\begin{aligned}(\sigma+1)\frac{\partial\psi}{\partial t} &= \sigma(r(x)-1)\psi + ia\nabla^2\psi - i\sigma\Omega\psi \\ &\quad -\frac{\sigma}{(1+\sigma)^2}(\Omega + a\nabla^2)^2\psi - \sigma(1+i\alpha)n\psi \qquad (11.6)\\ \frac{\partial n}{\partial t} &= -bn + |\psi|^2.\end{aligned}$$

The complex order parameter ψ is the scaled envelope of the electric field and n is a scaled relative carrier density. The latter acts as a mean-flow and has a profound influence in destabilizing the system, leading to a very complicated linear growth behavior of the traveling wave solutions. Here σ is the scaled cavity loss coefficient, a is proportional to the inverse of the Fresnel number of the laser and measures the characteristic length scale in the transverse dimension relative to the wavelength of light, Ω is the dimensionless detuning of the laser frequency from the gain peak, ∇^2 is either a one-dimensional or two-dimensional Laplacian and b is the dimensionless ratio of the carrier recombination to polarization dephasing times in the Semiconductor Bloch equations [6]. This equation differs from the 2-level model by the presence of the constant α multiplying the nonlinear term in the ψ-equation. This constant is called the *Linewidth Enhancement* or α *factor* and it provides a strong amplitude-phase coupling which promotes a strong phase instability growth [11].

The dimensionless parameter $r(x)$ is a scaled pumping and this is the principal bifurcation parameter of the problem. When $r > r_c = 1$, the laser turns on. Figure 5 shows the results of a numerical simulation of the 1D CSH equation starting from noisy initial data when r exceeds r_c. In order to mimic the operation of a true broad-area semiconductor laser, we assume a finite pumping $r(x)$ in the transverse x-dimension. Regions outside the pump are absorbing. We present the data in a manner where we can compare it to the experimental data in Figure 2. The graph in Figure 5 is the numerical nonlinear dispersion of the CSH equation and is computed by taking the spatial Fourier transform of $\psi(x,t)$ and accumulating the spectrum in time after all initial transients have died out. We see the same parabolic shape as in the experiment - here however, only a single longitudinal mode is oscillating. The dashed curve in Figure 5 is the analytic dispersion curve of the CSH given below with γ equal to zero. The weakly turbulent output corresponds to the fact that the output laser beam is steering randomly while undergoing dynamic filamentation. Analysis of the CSH equation for values of α corresponding to a real semiconductor laser indicates that its traveling wave solutions, $\psi = Ae^{i(\vec{k}\cdot\vec{x}-\omega t)}, n = A^2/b$, where $A^2 = b(r-1-(\frac{\Omega-ak^2}{1+\sigma})^2)$, are nowhere stable. In practice, the average period of the beam steering is a few hundred picoseconds.

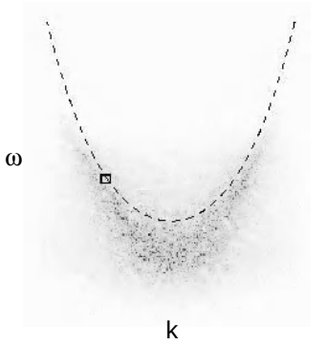

FIGURE 5. Numerical computed nonlinear dispersion of the semiconductor CSH equation. The dashed curve is the corresponding analytic dispersion curve given in the text.

3.1 The control

One would like to devise a control scheme which would lock the solution to a single point on the nonlinear dispersion curve. Optically, it is natural to introduce spatial filtering of the far-field $(k-\omega)$ and temporal filtering (a Fabry-Perot etalon) in an experimental setting. This is indicated schematically in Figure 2b. Our control which is added to the RHS of the

ψ-equation, above has the form,

$$\epsilon_\psi = -\gamma \left[\psi(\mathbf{x},t) - (1-R) \sum_{n=1}^{\infty} R^{n-1} \tilde{\psi}_n(\mathbf{x}, t-n\tau) \right] \tag{11.7}$$

This combined spatial/spectral optical feedback control acts as an added loss term for all but the selected TW solution (a specific k and ω on the dispersion curve). It compares the immediate output of the laser $\psi(\vec{x},t)$ with a time delayed spatially and spectrally filtered feedback. The term $\tilde{\psi}_n$ is constructed by taking the Fourier transform in space of ψ, multiplying it by a filter function and inverting the Fourier transform. The multiple-pass action of the Fabry etalon is represented by the summed term and the delay time is determined by the frequency once the wavenumber k is specified on the analytic dispersion curve. The strength of the feedback is measured by the parameter γ.

The semiconductor CSH with the added feedback can be analyzed and its traveling wave solutions explicitly found. The added delay introduces a further infinity of solutions but most of these are not physically realizable. The analytic dispersion relation for the full system is given by

$$\begin{aligned} (\sigma+1)\omega &= -ak_c^2 - \sigma\Omega - \sigma\alpha \left[r - 1 - \left(\frac{\Omega - ak_c^2}{1+\sigma} \right)^2 \right. \\ &\quad \left. + \frac{\gamma}{\sigma} \left(1 - (1-R) \left(\frac{\cos(\omega\tau) - R}{1 + R^2 - 2R\cos(\omega\tau)} \right) \right) \right] \\ &\quad + \frac{\gamma(1-R)\sin(\omega\tau)}{1 + R^2 - 2R\cos(\omega\tau)}. \end{aligned} \tag{11.8}$$

Given a wavenumber k, this transcendental equation needs to be solved graphically for ω and the only physically allowed solutions are those which yield positive values for the intensity of the new TWs with feedback control included,

$$\begin{aligned} A^2 &= b \left[r - 1 - \left(\frac{\Omega - ak_c^2}{1+\sigma} \right)^2 \right. \\ &\quad \left. + \frac{\gamma}{\sigma} \left(1 - (1-R) \left(\frac{\cos(\omega\tau) - R}{1 + R^2 - 2R\cos(\omega\tau)} \right) \right) \right]. \end{aligned} \tag{11.9}$$

Figure 6 shows such a graphical solution where the parameters used are indicated in the caption. The open diamond corresponds to the TW solution of the original isolated laser at the selected wavenumber k. If this solution is achieved, the feedback control ϵ_ψ would be identically zero. The "+" and "*" symbols denote the additional solutions introduced by the delay feedback term; the "+" ones are the only physical solutions for the

reason stated above. The situation depicted here corresponds to a choice of physical parameters such that there are 3 additional physically accesible solutions at different frequencies for the selected k. These solutions correspond to a finite amount of energy in the feedback loop.

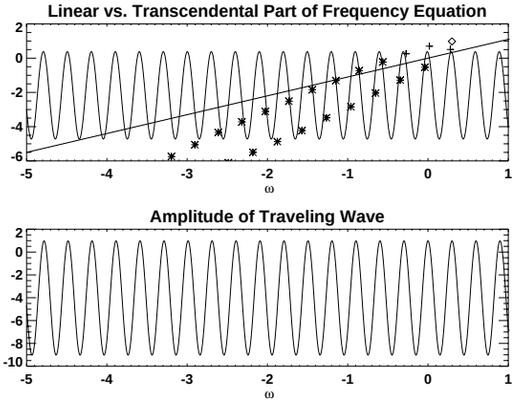

FIGURE 6. Graphical solution of the transcendental equation (1:8). The symbols are the computed values of A^2 from (1:9). Parameters are: $a = 0.01$, $b = 0.01$, $\Omega = 0.001$, $r = 2$, $\sigma = 0.1$, $\alpha = -5$, $k_c = 4$, $\gamma = 0.5$, $R = 0$

Figure 7 shows a numerical simulation starting from noise (remote from the the desired control state - the open diamond). The center of the transverse 1D output ($x = 0$) is being tracked in time. After switch-on, the system appears to settle initially into a dynamic state. However, the spectrally resolved output shows that this state corresponds to beating between the two leftmost "+" symbols and is a metastable solution. As the latter are known from the analysis, we can determine their stability. The latter are weakly unstable and the system eventually switches to the desired control solution (open diamond).

We find that this scheme is remarkably robust and the control solution was achieved for all conditions tested as long as the feedback strength exceeds about 5%. In 2D there is the added flexibility to try and control the system with more complex filters in order to stabilize more general patterns. We have also demonstrated the control for a variety of situations in 2D. Figure 8 shows a snapshot of the 2D weakly turbulent near- (top) and far- (bottom) fields for a laser pumped over a finite square cross-section i.e $r(x, y) > 1$ in the central square region. The far-field (k_x, k_y) spectrum is extremely broad. After the control has acted for a while, the far-field collapses down to a narrow spectral feature whose finite width is due to the TW being a finite supported rather than infinitely extended object. The corresponding real part of the near-field (upper right hand corner)

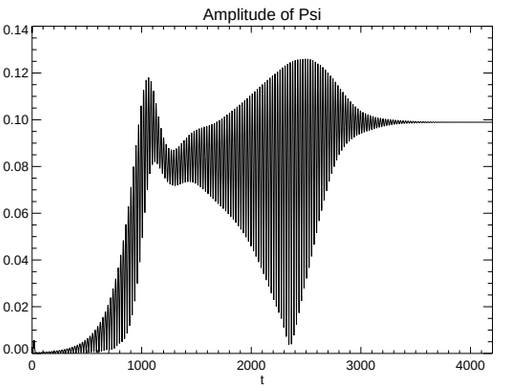

FIGURE 7. Demonstration of the action of the control showing a plot of the center ($x = 0$) of the transverse evolving pattern.

looks like parallel stripes corresponding to a uniform intensity illumination at a finite off-axis angle.

4 Summary and conclusions

Two nonlinear optical phenomena where large scale coherent structures play a prominent role have been discussed. Both topics are extremely rich and are likely to pose major challenges in nonlinear science in the next decades. The results presented here represent an initial attempt to understand the underlying nonlinear phenomena. Much remains to be explained and discovered. For example, we know that the high power femtosecond atmospheric pulse propagation problem involves initial data with hundreds to thousands of critical powers. Consequently, we expect a transverse modulational instability to give rise to a highly dynamic state involving simultaneous collapsing filaments. Can these form some type of weakly turbulent cooperative state which is sufficiently robust to penetrate through a nonhomogeneous, pressure varying atmosphere? The severe space and time compression accompanying these recurring collapses means that higher order terms must be included in our physical model. In some instances it will be necessary to discard the envelope approach and revert to a full 3D vector Maxwell description. High power semiconductor lasers are extremely complex nonlinear systems. We have made significant progress in being able to compute the many-body microscopic semiconductor optical response from a first principles base. This is only half of the battle. Designing schemes to stabilize high-power lasers remains a major theoretical and computational challenge. We know that very weak optical feedback can destabilize these

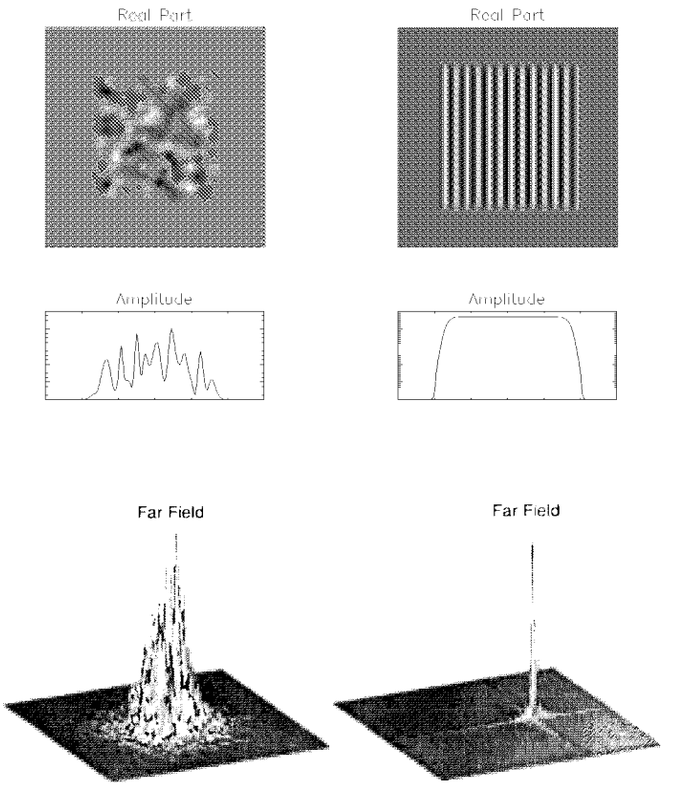

FIGURE 8. Control in 2D of a finitely pumped surface emitter.

lasers and that plasma and lattice heating are major players in determining performance. Nonlinear optics remains a vitally important technology area which relies heavily on phenomenology despite its critical importance. Computational nonlinear optics is an unexplored area of research open to exploitation by numerical analysts and computational scientists alike.

Acknowledgments: The research described in this chapter represents the efforts of a number of colleagues, postdoctoral fellows and students over the past four years. Ewan Wright has been a continual contributor and motivator in the ultrashort pulse work over the years. Alan Newell, Greg Luther and Quan Yuan Feng were major players in the early phases of this problem when the focus was on critical collapse in water and laser eye

damage. More recently, my student Michal Mlejnek, has been responsible for bringing this problem to a new level and the atmospheric propagation work highlighted here is his Ph.D. thesis work. Joceline Lega and Alan Newell were the driving forces behind the laser patterns work. Joceline and my student, David Hochheiser, were the key players in the control work described here.

Of course, none of this research could be carried out without the continued generous support of the Air Force Office of Scientific Research (AFOSR). The work described here was supported by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF under grant numbers F49620-97-1-0002, F49620-94-1-0463, F49620-95-1-0454, F49620-98-1-0227.

5 References

[1] A. Brodeur, C.Y. Chien, F.A. Ilkov, S.L. Chin, O.G. Kosareva, and V.P. Kandidov. Moving focus in the propagation of ultrashort laser pulses in air. *Optics Lett.*, 22:304, 1997.

[2] A. Braun, G. Korn, X. Liu, D. Du, J. Squier, and G. Mourou. Self-channeling of high-peak-power femtosecond laser pulses in air. *Optics Lett.*, 20:73, 1995.

[3] L. Berge, E.A. Kuznetsov, and J.J. Rasmussen. Defocusing regimes of nonlinear waves in media with negative dispersion. *Phys. Rev. E*, 53:R1340, 1996.

[4] L. Berge and J.J. Rasmussen. Multisplitting and collapse of self-focusing anisotropic beams in normal/anomalous dispersive media. *Physics of Plasmas*, 3:824, 1996.

[5] L. Berge, J.J. Rasmussen, E.A. Kuznetsov, E.G. Shapiro, and S.K. Turitsyn. Self-focusing of chirped optical pulses in media with normal dispersion. *J. OSA B*, 13:1879, 1996.

[6] W.W. Chow, S.W. Koch, and M. Sargent III. *Semiconductor Laser Physics*. Springer-Verlag, New York, 1994.

[7] P. Chernev and V. Petrov. Numerical simulation of nonlinear pulse propagation in the self-focusing limit. *Nonlinear Optics, Principles, Materials, Phenomena and Devices*, 2:311, 1992.

[8] S. Dyachenko, A.C. Newell, A. Pushkarov, and V.E. Zakharov. Optical turbulence: weak turbulence, condensates and collapsing filaments in the nonlinear schrodinger equation. *Physica D*, 57:96, 1992. and references therein.

[9] Q. Feng, J.V. Moloney, A.C. Newell, E.M. Wright, K. Cook, P.K. Kennedy, D.X. Hammer, B.A. Rockwell, and C.R. Thompson. Theory and simulation on the threshold of water breakdown induced by focused ultrashort laser pulses. *IEEE JQE*, page 127, 1997.

[10] G. Fibich and G.C. Papanicolaou. Self-focusing in the presence of small time dispersion and nonparaxiality. *Optics Lett.*, 22:1379, 1997.

[11] D. Hochheiser, J.V. Moloney, and J. Lega. Controlling optical turbulence. *Phys. Rev. A*, 55:R4011, 1997.

[12] J. Lega, P.V. Jakobsen, J.V. Moloney, and A.C. Newell. Nonlinear transverse modes of large-aspect-ratio homogeneously broadened lasers: I. Pattern analysis near and beyond threshold. *Phys. Rev. A*, 49:4201, 1994.

[13] J. Lega, J.V. Moloney, and A.C. Newell. Swift-hohenberg equation for lasers. *Phys. Rev. Lett.*, 73:2978, 1994.

[14] J. Lega, J.V. Moloney, and A.C. Newell. Universal description of laser dynamics near threshold. *Physica D*, 83:478, 1995.

[15] G.G. Luther, A.C. Newell, and J.V. Moloney. The effects of normal dispersion on collapse events. *Physica D*, 74:59, 1994.

[16] J.H. Marburger. Self-focusing: Theory. In J.H. Sanders and S. Stenholm, editors, *Progress in Quantum Electronics*, page 35. Pergamon Press, 1977.

[17] M. Mlejnek, E.M. Wright, and J.V. Moloney. Dynamic spatial replenishment of femtosecond pulses propagating in air. *Optics Lett.*, 23:382, 1998.

[18] M. Mlejnek, E.M. Wright, and J.V. Moloney. Femtosecond pulse propagation in argon - a pressure dependence study. *Phys. Rev. E*, 1998.

[19] J.E. Rothenberg. Pulse splitting during self-focusing in normally dispersive media. *Optics Lett.*, 17:583, 1992.

[20] J.K. Ranka, R.W. Shirmer, and A.L. Gaeta. Observation of pulse splitting in nonlinear dispersive media. *Phys. Rev. Lett.*, 77:R1340, 1996.

[21] Y. R. Shen. Self-focusing: Experimental. In J.H. Sanders and S. Stenholm, editors, *Progress in Quantum Electronics*, page 1. Pergamon Press, 1977.

[22] L. Woeste, C. Wedekind, H. Wille, P. Rairoux, B. Stein, S. Nikolov, C. Werner, S. Niedermeier, F. Ronneberger, H. Shillinger, and R. Saurbrey. Femtosecond atmospheric lamp. *Laser und Optoelectronik*, 29:51, 1997.

[23] N. Yu, R.K. DeFreez, D.J. Bossert, G.A. Wilson, R.A. Elliott, S.S. Wang, and H.G. Winful. Spatiospectral and picosecond spatiotemporal properties of a broad area operating channeled-substrate-planar laser array. *Applied Optics*, 30:2503, 1991.

[24] X.M. Zhao, J.-C. Diels, C.Y. Wang, and J.M. Elizondo. Femtosecond ultraviolet laser pulse induced lightning discharges in gases. *IEEE JQE*, page 599, 1995.

12

Solitons in Optical Media with Quadratic Nonlinearity

Boris A. Malomed

ABSTRACT A brief review of basic models describing second harmonic generation (SHG) in dispersive and diffractive optical systems, and of fundamental soliton solutions to these models is given. Included are two- and three-wave models, gap solitons, linearly coupled waveguides with an intrinsic quadratic nonlinearity, multidimensional "light bullets", two-soliton bound states (which are always unstable in the usual SHG models, but may be stable in the SHG models that give rise to gap solitons), and modulational instability of cw states. Both analytical results, obtained by means of the variational approximation (VA), and numerical results are presented. In practically all the cases, VA demonstrates a fairly good agreement with direct simulations.

1 Introduction

Nowadays, optical solitons constitute a huge research area that has been attracting a great deal of interest. Several fundamental monographs have summarized the results of theoretical and experimental studies obtained, chiefly, for the solitons propagating in glass fibers [1]. This type of solitons is referred to as being in the *temporal* domain.

In generating optical solitons, the crucial role belongs to nonlinear properties of the medium. Usually, the optical nonlinearity (the Kerr effect) is quite weak. In optical fibers, this can be compensated by choosing the carrier wavelength of light close to the zero-dispersion point, where even a weak nonlinearity can compete with the dispersion. However, the search for physical mechanisms that are able to induce a stronger optical nonlinearity is a very relevant problem.

An important result in this direction was obtained when it was experimentally demonstrated that a classical nonlinear optical effect in the form of the *second harmonic generation* (SHG) induced by a quadratic ($\chi^{(2)}$), rather than cubic ($\chi^{(3)}$), nonlinearity of the medium, can be used to induce a very strong effective $\chi^{(3)}$ nonlinearity through the *cascading* mechanism (see details and references to the original works in the review [2]). This

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 247−262, 2000.

means that a pump wave at the *fundamental harmonic* (FH), generates its *second harmonic* (SH), i.e., a wave with the double frequency, and then, because the SHG process is reversible, interaction of the remaining FH with the generated SH induces the inverse down-conversion of SH into FH. As a result, the incoming FH is transformed into itself, but with a large phase shift generated by the SHG process and its inverse. The nonlinear phase shift generated this way is the same as would be generated by a very strong *effective* Kerr nonlinearity, provided that the *phase-matching* conditions (see below) for FH and SH are met.

SHG was the first effect that gave rise to the development of nonlinear optics forty years ago [3], but, until recently, SHG was studied without consideration of its interplay with diffraction or temporal dispersion of light in the material, therefore solitons could not be supported by this process. On the other hand, the mathematical model of the SHG generation is exactly integrable by means of the *inverse scattering transform*, in the case when the model ignores the diffraction or dispersion, but takes into regard a group-velocity difference between the harmonics [4]. Still earlier, the first theoretical works had appeared which correctly predicted that the interplay of SHG and dispersion gives rise to two-wave solitons, in which a stable dynamical balance is maintained between the SHG process and its inverse, as well as between the effective nonlinear self-phase modulation of FH, through the cascading mechanism, and the chromatic dispersion [5]. Another outstanding early work [6] has demonstrated that the $\chi^{(2)}$ nonlinearity does not lead to a wave collapse in any physical dimension, hence stable *multidimensional* solitons with the mutually trapped harmonics should be possible.

In the last years, the development of this field, both theoretical and experimental, was extremely rapid. Recently studied cascading effects in the quadratically nonlinear media were summarized in the review article [2]. The most important physical results were the experimental observations of the $\chi^{(2)}$ solitons. Until very recently, no solitons of this type have been observed in the temporal domain, while many experiments were reported for the *spatial solitons*, viz., stationary self-trapped cylindrical light beams in the three-dimensional (3D) bulk [7], or self-supporting light stripes in a two-dimensional (2D) $\chi^{(2)}$ waveguide [8]. However, an observation of narrow *temporal* $\chi^{(2)}$ solitons, with a width of 58 fs at the carrier wavelength 527 nm, has just been reported [9]. The medium in which this experiment was done is a so-called BBO crystal.

Besides the great interest that the $\chi^{(2)}$ solitons have as a fascinating object for fundamental research, many hopes are also related to the potential they have for applications. The most important applied problem is all-optical switching in multichannel optical communication systems, when one beam of light has to steer another one (in particular, the signal beam must be able to switch the control one between different positions) [2].

An objective of this chapter is to give a concise review of the basic

theoretical models that describe various types of the $\chi^{(2)}$ systems, and of the most important results obtained for solitons in those models (however, dissipative effects are not included). The review is strongly biased to displaying theoretical results, as the theory is going far ahead of experiments in this area (note that 21 years had elapsed between the prediction of the $\chi^{(2)}$ solitons [5] and their experimental observation [7]). Nevertheless, there is a strong hope that further development of the experimental techniques will make it possible to observe the many interesting effects that recently have been theoretically predicted.

2 The basic theoretical models

Mathematical models of SHG processes in diffractive and dispersive media are derived from the Maxwell's equations in a dielectric medium, supplemented by terms that describe a quadratically nonlinear response of the medium to the applied field. As a result, one can derive rather simple models [2] . The simplest among them describes the SHG process in the 1D medium, FH and SH being represented each by a single component (u and v, respectively)

$$iu_z + \frac{1}{2}u_{\tau\tau} - u + u^* v = 0, \tag{12.1}$$

$$2iv_z + \frac{1}{2}\delta v_{\tau\tau} - \alpha v + \frac{1}{2}u^2 = 0. \tag{12.2}$$

Here z is the coordinate along which the light propagates, and $\tau \equiv t - z/V_{\rm gr}$, where t is the physical time, and $V_{\rm gr}$ is the mean group velocity of the carrier wave. The coefficient $\alpha > 0$ measures a *mismatch* between the two harmonics (the cases $\alpha > 4$ and $\alpha < 4$ are referred to as positive and negative mismatch, respectively), and the quadratic terms account for the FH-SH conversion. The SHG process may be effective if one can find a value of the FH frequency ω such that, according to the dispersion relation of the material medium, $k = k(\omega)$, the SH wave number (also called *propagation constant*), $k(2\omega)$, is close to $2k(\omega)$. In the model (12.1, 12.2), it is assumed that, in the lowest approximation, $k(2\omega) = 2k(\omega)$, and a remaining mismatch in the wave numbers is proportional to $\alpha - 4$, $\alpha = 4$ corresponding to the exact-matching point.

The form in which Eqs. (12.1) and (12.2) are written assumes that they govern the wave propagation in the temporal domain, e.g., propagation in optical fibers. In the spatial domain, the same equations (12.1) and (12.2) govern time-independent light distributions in planar waveguides with the quadratic nonlinearity where τ is replaced by the transverse coordinate x. In the temporal domain, Eq. (12.1) implies that FH has *anomalous* dispersion, then δ is a relative dispersion coefficient at SH. In practically

all the physically realistic cases, the dispersion coefficient (defined as in Eqs. (12.1) and (12.2) has a tendency to decrease with decreasing wavelength, therefore one always has $\delta < 1$. Moreover, the optical media usually have a zero-dispersion point (ZDP), beyond which the dispersion becomes normal. If SH is located beyond ZDP, then $\delta < 0$. Contrary to this, in the spatial domain the second terms in Eqs. (12.1) and (12.2) represent diffraction rather than dispersion. Because the diffraction coefficient is the same for all the wavelengths, $\delta \equiv 1$ in this case.

In the general case ($\delta < 1$), the system of Eqs. (12.1) and (12.2) does not have any nontrivial symmetry. However, in the special case $\delta = 1$, corresponding to the model in the spatial domain, the system is invariant with respect to the Galilean transformations with an arbitrary real parameter c

$$u \to u \exp\left(-\frac{1}{2}ic^2 z + ic\tau\right),\ v \to v \exp\left(-ic^2 z + 2ic\tau\right),\ \tau \to \tau - cz. \tag{12.3}$$

The purport of this invariance is that, given a particular stationary solution $u(\tau)$, $v(\tau)$ to Eqs.(12.1) and (12.2), a whole family of *moving* solutions can be generated by the transformation (12.3).

The next step is to introduce a three-wave (3W) model, which has two different FH components, u_1 and u_2, physically corresponding to orthogonal polarizations (the real SHG crystals are birefringent, one polarization being ordinary and the other one extraordinary). This modified model implies that two FH quanta with the orthogonal polarizations merge into one SH quantum with a fixed polarization. This type of the SHG process is called *type-II* interaction, while the simpler interaction corresponding to the model (12.1), (12.2), when two identical FH quanta merge into a SH one, is called *type-I* interaction. The 3W model is important because, in the real experiment, it is much easier to achieve nearly ideal matching between FH and SH, using the crystal's birefringence to cancel the wave number mismatch. A convenient form of the three-wave model is, in the simplest case,

$$i(u_1)_z + \frac{1}{2}(u_1)_{xx} - \beta u_1 + u_2^* v = 0, \tag{12.4}$$

$$i(u_2)_z + \frac{1}{2}(u_2)_{xx} - \frac{1}{\beta}u_2 + u_1^* v = 0, \tag{12.5}$$

$$2iv_z + \frac{1}{2}v_{xx} - \alpha v + u_1 u_2 = 0, \tag{12.6}$$

where the real coefficient β measures the birefringence [10] (see also [11]). There are more general versions of the 3W model, which correspond to a situation when the three waves have different frequencies $\omega_{1,2,3}$ and wave numbers $k_{1,2,3}$, subject to the resonant conditions $\omega_1 + \omega_2 - \omega_3 = 0$, $k_1 + k_2 - k_3 = q$, q being a small wave number mismatch. These general 3W models were introduced in [12].

As it was already mentioned above, the fundamental physical condition providing for an effective SHG process is the wavenumber matching. A way to achieve it is to use a system with a stronger temporal dispersion: in that case, a small change of the FH frequency ω gives rise to a larger change in the mismatch $k(2\omega)-2k(\omega)$, helping one to find a point where the mismatch nearly vanishes. It is well known that, if the intrinsic dispersion of a fiber or waveguide is weak, it can be strongly enhanced by a linear coupling between counterpropagating waves. Physically, this can be realized in terms of a grating written on the fiber or waveguide, that gives rise to the resonant (Bragg) scattering of light. A four-wave model combining SHG and the Bragg scattering was put forward independently in [13] and [14]. Following the notation adopted in [13], the model can be cast into the following form

$$\left(i\frac{\partial}{\partial t}+i\frac{\partial}{\partial z}+\omega\right)u_{+}+u_{+}^{*}v_{+} = -u_{-}, \tag{12.7}$$

$$\left(i\frac{\partial}{\partial t}-i\frac{\partial}{\partial z}+\omega\right)u_{-}+u_{-}^{*}v_{-} = -u_{+}, \tag{12.8}$$

$$\left(ic\frac{\partial}{\partial t}+i\frac{\partial}{\partial z}+q+2c\omega\right)v_{+}+u_{+}^{2} = -\kappa v_{-}, \tag{12.9}$$

$$\left(ic\frac{\partial}{\partial t}-i\frac{\partial}{\partial z}+q+2c\omega\right)v_{-}+u_{-}^{2} = -\kappa v_{+}\,, \tag{12.10}$$

where the subscripts $\pm$ pertain to the amplitudes of the right- and left-propagating waves, the reference frequency ω is introduced in order to obtain a time-independent soliton solution later, q is the mismatch parameter, and c is the SH group velocity relative to FH. The terms on the right-hand side account for the linear coupling induced by the Bragg scattering, the coefficient κ taking into regard the fact that the FH and SH coupling coefficients are different. The intrinsic dispersion (or diffraction) is neglected as the effective dispersion generated by the linear coupling is much stronger. The spectrum of the linearized version of these equations contains a gap, inside which solitons (the so-called *gap solitons*, GS's) may exist, that is why models of this type are frequently called GS models.

A simpler 3W model that combines the SHG, Bragg scattering, and diffraction can also be formulated [15]

$$i(u_1)_z+i(u_1)_x+u_2^{*}v = -u_2, \tag{12.11}$$

$$i(u_2)_z-i(u_2)_x+u_1^{*}v = -u_1, \tag{12.12}$$

$$2iv_z+Dv_{xx}-qv+u_1u_2 = 0. \tag{12.13}$$

This model is formulated in terms of the stationary distribution of light in a planar waveguide, the *walkoff* terms $\pm(u_{1,2})_x$ corresponding to different propagation directions of the two FH waves. It is assumed that their wave vectors make equal angles with the Bragg grating, while the SH carrier

wave vector is parallel to the grating, therefore the SH wave does not feel the grating, and the diffraction term with an effective coefficient D must be kept for this harmonic.

Another interesting possibility is to consider a system consisting of two parallel $\chi^{(2)}$ waveguides. If the separation between them is small enough, the light can *couple* (tunnel) between the two cores. In this case, the coupling between the two subsystems is also linear, and the corresponding four-wave model (also written in the spatial domain) is [16]

$$iu_{1z} + i\delta u_{1x} + \frac{1}{2}u_{1xx} - u_1 + u_1^* v_1 = -Qu_2, \tag{12.14}$$

$$2iv_{1z} + 2i\delta v_{1x} + \frac{1}{2}v_{1xx} - \alpha v_1 + \frac{1}{2}u_1^2 = -Kv_2, \tag{12.15}$$

$$iu_{2z} - i\delta u_{2x} + \frac{1}{2}u_{2xx} - u_2 + u_2^* v_2 = -Qu_1, \tag{12.16}$$

$$2iv_{2z} - 2i\delta v_{2x} + \frac{1}{2}v_{2xx} - \alpha v_2 + \frac{1}{2}u_2^2 = -Kv_1, \tag{12.17}$$

where α is the usual mismatch parameter (cf. Eqs. (12.1) and (12.2)), Q and K are the FH and SH linear coupling coefficients, and the walkoff parameter δ (that can be omitted in the simplest version of the model) takes into regard a possible misalignment of the light beams in the two waveguides.

2D and 3D models with the quadratic nonlinearity are of great interest too, due to the above-mentioned fact that they are collapse-free [6]. The most fundamental one is a straightforward multidimensional generalization of the simplest 1D model (12.1), (12.2), which was proposed and studied in detail in the recent works [29]

$$iu_z + \nabla_\perp^2 u + u_{\tau\tau} - u + vu^* = 0\,, \tag{12.18}$$

$$2iv_z + \nabla_\perp^2 v + \delta v_{\tau\tau} - \alpha v + \frac{1}{2}u^2 = 0\,, \tag{12.19}$$

where the operator $\nabla_\perp^2$ acts on the transverse coordinate(s). A very interesting peculiarity of the model is its *spatiotemporal anisotropy*: while one can define a spatiotemporal Laplacian $\nabla_\perp^2 + \partial^2/\partial\tau^2$ in Eq. (12.18), and the corresponding polar coordinates, these coordinates are not relevant for Eq. (12.19), as, in any physically realistic case, $\delta < 1$. Moreover, as it was mentioned above, the relative dispersion coefficient δ may become negative, in which case Eq. (12.19) will be of the hyperbolic type, while Eq. (12.18) is elliptic.

Exact soliton solutions to the above models are very rare. Therefore, most studies of solitons are based on direct numerical simulations. The only general analytical approach relies on the *variational approximation* (VA), see the next section. VA is based on a Lagrangian representation

of the underlying equations. For instance, Eqs. (12.18) and (12.19) can be derived from the Lagrangian

$$L = \int [i\left(u^* u_z + 2v^* v_z\right) - \frac{1}{2}\left(|u_\tau|^2 + \delta|v_\tau|^2\right) - \left(|u|^2 + \alpha|v|^2\right) - \frac{1}{2}\left(|\nabla_\perp u|^2 + |\nabla_\perp v|^2\right) + u^2 v^*] d\tau d\mathbf{r}_\perp + \text{c.c.} \quad (12.20)$$

Another important global feature of the same model is its dynamical invariant (integral of motion),

$$E = \int \left(|u|^2 + 4|v|^2\right) d\tau d\mathbf{r}_\perp. \quad (12.21)$$

This quantity has the physical meaning of the net number of quanta in a given wave pulse. In nonlinear optics, it is usually called "energy" or (in the context of the $\chi^{(2)}$ models) also the Manley-Row invariant.

The next section will summarize basic results obtained for solitons in the $\chi^{(2)}$ models. Before proceeding to this, it is relevant to mention simpler cw (continuous wave) solutions, e.g., $u = A\exp(ikz - i\omega\tau)$ and $v = B\exp(2ikz - 2i\omega\tau)$ in the case of Eqs. (12.1) and (12.2). Here, the amplitude A and the frequency ω (for instance) can be chosen as arbitrary parameters, then the remaining parameters B and k are to be found from the equations. An important and relatively simple problem is analysis of the *modulational instability* (MI) of the cw solutions, i.e., their instability against perturbations that make $|u|$ and $|v|$ different from a constant. For the simplest model (12.1), (12.2), the MI was investigated in [17]. The result is that *all* the cw solutions are unstable in the case $\delta > 0$. In the case $\delta < 0$, when the FH and SH dispersions are opposite, *stable* cw solutions are possible. Very recently, the MI was also studied in the multidimensional model (12.18), (12.19) [18]. In this case, stable cw solutions are not possible even in the case $\delta < 0$.

3 The solitons

In the limit of a large mismatch, $\alpha \gg 1$, the SH field can be eliminated: as it follows from Eq. (12.2), $v \approx u^2/2\alpha$, and substitution of this expression into Eq. (12.1) transforms it into the nonlinear Schrödinger (NLS) equation, $iu_z + (1/2)u_{\tau\tau} - u + (1/2\alpha)|u|^2 u = 0$, which has commonly known soliton solutions. More consistently, in the case $\alpha \gg 1$ a stationary soliton solution to Eqs. (12.1) and (12.2) can be obtain in the form of an expansion in powers of $1/\alpha$, starting from the NLS soliton (see, e.g., [20]; hereafter, the case $\delta = 1$ is considered)

$$u = \pm 2\sqrt{\alpha}\left[\mathrm{sech}(\sqrt{2}\tau) + 2\alpha^{-1}\sinh^2(\sqrt{2}\tau)\mathrm{sech}^3(\sqrt{2}\tau) + ...\right], \quad (12.22)$$

$$v = 2\mathrm{sech}^2(\sqrt{2}\tau) - 4\alpha^{-1}\left[5 - 4\cosh^2(\sqrt{2}\tau)\right]\mathrm{sech}^4(\sqrt{2}\tau) + ... \,. \quad (12.23)$$

The first soliton solution to Eqs. (12.1), (12.2) was found in 1974 for the particular case equivalent to $\alpha = \delta > 0$ [5]

$$u = \pm 3\sqrt{\delta}\mathrm{sech}^2\left(\tau/\sqrt{2}\right), \; v = \frac{3}{2}\mathrm{sech}^2\left(\tau/\sqrt{2}\right). \quad (12.24)$$

This is, in fact, the *single* soliton solution to all the models described in the previous section that can be found in an exact form. All the other solutions should be sought for numerically. In particular, the model (12.1), (12.2) has a family of stationary real solutions $u(\tau)$, $v(\tau)$ (α is a parameter of the family) [19]. However, despite the lack of the exact solutions, quite an effective analytical approximation was developed on the basis of VA in [20]. The main ingredient of VA is to adopt an *ansatz*, i.e., a trial functional form of the solution sought for. As is well known, the only tractable ansatz that admits the two components of the soliton to have *different widths*, which is an obvious property of the general solution, should be based on Gaussians

$$u = A\exp\left(-2a\tau^2\right), \; v = B\exp\left(-2b\tau^2\right), \quad (12.25)$$

where, A, B and a,b are indefinite parameters. The ansatz is to be inserted into the integral representation of the Lagrangian (the 1D version of (12.20)), after which the integration over τ in the Lagrangian can be performed explicitly, leading to an *effective Lagrangian*, which is a function of the indefinite parameters. Finally, these parameters are determined by the variational equations derived from the effective Lagrangian.

In the case of the underlying model (12.1), (12.2) and the ansatz (12.25), these equations can be cast into the following form [20] (which is, in fact, a generic example of what can be obtained by means of VA for the $\chi^{(2)}$ solitons)

$$\begin{aligned} A^2 &= (ab)^{-1/2}(1+a)(2a+b)(b+\alpha), \\ B &= (2a)^{-1/2}(1+a)\sqrt{2a+b}, \\ b &= 4a^2(1-a)^{-1}, \end{aligned} \quad (12.26)$$

while the remaining parameter a is determined by a cubic equation

$$20a^3 + (4-3\alpha)a^2 + 4\alpha a - \alpha = 0. \quad (12.27)$$

Meaningful solutions must satisfy conditions $A^2 > 0$ and $a, b > 0$. It then follows from (12.26) that only the solutions to Eq. (12.27) with $a < 1$ should be kept.

It is interesting to note that Eq. (12.27) has an exact solution $a = 1/5$ at $\alpha = 1$, i.e., just in the case when the particular exact solution (12.24) is available. According to Eqs. (12.26), in this case the amplitudes of the FH and SH components of the soliton predicted by VA are $A = 6\sqrt{3}/5 \approx 2.08, B = 3\sqrt{6}/5 \approx 1.47$, while the exact solution yields the amplitudes $u(0) = 3/\sqrt{2} \approx 2.12, v(0) = 1.5$. Thus, the accuracy provided by VA is quite acceptable, and this holds in the general case, as is illustrated by Fig. 1. Simultaneously, these figures display basic features of the general family of the stationary soliton solutions to Eqs. (12.1), (12.2).

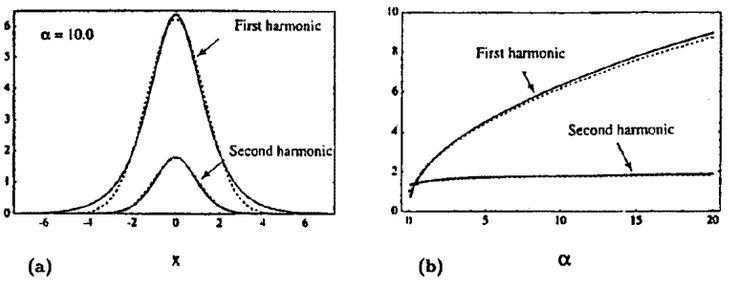

FIGURE 1. The stationary soliton solutions to Eqs. (12.1), (12.2): (a) Comparison between the shapes of the FH and SH components of the soliton, as obtained numerically (solid curves) and produced by the variational approximation (dashed curves) at $\alpha = 10$; (b) FH and SH amplitudes (peak values) vs. the mismatch α, as found numerically (solid curves) and predicted by the variational approximation (dashed curves).

The next fundamental problem is stability of these solitons. Numerical simulations reported in many papers demonstrate that, in almost all the cases, the solitons are indeed robust, and they can readily self-trap from localized initial pulses in simulations of the full equations (12.1) and (12.2). Detailed stability analysis of the solitons was presented in [21] , based on direct consideration of the eigenmodes of infinitesimal perturbations around the soliton. The final result is that, in the case $\delta = 1$, the stationary solitons are *stable* at $\alpha > \alpha_0 \approx 0.212$, and unstable in the opposite case, i.e., in fact, they are unstable only at a large negative mismatch.

Note that there is a general *Vakhitov-Kolokolov* (VK) stability criterion for solitons in a vast class of models of the NLS type [22]. If the soliton has the form $u(z, \tau) = \exp(ikz)\ U(\tau)$ (or obviously modified in multicomponent models, e.g., the $\chi^{(2)}$ one), its propagation constant k is a function of the energy $E = \int_{-\infty}^{+\infty} |U(\tau)|^2 d\tau$, or the properly redefined energy in the

multicomponent model (e.g., (12.21)). The VK criterion states that a *necessary* condition for the stability of the soliton is $dk/dE > 0$. In the form in which Eqs. (12.1) and (12.2) are written, the propagation constant of the stationary soliton was transformed into the mismatch parameter α. It is possible to reformulate the model so that the propagation constant will appear explicitly. The eventual result is that the stability region of the solitons predicted by the VK criterion *exactly* coincides with the above-mentioned one ($\alpha > 0.212$) produced by the direct analysis [21]. This fact strongly suggests that the VK criterion, having a fairly simple form, may be a very efficient tool for the analysis of stability of the $\chi^{(2)}$ solitons.

Besides the simple single-humped solitons illustrated by Fig. 1, the model (12.1), (12.2) and its generalizations also have more complicated double-humped localized solutions, that can be interpreted as bound states of the solitons [23]. However, it proves that, except for the GS models (see below), the two-soliton states are always dynamically unstable. Another class of solutions that were also studied but turn out to be unstable (because of the modulational instability of the cw background, see above) are *dark solitons* and their bound states [23].

In the case when the solitons are stable, it is interesting to consider their internal vibrations. From the *linearized* version of Eqs. (12.1) and (12.2), it follows that two branches of the dispersion relations for the small perturbation $\sim \exp(ikz - i\omega\tau)$ are

$$k_1 = -\left(\frac{1}{2}\omega^2 + 1\right),\ k_2 = -\frac{1}{2}\left(\frac{1}{2}\omega^2 + \alpha\right). \tag{12.28}$$

If the soliton has a nontrivial eigenmode of internal vibrations, the corresponding eigenvalue $\tilde{k}$ of the propagation constant must be separated from the continuum spectra (12.28), otherwise the mode is subject to strong damping through resonant emission of radiation. A detailed numerical analysis of the linearized problem developed in [24] has demonstrated that, in the stable region $\alpha > 0.212$, the soliton in the model (12.1), (12.2) with $\delta = 1$ always has exactly *one* nontrivial eigenmode. Moreover, the loss of the stability of the soliton at the critical point $\alpha \approx 0.212$ can be naturally explained by a bifurcation that happens to this eigenmode: its squared eigenvalue $\tilde{k}^2$ changes the sign at the critical point. Another noteworthy result of [24] is that the same soliton also has a *quasi*mode, which is a mode whose propagation eigenvalue is *inside* the continuum spectrum (12.28), but quite close a spectrum's border. Although, theoretically, this quasimode must quickly decay into radiation, in reality direct simulations demonstrate that the decay is extremely slow, so that the quasimode is practically stable. Moreover, the eigenvalue $\tilde{k}$ of the genuine eigenmode also turns out to be close to the border of the continuum spectrum, lying on its other side, and, as a result, the simulations demonstrate very long-lived beatings between the genuine mode and quasimode. A general

inference concerning the internal vibrations of the stable soliton is that the vibrations are amazingly robust, showing almost no radiative damping.

Above, only real z-independent stationary solutions to Eqs. (12.1), (12.2) were considered, which form a one-parameter family with the parameter α. A natural generalization is to extend it to a *two*-parametric family, the second parameter being a soliton's inverse velocity c in the case of the temporal solitons, or a spatial *walkoff* in the case of the spatial solitons:

$$u(z,\tau) = e^{ikz}U(\tau - cz),\ v(z,\tau) = e^{2ikz}V(\tau - cz).$$

As it was mentioned above, this generalization is trivial only in the particular case $\delta = 1$, when Eqs. (12.1) and (12.2) have the invariance (12.3). In the case $\delta \neq 1$ (corresponding, actually, to the temporal solitons), the generalization was considered in detail in the works [25]. It was found that these "walking" solitons exist in a broad region of parameters, and they are stable almost everywhere, except for a narrow stripe.

Once moving solitons are available, it is natural to consider collisions between them. This was done, in the work [26], only in the case $\delta = 1$, corresponding to the spatial solitons. The "moving" solitons were generated by means of the Galilean transformation (12.3) which is valid in this case. For the spatial solitons (optical beams), the collision problem is practically important, as it is related to designing schemes for all-optical switching and steering. A general conclusion obtained in [26] on the basis of simulations of Eqs. (12.1) and (12.2) is that, if the relative velocity of the colliding solitons exceeds a certain critical value, the collision is quasielastic, i.e., the solitons reappear after the collision practically unscathed; below the critical value, the collision is strongly inelastic, resulting in fusion of the solitons into a single pulse with very strong internal vibrations, similar to the vibrating soliton described above. This conclusion is typical for collisions of solitary pulses in *non*integrable nonlinear-wave models [27].

Solitons in the 3W model (12.4) - (12.6) have also been studied in detail, including the existence and stability problems, collisions between the solitons, etc. [10, 11]. In particular, application of VA, using a Gaussian ansatz similar to (12.25), to the 3W solitons yields, as well as in the case of the model (12.1), (12.2), very accurate results as compared to numerically generated shapes [10], and the VK stability criterion ([22]) can be properly adapted to the 3W model [11]. In a limited range, a bistability of the 3W solitons was discovered in [11].

The study of soliton solutions to the GS model (12.7)-(12.10) [13] reveals noteworthy features: the gaps where the solitons may exist turn out to be only *partly* filled with the actual soliton solutions, and the simulations demonstrate that two-humped solitons may be both unstable and *stable* (while they are always unstable in the other $\chi^{(2)}$ models), see Fig. 2. An analytical approach to this model is available in the limit when the SH coupling coefficient κ in Eqs. (12.9), (12.10) is large, allowing one to eliminate the SH fields, $v_{\pm} \approx -\kappa^{-1}u_{\pm}^2$. Substitution of this into the FH

equations (12.7) and (12.8) leads to coupled equations with a cubic nonlinearity, which resemble the Thirring model. An analytical (single-humped) soliton solution can be obtained in this approximation. The analytical approximation shares the above-mentioned feature of the numerical solutions, according to which the gap available for the solitons is only partly filled by the solutions.

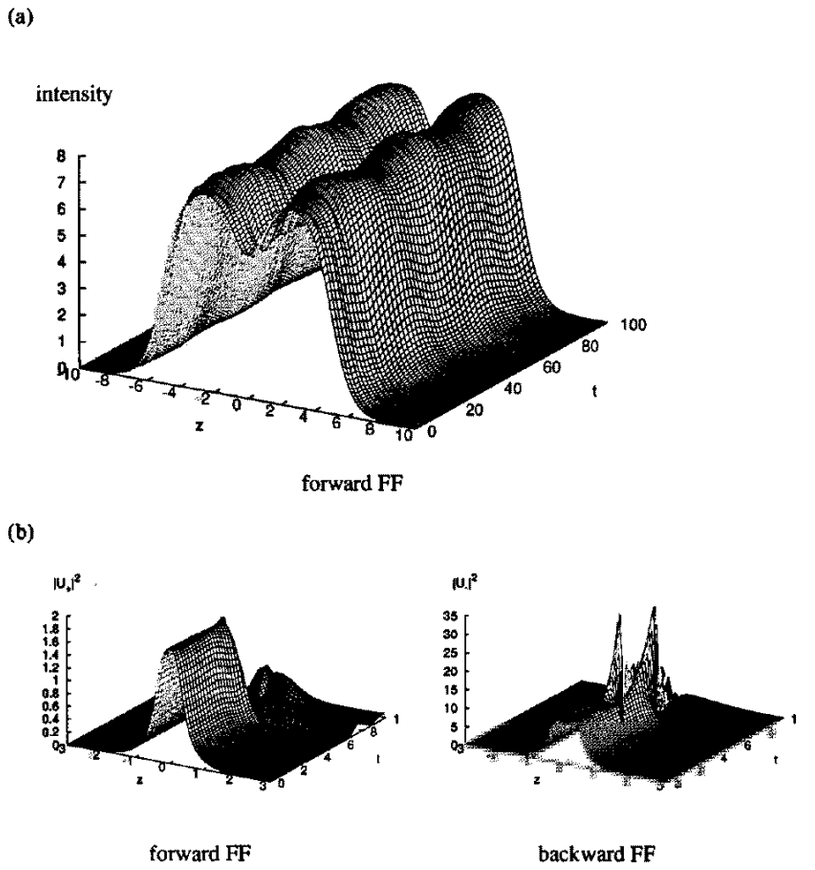

FIGURE 2. The stationary soliton solutions to Eqs. (12.1), (12.2): (a) Examples of stable (a) and unstable (b) two-humped solitons revealed by simulations of the gap-soliton model (12.7)-(12.10).

The properties of the $\chi^{(2)}$ gap solitons can be further clarified within the framework of the 3W model (12.11)-(12.13) (see also [28]). In this model,

a simple VA based on the ansatz

$$\begin{aligned} u_{1,2} &= \pm e^{ikz}\left[A_{\rm r}\,\mathrm{sech}\,(\mu x)\,\pm iA_{\rm i}\sinh(\mu x)\,\mathrm{sech}^2(\mu x)\right], \\ v &= Ve^{2ikz}\,\mathrm{sech}\,(\mu x) \end{aligned} \tag{12.29}$$

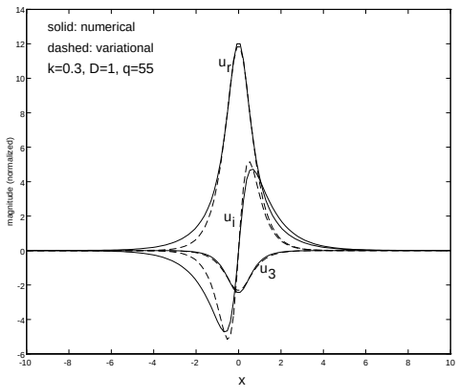

FIGURE 3. Typical examples of *stable* gap solitons in the three-wave model (12.11)-(12.13): a single-humped soliton, the solid and dashed curves showing, respectively, the results generated by the variational approximation and by direct simulations ($u_{\rm r}$ and $u_{\rm i}$ are the real and imaginary parts of the FH component, and u_3 is the SH component of the soliton, see Eq. (12.29)). The values of the parameters are: $k = 0.3$, $D = 1$, $q = 55$.

where $A_{\rm r,i}$, B and μ are free real parameters, cf. Eqs. (12.25), while k is a given propagation constant, yields predictions that compare fairly well with the direct simulations, see Fig. 3 [15]. Note that the solitons of this type are very different from those illustrated by Fig. 1a. The 3W gap solitons exist in the wavenumber interval $|k| < 1$; as $k \to +1$, they disappear developing a singularity, i.e., via a *collapse*. In the same 3W model, direct simulations yield a family of two-humped solitons, see Fig. 4. The single- and double-humped solitons have their own (actually, quite close) stability borders: they are stable if the mismatch is not too small. However, the stability borders do not exist with $k < 0$, as *all* the single- and two-humped solitons are stable in this case. At a very large mismatch, the two-humped solitons disappear via a collapse.

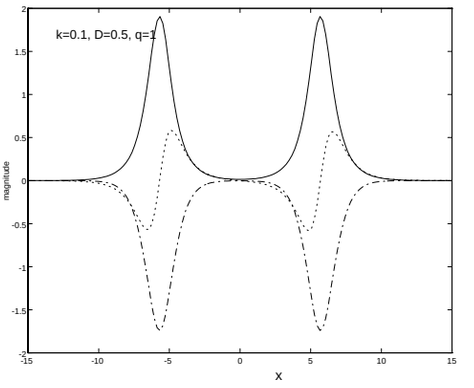

FIGURE 4. Typical examples of *stable* gap solitons in the three-wave model (12.11)-(12.13): a double-humped soliton (a bound state of two single-humped solitons), the solid and dotted curves being the real and imaginary parts of the FH field, while the dot-dashed curve is the real SH field. The values of the parameters are: $k = 0.1$, $D = 0.5$, $q = 1$.

4 Conclusion

Due to the length limitations, several important results are left aside, but detailed expositions are available in the original works. In the model (12.4)-(12.17) of the two linearly coupled $\chi^{(2)}$ waveguides, an effect that has no analog in the other $\chi^{(2)}$ models is a bifurcation that makes obvious symmetric solitons unstable and creates, instead, *asymmetric* solitons with $u_1 \neq u_2$, $v_1 \neq v_2$. In particular, in the most important case, $K = Q$, it can be shown that, at $q = 1$ (this is the case when Eqs. (12.1) and (12.2) have the exact solution (12.24)), the asymmetric solitons exist exactly in the interval $-1 < Q < +5/13$. In the other cases, the asymmetric solitons can be found with a high accuracy by means of VA. In all the cases when the asymmetric solitons exist, they are found to be stable. Further details are given in [16].

In the 2D and 3D models (12.18, (12.19), both VA and direct simulations produce stable fully localized multidimensional solitons, or *light bullets*, provided that $\delta \geq 0$ (*including* the case $\delta = 0$). In particular, the simple variational ansatz (cf. Eq. (12.25)) $u = A \exp\left(-ar_\perp^2 - \aleph\tau^2\right)$, $v = B \exp(-br_\perp^2 - \beta\tau^2)$, where $r_\perp$ is the radial variable in the transverse plane, yields very reasonable analytical results, if compared to the direct simulations, and the VK stability criterion ([22]) proves also to be relevant. In the case $\delta < 0$, the light bullets do not exist, slowly decaying into radiation. More details can be found in the works [29]. Very recently, the first experimental observation of a light bullet was reported in [31]. In this experiment,

the bullet was created in a bulk SHG sample of size 1 cm. The bullet was localized in the longitudinal (i.e., temporal) and one transverse directions, but delocalized in the other transverse direction. Thus, this spatiotemporal soliton may be regarded as a quasi-two-dimensional light bullet. The work aimed at generation of a fully localized three-dimensional light bullet in the SHG material is now in progress [32].

Still another interesting modification of the theory is soliton-like states in discrete lattices with the $\chi^{(2)}$ nonlinearity, which represent a stack of parallel waveguides, the field being cw inside each core. Many results, including the so-called staggered and unstaggered solitons, motion and collisions of the solitons in the lattices, etc., can be found in [30].

Acknowledgments: I appreciate valuable discussions of various issues considered in this review with P. Drummond, Y. Kivshar, F. Lederer, W. Mak, U. Peschel, L. Torner, and F.W. Wise.

5 References

[1] G.P. Agrawal. *Nonlinear Fiber Optics* (Academic Press: San Diego, Boston, New York, 1995); A. Hasegawa and Y. Kodama. *Solitons in Optical Communications* (Oxford University Press: Oxford (UK), 1995).

[2] G.I. Stegeman, D.J. Hagan, L. Torner, Opt. Quant. Electr. **28**, 1691 (1996).

[3] Y.R. Shen. *Principles of Nonlinear Optics* (Wiley: New York, 1984).

[4] D.J. Kaup, Stud. Appl. Math. **59**, 25 (1978).

[5] Yu.N. Karamzin and A.P. Sukhorukov, Pis'ma ZhETF **20** , 734 (1974) [JETP Lett. **20**, 339 (1974)]; Zh. Eksp. Teor. Fiz. **68**, 834 (1975) [JETP Lett. **41**, 414 (1976)].

[6] A.A. Kanashov and A.M. Rubenchik, Physica D **4**, 122 (1981).

[7] W. E. Torruellas, Z. Wang, D. J. Hagan, E. W. Van Stryland, G. I. Stegeman, L. Torner, and C. R. Menyuk, Phys. Rev. Lett. **74**, 5036 (1995).

[8] R. Schiek, Y. Baek, and G. I. Stegeman, Phys. Rev. E **53**, 1138 (1996).

[9] P. Di Trapani, D. Caironi, G. Valiulis, A. Dubietis, R. Danielius, and A. Piskarskas, Phys. Rev. Lett. **81**, 570 (1998).

[10] U. Peschel, C. Etrich, F. Lederer, and B.A. Malomed, Phys. Rev. E **55**, 7704 (1997).

[11] A. V. Buryak, Yu. S. Kivshar, and S. Trillo, Phys. Rev. Lett. **77**, 5210 (1996).

[12] H. T. Tran, Opt. Commun. **118**, 581 (1995); B. A. Malomed, D. Anderson, and M. Lisak, Opt. Commun. **126**, 251 (1996).

[13] T. Peschel, U. Peschel, F. Lederer, and B.A. Malomed, Phys. Rev. E. **55**, 4730 (1997).

[14] C. Conti, S. Trillo, and G. Assanto, Phys. Rev. Lett. **78**, 2341 (1997); H. He and P.D. Drummond, Phys. Rev. Lett. **78**, 4311 (1997).

[15] W.C.K. Mak, B.A. Malomed, and P. L. Chu, Phys. Rev. E **57**, 1092 (1998), and **58**, 6708 (1998).

[16] W.C.K. Mak, B.A. Malomed, and P. L. Chu, Phys. Rev. E **55**, 6134 (1997), and **57**, 1092 (1998).

[17] S. Trillo and P. Ferro, Opt. Lett. **20**, 438 (1995); H. He, P.D. Drummond, and B.A. Malomed, Opt. Commun. **123**, 394 (1996).

[18] Z.H. Musslimani and B.A. Malomed, Physica D **123**, 235 (1998).

[19] A.V. Buryak and Yu.S. Kivshar, Opt. Lett. **19**, 1612 (1994).

[20] V.V. Steblina, Yu.S. Kivshar, M. Lisak, and B. A. Malomed, Opt. Commun. **118**, 345 (1995).

[21] D.E. Pelinovsky, A.V. Buryak, and Yu.S. Kivshar, Phys. Rev. Lett. **75**, 591 (1995).

[22] M.G. Vakhitov and A.A. Kolokolov, Radiophys. Quantum Electr. **16**, 783 (1973).

[23] A.V. Buryak and Yu.S. Kivshar, Phys. Lett. A **197**, 407 (1995); H. He, M.J. Werner, and P.D. Drummond, Phys. Rev. E **54**, 896 (1996).

[24] C. Etrich, U. Peschel, F. Lederer, B.A. Malomed, and Yu. S. Kivshar, Phys. Rev. E **54**, 4321 (1996).

[25] L. Torner, D. Mazilu, and D. Mihalache, Phys. Rev. Lett. 77, 2455 (1996); C. Etrich, U. Peschel, F. Lederer, and B.A. Malomed, Phys. Rev. E **55**, 6155 (1997).

[26] C. Etrich, U. Peschel, F. Lederer, and B.A. Malomed, Phys. Rev. A **52**, R3444 (1995).

[27] Yu.S. Kivshar and B.A. Malomed, Rev. Mod. Phys. **61**, 763 (1989).

[28] S. Trillo, Opt. Lett. **21**, 1732 (1996).

[29] B.A. Malomed, P. Drummond, H. He, A. Berntson, D. Anderson, and M. Lisak, Phys. Rev. E **56**, 4725 (1997); D. Mihalache, D. Mazilu, B.A. Malomed, and L. Torner, Opt. Commun., **152**, 365 (1998); D. Mihalache, D. Mazilu, J. Dörring, and L. Torner, Opt. Comm. **159**, 129 (1999).

[30] O. Bang, P.L. Christiansen, and C.B. Clausen, Phys. Rev. E **56**, 7257 (1997); T. Peschel, U. Peschel, and F. Lederer, Phys. Rev. E **57**, 1127 (1998); S. Darmanyan, A. Kobyakov, E. Schmidt, and F. Lederer, Phys. Rev. E **57**, 3520 (1998); P.D. Miller and O. Bang, Phys. Rev. E **57**, 6038 (1998).

[31] X. Liu, L.J. Qian, and F.W. Wise, Phys. Rev. Lett. **82**, 4631 (1999).

[32] F.W. Wise, private communication.

13

Nonlinear Models for the Dynamics of Topological Defects in Solids

Yuri S. Kivshar
Hartmut Benner
Oleg M. Braun

ABSTRACT We discuss several physical systems, where the nonlinear dynamics of topological defects is described by quasi-one-dimensional kink solutions of the generalized Frenkel-Kontorova model and its continuous approximations (the sine-Gordon equation). This includes dislocations, long Josephson junctions, magnetic chains, adsorbed layers of atoms, hydrogen-bonded chains, DNA-type chains, etc. We briefly review different properties of kinks and describe experimental verifications of kink dynamics.

1 Introduction

Theoretical physics deals with physical models. One of the main requirements for a good physical model is simplicity. Universal models which can be applied to describe a variety of effects of different physical origin are rare and therefore of key importance. Such models attract special attention and can be employed to describe the basic physical concepts in the simplest way. A classical example is the theory of linear oscillations based on the model of a mathematical pendulum – this universal model serves as a paradigm with basic features of many different physical systems.

In solids, which are described by many degrees of freedom, all microscopic models are quite complicated. However, a very simple model, which describes a chain of classical particles harmonically coupled to their nearest neighbours and subjected to a periodic on-site potential, has become one of the fundamental and universal models of *low-dimensional nonlinear physics*. In spite of the fact that a link with the classical problem is not often stated explicitly in various applications, many kinds of nonlinear models involving the dynamics of discrete nonlinear chains are based (directly or indirectly) on the classical formulation introduced six decades ago by Frenkel and Kontorova (1938; 1939), who described the structure and dynamics of a crystal lattice in the vicinity of a dislocation core. As a matter of fact, this was the first example in solid-state physics when the

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 265−291, 2000.

dynamics of an extended two-dimensional defect in a bulk was described by a simple one-dimensional model. Similar ideas were employed in many different models of low-dimensional solid-state physics, providing a link with the theory of solitary waves developed much later.

The simplicity of the Frenkel-Kontorova (FK) model, due to the assumptions of linear inter-atomic forces and a sinusoidal on-site (substrate) potential, as well as its surprising capability to describe a broad spectrum of physically important nonlinear phenomena, such as propagation of charge-density waves, the dynamics of absorbed layers of atoms on crystal surfaces, commensurate-incommensurate phase transitions, domain walls in magnetically ordered structures, etc., has attracted a great deal of attention from physicists working in solid-state physics and nonlinear physics.

One of the important features that can explain *why* the FK model has attracted much attention in different branches of solid state physics is the fact that in the continuum-limit approximation the model reduces to the exactly integrable sine-Gordon (SG) equation, which has a number of unique properties and allows exact solutions for nonlinear waves and their interaction. In particular, the SG equation yields an example of a fundamental nonlinear model for which we know almost everything about the dynamics of *nonlinear excitations*. As is known, the SG system has three different types of elementary excitations, namely *phonons, kinks* (topological solitons), and *breathers* (dynamical solitons), whose dynamics determines the general behaviour of the system as a whole. And, although the FK model is inherently *discrete* and not exactly integrable, one may get deep physical insight and may markedly simplify the understanding of its nonlinear dynamics using the language of SG quasi-particles as weakly interacting nonlinear excitations.

In this Chapter we review several different nonlinear models of solids from a rather general point of view, considering generalized (discrete and continuous) versions of the FK model and its topological excitations – kinks. We show how the FK model appears in the context of the theory of dislocations, Josephson junctions, surface physics, and also as a simplified model of hydrogen-bonded chains, ferro- and antiferromagnetic systems, DNA-type chains, etc. We also describe the basic properties of kinks and give some examples of experimental verifications of the kink dynamics.

2 The FK model and the SG equation

A simple model which describes the dynamics of a chain of particles (atoms) interacting with their nearest neighbours and subjected to a periodic on-site (substrate) potential was first mentioned, to the best of our knowledge, by Dehlinger (1929) and then independently introduced by Frenkel and Kontorova (Frenkel and Kontorova, 1938; 1939). It can schematically be

presented as a chain of particles on a periodic substrate potential. The dynamics of such a chain is described by the Hamiltonian

$$\mathcal{H} = \sum_n \left\{ \frac{m_a}{2} \left(\frac{dx_n}{dt} \right)^2 + V_{\rm int}(x_{n+1} - x_n) + V_{\rm sub}(x_n) \right\}, \tag{13.1}$$

where m_a is the particle mass, and x_n is the co-ordinate of the n-th particle. In the simplest case the interaction between neighbouring particles is assumed to be harmonic, $V_{\rm int}(x_{n+1} - x_n) = \frac{1}{2} g (x_{n+1} - x_n - a_0)^2$, where $g = V''_{int}(a_0)$ is the elastic constant, and a_0 is the equilibrium distance of the interatomic potential. Also the substrate potential is expanded into the Fourier series $V_{\rm sub}(x_n) = \frac{1}{2} U_0 [1 - \cos(2\pi x_n / a_s)]$, where U_0 is the maximum energy and a_s is the period.

After renormalization, the Hamiltonian (13.1) gives the equation of motion

$$\frac{d^2 x_n}{dt^2} + \sin x_n - g(x_{n+1} + x_{n-1} - 2x_n) = 0, \tag{13.2}$$

and we notice that this equation does not include the parameter a_0 corresponding to the equilibrium distance of the inter-atomic potential. In the following we only consider the commensurate case $a_0 = a_s$, where in the ground state each minimum of the substrate potential is occupied by one atom, so that we can introduce the variables $x_n = n a_s + u_n$.

Linear excitations (e.g. phonons or magnons), $u_n(t) \propto \exp[i\omega(k)t - ikn]$, are characterized by the dispersion law

$$\omega^2(k) = \omega_{\rm min}^2 + 2g(1 - \cos k), \tag{13.3}$$

where k is the wave number ($|k| \leq \pi$) and $\omega_{\rm min} = 1$. The spectrum is characterised by the cut-off frequency $\omega_{\rm max} = \sqrt{\omega_{min}^2 + 4g}$. When the particle displacements are not small, the linear approximation is no longer valid, and the model (13.2) can support nonlinear excitations, *kinks* and *breathers*. To show this, let us first consider the continuum limit of the FK model,

$$u_{tt} - u_{xx} + \sin u = 0, \tag{13.4}$$

where the spatial variable x is measured in units of $d = a_s \sqrt{g}$.

As a matter of fact, Eq. (13.4) was the first partial differential equation whose multi-soliton properties were recognized. Indeed, in a transformed form Eq. (13.4) was originally considered by A. Enneper (Enneper, 1870) in the differential geometry of surfaces of constant negative Gaussian curvature. The study of Eq. (13.4) in the context of differential geometry revealed very interesting properties, including the possibility to generate from one known solution new (unknown) solutions by means of the *Bäcklund transformation* (Bäcklund, 1882).

In physics, Eq. (13.4) has found its first application in dislocation models, and kinks and breathers of the SG equation have first been introduced

by A. Seeger and co-workers more than forty years ago (Kochendörfer and Seeger, 1950; Seeger and Kochendörfer, 1951; Seeger et al., 1953; see also Seeger, 1980; Döttling et al., 1990). The original German names for the kinks and breathers were "translatorische und oszillatorische Eigenbewegungen", and from a historical point of view it is interesting to note that this preceded the discovery of solitonic properties of the Korteweg-de Vries equation (Zabusky and Kruskal, 1965) by more than a decade. Perring and Skyrme (1962) introduced the SG equation as a simple one-dimensional model of the scalar field theory. Almost simultaneously, the SG equation appeared in the theory of weak superconductivity to be the main nonlinear equation describing the so-called long Josephson junctions (see e.g. Josephson, 1965), where a kink describes a quantum of magnetic flux, a *fluxon*. The two next important steps in the history of the SG equation were the emphasis of its pedagogical importance in terms of a very simple chain of coupled pendulums (the mechanical analog of the FK chain) introduced by A. Scott (1969), and the solution of the related inverse scattering transform problem obtained by M. Ablowitz *et al.* (1973). Later, the SG equation was proved to be completely integrable, and its properties have been described in many review papers and books (see e.g. Zakharov *et al.*, 1980). Here we only mention the main properties of Eq. (13.4) and its solutions.

Elementary excitations of SG systems are *phonons*, *kinks*, and *breathers*. *Phonons*, or continuous waves in the linear limit, take the standard form $u(x,t) \propto \exp(i\omega t - ikx)$ and are characterised by the dispersion relation, $\omega^2(k) = 1 + k^2$, which is a long-wave expansion of Eq. (13.3).

Kinks, or topological solitons, appear due to a degeneracy of the system ground state. Indeed, a kink can be understood as the solution connecting two nearest identical minima of the periodic on-site potential,

$$u_k(x,t) = 4\tan^{-1}\exp[-\sigma\gamma(v)(x - vt)], \tag{13.5}$$

where $\sigma = \pm 1$ stands for the *topological charge*. The kink velocity v is measured in units of the sound velocity c and determines the kink width, $\gamma(v) = 1/\sqrt{1 - v^2}$. The latter result follows from the relativistic invariance of the SG model and may be considered as a Lorentz contraction of the kink width. It allows one to introduce the rest mass of the kink, which in dimensionless units is written as $m = 2/(\pi^2\sqrt{g})$, and its rest energy, defined as $\epsilon_k = mc^2 = 8\sqrt{g}$.

A breather, or dynamical soliton, has the form

$$u_{\rm br}(x,t) = 4\tan^{-1}\left\{\left(\frac{\sqrt{1-\Omega^2}}{\Omega}\right)\frac{\sin(\Omega t)}{\cosh(x\sqrt{1-\Omega^2})}\right\}, \tag{13.6}$$

and describes a nonlinear state oscillating with the internal frequency Ω, which is located within the gap of the linear spectrum, $0 < \Omega < \omega_{\rm min}$, with

the amplitude $u_{\max} = 4\tan^{-1}(\sqrt{1-\Omega^2}/\Omega)$, the oscillation being localised to the spatial scale $\sim d/\sqrt{1-\Omega^2}$. The breather energy is $\epsilon_{\rm br} = 2\epsilon_k\sqrt{1-\Omega^2}$, so that $0 < \epsilon_{\rm br} < 2\epsilon_k$. The solution for a moving breather is easily obtained from Eq. (13.6) by means of the Lorentz transformation. In the limit of low frequencies, $\Omega \ll 1$, the breather can qualitatively be treated as a kink-antikink pair.

In the framework of the SG model kinks and breathers move freely along the chain without dissipation losses. Moreover, the only effect of their collision is a phase shift (see e.g. Zakharov *et al.*, 1980). That is why kinks and breathers can be treated as *nonlinear quasi-particles* of the SG model. Such an approach is still valid for nearly integrable modifications of the SG equation, when the model includes small perturbations, like those appearing when the SG equation is derived from the primary FK model in the quasi-continuum approximation, assuming discreteness effects to be small.

3 Physical models

The model described by Hamiltonian (13.1) appears in many problems of solid state physics and has a simple physical meaning. A complex system of interacting particles is treated as *two coupled sub-systems*; one is modelled by a chain of moving particles, provided its dynamics is quasi-one-dimensional, the other one is treated as a "frozen" substrate. The effective particles of the chain can model the dynamics of real atoms (e.g. in the theory of absorbed atoms on a crystal surface), spins (magnetic systems), clusters of atoms (the so-called bases of the DNA-like chains), etc. Including more degrees of freedom leads usually either to models describing a system of coupled chains, or to two-dimensional models. Below we discuss several examples where the SG equation and the FK model are the basic models for describing the nonlinear dynamics.

3.1 Dislocations in solids

The first application of the FK model was in the *theory of dislocations in metals* (Frenkel and Kontorova, 1938; 1939; Frank and van der Merwe, 1949a,b; Seeger and Kochendörfer, 1951; Atkinson and Cabrera, 1965). The importance of this problem cannot be overestimated, since dislocations are responsible for the mechanical properties of solids. The FK model was the first model that explained the dynamics of a dislocation core on an atomistic level and has resulted in simple formulas valid even quantitatively. (Note that a large-scale first-principle simulation of dislocation dynamics is a nontrivial problem even for modern parallel computers.)

In dislocation theory the FK model has a simple physical origin. Indeed, let us consider an additional semi-infinite plane of atoms inserted into a

perfect crystal lattice. After relaxation to the equilibrium state, the system has one *edge* dislocation. Then the layer of atoms perpendicular to the inserted plane divides the crystal into two different parts and plays the role of an interface layer. The atoms at the interface layer are subjected to an external potential resulting from the surrounding atoms of the lattice, and the interaction with these other atoms can be modelled by an effective periodic potential. This idea gave birth to the FK model.

First, the two-dimensional plane of atoms belonging to the interface layer can approximately be treated in the framework of a one-dimensional model as an atomic chain perpendicular to the edge dislocation line. In this model the dislocation core is described as a kink of the FK chain, and this explains the existence of activation barrier, the so-called Peierls-Nabarro (PN) barrier, for the dislocation motion. This barrier is known as the *primary* PN barrier.

The next step of modelling is to consider the dislocation line itself as an (other) FK chain placed in the external periodic potential which is nothing but the primary PN relief. The motion of the dislocation line is due to the creation of kink-antikink pairs, when a section of the dislocation is shifted to a neighbouring valley of the external (primary PN) potential. These kinks move in the *secondary* PN relief until they reach the end of the dislocation (the dislocation is a topological object and cannot be broken) or annihilate with another kink of opposite topological charge.

The main advantage of this phenomenology is that it predicts correctly the existence of the PN barrier which determines the dynamics of dislocations. Besides, plastic deformations in metals may also result from local defects known as *crowdion* (Paneth, 1950; Frenkel, 1972). The crowdion corresponds to a configuration where one extra atom is inserted into a closely packed row of atoms in a metal with ideal crystal lattice. In many such cases the crystalline potential is organized in such a way that the atoms can move only along the row direction, and the inserted atom (together with its neighbouring atoms) forms a one-dimensional configuration which can be treated as a kink of the FK model. Usually the size of such a configuration is about ten lattice units, so that the kink is characterised by a very small PN barrier and moves along the row almost freely. Therefore, the crowdions play an important role in plasticity of metals. Crowdion problems were recently studied by Landau et al. (1993) and Kovalev et al. (1993).

3.2 Magnetic chains

A kink in magnetic models describes a domain wall (DW), i.e. a localised region between two domains of magnetization (two equivalent ground states). The simplest one-dimensional classical model of a magnetic chain is de-

scribed by the following Hamiltonian (Mikeska and Steiner, 1991)

$$\mathcal{H} = -J \sum_n S_n S_{n+1} - D \sum_n (S_n^x)^2 \,. \tag{13.7}$$

The first term is the isotropic exchange energy and the second one the single-ion anisotropy. If J and D are positive, the system has the x-axis as the preferred direction in the spin space. For minimum energy all spins should have maximum x-components. This leads to a degeneracy of the ground state, and two equivalent ground states are connected by a static domain wall. Its form represents a compromise between the counteracting effects of exchange interaction J ($\Rightarrow$ wide DW) and anisotropy D ($\Rightarrow$ narrow DW). In the classical approximation spins are treated as vectors of length S, and one works in spherical co-ordinates given as $S_n = S(\sin\Theta_n \cos\Phi_n, \sin\Theta_n \sin\Phi_n, \cos\Theta_n)$. The continuum approximation leads to continuous variables given as $\Theta_n \to \Theta(z)$ and $\Phi_n \to \Phi(z)$, so that a static equation for $\Phi(x)$ defines the structure of the domain wall. In the case of a strong planar anisotropy and a transverse magnetic field, the equation of motion for the angle Φ results in the SG equation (Mikeska, 1978).

Considering a *ferromagnetic* (FM) chain with easy-plane anisotropy in a transverse magnetic field H at sufficiently low temperature, there exists one ground state where all spins are aligned in the H-direction. A soliton represents a 2π-twist of the ordered spin chain which can freely move along the chain without dispersion. In an *antiferromagnetic* (AFM) chain the ground state corresponds to a spin-flop configuration, i.e. the two sublattices, apart from a slight canting, are aligned antiparallel to each other and perpendicular to H. Therefore, there exist two degenerate, but topologically inequivalent ground states which differ by the sign of spin orientation. Thus, an AFM soliton represents a π-twist reversing both sublattices when passing along the chain. The difference in topology results in different spatial and temporal behaviour of the spin components and largely affects the experimental evidence. While in the FM case spin fluctuations only originate from the 2π-soliton itself, it is flipping of all spins *between* two neighbouring π-solitons (the so-called "flipping mode") which yields the essential contribution in the AFM case. In either case, spin fluctuations in the chain direction are largely suppressed by the strong planar anisotropy.

The SG model is valid provided the magnetisation vector rotates only in the easy plane. However, for $\lambda < 3$, where $\lambda = 2DS/\mu H$, the SG soliton is unstable against out-of-plane fluctuations. The system dynamics should be described by two coupled equations for two angles Θ and Φ (Kosevich et al., 1990).

3.3 Josephson junctions

Among many physical systems, one of the closest correspondence to the SG equation has been found in the problem of fluxon dynamics in long

Josephson junctions (JJs or Josephsons transmission lines). A SG kink in this case is often called a "fluxon" since it carries a magnetic flux quantum $\Phi_0 = h/2e = 2.07 \times 10^{-15}$ Wb moving between two superconducting electrodes. Several comprehensive reviews on fluxon dynamics in JJs have been published in the past (McLaughlin and Scott, 1978; Pedersen, 1986; Parmentier, 1993) and some of the new topics have been addressed more recently (Pedersen and Ustinov, 1995; Ustinov and Parmentier, 1996). This is an active area of research, especially after the discovery of high-T_c superconductors.

A fluxon in JJs carries a magnetic flux Φ_0 generated by a circulating supercurrent, often called Josephson vortex, which is located between two superconducting films separated by a few nanometers thin layer of insulator. As shown by McLaughlin and Scott (1978), the fluxon corresponds to a 2π-kink of the quantum phase difference φ between the two superconducting electrodes of the junction. The perturbed sine-Gordon equation which describes the dynamics of the system reads in normalized form

$$\varphi_{xx} - \varphi_{tt} - \sin\varphi = \alpha\varphi_t - \beta\varphi_{xxt} - \gamma. \tag{13.8}$$

Time t is measured in units of ω_0^{-1}, where ω_0 is the Josephson plasma frequency, the spatial co-ordinate x is measured in units of the Josephson penetration depth λ_J, α is dissipative term due to quasi-particle tunnelling (normally assumed ohmic), β is a dissipative term due to surface resistance of the superconductors, and γ is a normalised bias current density. To account for the behaviour of a real system, Eq. (13.8) must be solved together with appropriate boundary conditions which depend on the junction geometry and take into account the magnetic field applied in the plane of the junction.

A typical property of real JJs is that the parameters α, β and γ are rather small. An important solution of Eq. (13.8) with zero r.h.s. is the kink $\varphi = 4\tan^{-1} e^{\xi}$, where $\xi = (x - vt)/\sqrt{1 - v^2}$. The velocity v is determined by the balance between losses and input force, $v = 1/\sqrt{1 + (4\alpha/\pi\gamma)^2}$. The velocity saturates for large values of γ/α, and a fluxon behaves like a relativistic particle with its velocity close to the velocity c of linear waves in the junction, the Swihart velocity.

Fluxons in coupled JJs have recently become a subject of intensive theoretical and experimental investigations. The discovery of the intrinsic Josephson effect in some high-temperature superconductors such as $Ba_2Sr_2CaCu_2O_{8+y}$ (BSCCO) convincingly showed that these materials are essentially natural super-lattices of Josephson junctions formed on atomic scale (Kleiner *et al.*, 1992; Kleiner and Müller, 1994; Lee *et al.*, 1995). The spatial period of such a super-lattice is only 15 Å, so the Josephson junctions are extremely densely packed. The superconducting electrodes are formed by copper oxide bilayers as thin as 3 Å and are separated by the Josephson super-lattices with many active layers.

For the first time, the fluxon dynamics in two inductively coupled long JJs was considered theoretically by Mineev *et al.* (1981). The perturbation approach for small coupling has been further explored by Kivshar and Malomed (1988) and Grønbech-Jensen *et al.* (1990, 1993). A very important step towards quantitative comparison with real experiments was made by Sakai *et al.* (1993) who derived a model for arbitrary strong coupling between the junctions.

Finally, discreteness appears in the case of a parallel biased array of small JJs, which is described by a discrete SG equation,

$$\frac{dV_n}{dt} = \frac{1}{a^2}(\varphi_{n-1} - 2\varphi_n + \varphi_{n+1}) - \sin\varphi - \alpha V_n + \gamma; \quad \frac{d\varphi_n}{dt} = V_n, \quad (13.9)$$

where $1 \leq n \leq N$. These equations are just the Kirchhoff circuit-law equations for an array of N discrete JJ elements interconnected via a parallel resistance/inductance combination. The phases at the virtual points $n = 0$ and $n = N+1$ are defined through boundary conditions depending on the geometry of the array. This model has been analysed theoretically by Ustinov *et al.* (1993) and their predictions have later been confirmed experimentally by van der Zant *et al.* (1995).

3.4 Hydrogen-bonded chains

One more important model that can be described by the Hamiltonian (13.1) appears in the theory of proton conductivity of hydrogen-bonded chains. Hydrogen-bonded networks are quasi-one-dimensional clusters of molecular aggregates interacting with their first neighbours through hydrogen bonds. Schematically, this can be presented in the form ...X-H...X-H...X-H...X-H..., where the full line segments indicate covalent or ionic bonds, the dotted ones hydrogen bonds, and X a negative ion. The important idea of a simple physical model for such a nonlinear chain is based on the fact that protons move in a double-well potential resulting from hydrogen bonds with the heavy-ion lattice (oxygen lattice) which is assumed to be deformable (Antonchenko et al., 1983).

A local distortion of the oxygen lattice lowers the activation barrier for protons and thus promotes their motion. In order to describe this phenomenon, one- or two-component nonlinear models should include the proton sublattice which supports topological solitons (kinks), while the oxygen sublattice can be modelled as another sublattice or as an effective external potential for the proton motion. Several models of this kind have been proposed (Antonchenko et al., 1983; Zolotaryuk et al., 1984; Zolotaryuk, 1986; Peyrard et al., 1987; Hochstrasser et al., 1988; Pnevmatikos, 1988), and they give a simple and effective description of the proton mobility in hydrogen-bonded chains.

In the lowest-order approximation, the oxygen atoms are assumed to have fixed positions and to produce an effective substrate potential to the

mobile hydrogen atoms, for which a double SG equation with the effective on-site potential

$$V'_{\rm sub}(u) = V_0 \left[-\sin u + 2A \sin\left(\frac{u}{2}\right)\right] \tag{13.10}$$

can be derived (see, e.g., Pnevmatikos, 1988). Then, the mechanism of proton conductivity is explained by a migration transport of two types of defects, *ionic* and *bonding* (or Bjerrum) defects, along the chains. The defects are described by two types of kinks of the corresponding double SG model. A similar kind of *one-component* model for hydrogen-bonded systems is rather well investigated in the framework of the continuum approximation (see, e.g., Zolotaryuk and Pnevmatikos, 1990; Pnevmatikos et al., 1991; and references therein). At the same time, recent ideas in the theory of kink-induced proton conductivity involve more general properties of the FK type models, e.g. the discreteness of the proton chains and thermalized kink motion (Savin and Zolotaryuk, 1991), the effect of increased proton conductivity due to commensurate-incommensurate phase transitions (Christophorov and Gaididei, 1992), a complex chain of zigzag structure (Christiansen et al., 1997), a mass variation along the chain (Kalosakas et al., 1997), etc.

More rigorously, the dynamics of systems like hydrogen-bonded chains can properly be described by introducing *two interacting sublattices* for proton and oxygen atoms, respectively. In such a case, we should consider two-component generalisations of the FK model which describe two interacting chains of particles, one of them subjected to a substrate potential which is created by the other. Several models of this type have been introduced and studied in the continuum approximation (see e.g. Antonchenko et al., 1983; Zolotaryuk, 1984, 1986; Hochstrasser et al., 1988, Pnevmatikos, 1988). The dynamics of the two-component models has several new features in comparison with standard one-component models. For example, a new branch of the phonon spectrum appears in the band gap of linear excitations of the standard FK chain, so that the motion of kinks is stable only for small velocities which do not exceed the sound speed of acoustic phonons of the oxygen sublattice (see e.g. Zolotaryuk et al., 1984). As a matter of fact, the second phonon branch plays an important role in kink scattering by local impurities (Kivshar, 1991).

Similar two-component FK models describe more realistically the dynamics of other physical objects such as dislocations, crowdions, ad-atomic chains, chains of ions in superionic conductors, etc. In all such situations the second (heavy atom) subsystem corresponds to substrate atoms, so that the whole system may be treated again as a FK chain on a *deformable substrate*. A similar situation occurs for the physical models of molecular crystals and polymer chains as well as ferroelectric or ferroelastic chains where rotational and vibrational degrees of freedom are coupled (see e.g. Pouget and Maugin, 1984, 1985; Maugin and Miled, 1986, to cite a few).

3.5 Surface physics and adsorbed atomic layers

A broad area of different applications of the FK model is in surface science. The external substrate potential in the FK model is governed by the surface atoms of a crystal, while the atoms adsorbed on the surface are treated as those belonging to the chain. The adsorbed atoms are usually more mobile than the atoms of the substrate since the substrate atoms are trapped and vibrate near their equilibrium positions. Therefore, the approximation of a rigid substrate is adequate for this class of problems. In some cases, e.g. for the (112) plane of b.c.c. crystals, the surface atoms produce a furrowed potential. Then, the ad-atoms are located inside the furrows and, therefore, can be considered as one-dimensional chains. Another example are vicinal semiconductor surfaces, where the ad-atoms are adsorbed closely to the steps and, thus, are incorporated into the chains. In the latter case, the distance between the chains may be different depending on the angle of the vicinal surface chosen.

The models of adsorbed layers have several advantages when considered in the approximation of a FK chain. First of all, the assumption of a rigid substrate is a reasonable approximation. Then, the atomic concentration can be varied in a wide range, from $\theta = 0$ (adsorption of single atoms) to $\theta = 1$ and even more (when the "diameter" of the ad-atom is lower than the period of the external potential as, e.g., for lithium ad-atoms on transition metal surfaces). As a result, a large number of interesting physical phenomena can be analysed in terms of the FK model, including transitions between different commensurate structures and commensurate-incommensurate transitions. The parameters of the FK model can be estimated (or even evaluated from first principles) with good accuracy (see e.g. Braun and Medvedev, 1989). Also, these systems can often be studied experimentally by direct methods such as scanning tunnel microscopy.

In addition, the same model may be employed to describe a clean surface, if one treats the surface atoms as atoms of the effective FK chain, while the atoms of the underlying layer are assumed to produce an effective substrate potential. In particular, the FK model can be used to describe the *surface reconstruction* (Harten et al. 1985; Mansfield and Heeds, 1990) and the structure of *vicinal semiconductor surfaces* (Yang et al., 1991). Moreover, the FK-type models can also be used to describe processes of *crystal growth* (e.g. Franzosi et al., 1988).

In fact, one of the important restrictions of the standard FK model is the one-dimensional nature of the chain dynamics. In many physical situations the one-dimensional approximation is rather good. However, usually quasi-one-dimensional chains of atoms are not completely independent, forming a system of parallel chains. For example, the typical situation, when atoms are absorbed on stepped or furrowed surfaces of a crystal, can be described by a two-dimensional system of weakly coupled one-dimensional FK chains. Considering kinks of the primary FK chains as quasi-particles

subjected to a periodic PN potential, we may analyze collective excitations of the two-dimensional lattice model as "secondary" kinks which again can be described by a variant of the "secondary FK model". A system of interacting FK chains was analyzed, e.g. by Braun et al. (1988) and Braun and Kivshar (1990), and there are many papers devoted to the statistical mechanics of adsorbed layers.

One of the ways to make the FK model more realistic for describing a broader class of physical applications is to include an additional degree of freedom allowing the atoms to move in the direction *perpendicular to the chain*. The corresponding FK model with a transverse degree of freedom was proposed by Braun and Kivshar (1991). Interesting physical effects are possible in this type of models due to the existence of nontrivial zigzag-like excitations. More general models describing the dynamics in two-dimensional systems correspond to a *vector* generalization of the FK model, which is the most realistic model for two-dimensional arrays of atoms adsorbed on crystal surfaces; each atom has two degrees of freedom to move and is subjected to a two-dimensional external potential created by atoms of the surface. In fact, a variety of such models is generated by the symmetry properties of various substrate potentials (isotropic models with square, triangular, hexagonal lattices). We would like to mention those introduced by van der Merwe (1970), Snyman and van der Merwe (1974), Snyman and Snyman (1981), Abraham et al. (1984), Lomdahl and Srolovitz (1986).

3.6 Models of the DNA dynamics

Similar models can also play an important role in the interpretation of certain biological processes, such as *DNA dynamics and denaturation* (see Yomosa, 1983; Homma and Takeno, 1984; Yakushevich; 1989; Peyrard and Bishop, 1989; Dauxois et al., 1993; and also review papers by Zhou and Zhang, 1991; Gaeta et al., 1994). Definitely, the structure of a DNA-type double-helix chain is complex, but very general features can be modelled by the Hamiltonian (13.1), provided we assume that all bases of the DNA chain are identical. Then, the dynamics can again be described by the Hamiltonian (13.1), where m_a is the mass or the moment of inertia of a base, and $x_n^{(i)}$ is the generalized co-ordinate describing the base at site n on the i-th chain of the double helix. Usually, the DNA-type chain is described by two coupled equations for two coupled chains (double helix). There exist two types of models, planar and helicoidal ones. The planar model of torsional dynamics (Yakushevich, 1989) has originally been developed to describe travelling kinks, which are supposed to play a central role in the transcription of DNA. The two chains of the DNA helix structure are assumed to interact linearly, depending on the rotation of the end points of the bases which are coupled by H-bonds. Then, positions of the neighbouring bases

of different chains are described by the angles Θ_1, Θ_2 of rotation around the N-C bond, and the model is reduced to two coupled equations for the variables $\Psi_{1,2} = \frac{1}{2}(\Theta_1 \pm \Theta_2)$.

In the planar model of DNA vibrational dynamics (Peyrard and Bishop, 1989) the dynamics is associated with sugar puckering modes, and the coordinate $x_n^{(i)}$ describes the position of the base with respect to its distance from the axis of the double helix. For the variables $\Psi = \frac{1}{2}[x^{(1)} + x^{(2)}]$ and $\chi = [x^{(1)} - x^{(2)}]$, the model can be reduced to two independent equations,

$$m\ddot{\psi}_n = \beta(\psi_{n+1} + \psi_{n-1} - 2\psi_n) + V'(\psi_n) = 0,$$

$$m\ddot{\chi}_n = \beta(\chi_{n+1} + \chi_{n-1} - 2\chi_n),$$

where the potential is

$$V(\Psi_n) = \frac{1}{2}D\sum_n (e^{-\alpha\psi_n} - 1)^2.$$

The former equation is exactly the generalised FK model.

A modification of these models taking the helicoidal structure of DNA into account has also been suggested (Gaeta et al., 1994). In a double helix it happens that bases, which are one half-turn of the helix apart, end up spatially close to each other. Introducing "helicoidal" terms to the planar models to account for this effect will add a harmonic potential to the Hamiltonian (13.1) for the interaction between bases with numbers n and $n+h$. In the case of the model of vibrational dynamics, these higher-order inter-base interactions modify the FK model for the common displacement $\psi_n = \frac{1}{2}(x_1 + x_2)$, making the dispersion of linear waves more complicated. From the viewpoint of physics, the novel helicoidal terms produce extremal points in the linear dispersion allowing standard breather modes, i.e. localised oscillations responsible for local openings of the double helix which remain so.

4 Properties of kinks

4.1 On-site potential of general form

The standard model (13.1) assumes a sinusoidal potential $V_{\rm sub}(x_n)$. However, for some physical applications, the shape of the substrate potential deviates from the sinusoidal one. Indeed, the on-site substrate potential in the FK model is an *effective potential* resulting from couplings of the atoms in the chain with other degrees of freedom, e.g. with *substrate atoms*. Then, only in the lowest order of the Fourier expansion and in the case of a simple lattice with one atom per elementary cell, a sinusoidal potential can

be derived in a rigorous way. In all other physical situations, the periodic potential $V_{\rm sub}(x)$ deviates from the sinusoidal form.

In the general case, kinks can be easily described in the continuum approximation which is valid provided the discreteness effects are negligible. The equation of motion then becomes

$$u_{tt} - c^2 u_{xx} + V'_{\rm sub}(u) = 0. \tag{13.11}$$

Equation (13.11) is Lorentz invariant and, therefore, has always a stationary solution $u(x,t) = \phi(y)$, $y = \gamma[x - X(t)]$, $\gamma = (1 - v^2/c^2)^{-1/2}$, where the kink co-ordinate is defined as $X(t) = X_0 + vt$ and its velocity, $v = dX/dt$ varies in the interval $|v| < c$. The equation for the function $\phi(y)$, $d^2\phi/dy^2 = V'_{sub}(\phi)$, coincides with the equation of motion of an effective particle with co-ordinate ϕ in the potential $U[\phi(y)] = -V_{\rm sub}[\phi(y)]$ which can be easily integrated, and the shape of the kink can be expressed as

$$y = \mp \int_{x_m}^{\phi(y)} \frac{d\phi}{\sqrt{2V_{\rm sub}(\phi)}}. \tag{13.12}$$

Here, the upper sign corresponds to a kink (it describes a local contraction of the chain), the lower sign corresponds to an antikink, and the value x_m is the co-ordinate of the maximum of the substrate potential. Thus, the kink (anti-kink) solution connects two nearest neighbouring minima of of the substrate potential, say x_0 and $x_0 + 2\pi$. If the substrate potential has more than two minima per period, one may expect to find more than one type of kinks. One of the examples of the generalised potential is the so-called double SG potential (see e.g. Condat et al., 1983) $V_{\rm sub}(x) \propto -[\cos(x) + s\cos(2x)]$. We would like to mention also more general on-site potentials proposed by Peyrard and Remoissenet (1982).

Unlike the SG kink, a kink in the chain with a non-sinusoidal on-site potential may have an internal degree of freedom, the so-called kink shape mode (see e.g. Segur, 1983; Campbell et al., 1983; Braun et al., 1997). To explain the existence of such modes, one should linearise the equation of motion around the kink shape, $u(x) = u_k(x) + \Psi(x)e^{i\omega t}$. The function $\Psi(x)$ satisfies the linear eigenvalue equation

$$-c^2 \frac{d^2\Psi}{dx^2} + W(x)\Psi(x) = \omega^2 \Psi(x), \quad W(x) = \left.\frac{d^2 V_{\rm sub}(u)}{du^2}\right|_{u=u_k(x)}. \tag{13.13}$$

This equation has always a continuum spectrum of plane-wave solutions (i.e. *phonons*) with the frequencies $\omega > \omega_{\rm min}$, and also the so-called Goldstone mode $\Psi(x) = du_k/dx$ with $\omega = 0$ (in a discrete FK model the latter mode has a nonzero eigenvalue ω_{PN}). Besides these two modes, the eigenvalue spectrum can include one or more eigenfunctions with discrete eigenfrequencies within the gap $(0, \omega_{\rm min})$ or, depending on the shape of the substrate potential, with frequencies $\omega > \omega_{\rm max}$ (see details in Braun et

al., 1997a). Such modes are localized near the kink profile. They may be considered as excited states giving rise to internal oscillations of the kink shape. The shape modes can be excited during collisions between kinks, or due to the interaction of kinks with impurities, so that they play an important role in the kink dynamics.

4.2 Discreteness effects

The existence of kinks (and even their main properties) does not crucially depend on the discreteness of the primary model, so that the SG model considered above looks often as an acceptable approximation of the primary discrete chain, allowing to retain the basic properties of the system dynamics. However, a very specific property of the discrete lattice is the existence of the Peierls-Nabarro (PN) periodic potential $V_{PN}(X)$ for the kink motion, where X is the co-ordinate of the kink centre. The PN potential and its properties have originally been discussed in the context of dislocation theory (Peierls, 1940; Nabarro, 1947; Indenbom, 1958). To understand its origin we first note that in the continuum approximation the system is invariant with respect to any translation of the kink along the chain. In contrast to that, in a discrete model this invariance is absent, and only translations by the lattice spacing a_s and its integer multiples are allowed. The smallest energy barrier the kink has to overcome when moving through a lattice is known as the PN barrier, E_{PN}. The value of E_{PN} is equal to the difference between the two values of potential kink energy defined for two stationary configurations, a stable and an unstable (saddle) one.

The amplitude E_{PN} of the PN potential was calculated analytically for the quasi-continuum limit and for the weak-bond limit, as well as numerically (see references in Braun and Kivshar, 1998). One of the analytical approximations yields

$$E_{PN} = 32\pi^2 \frac{g}{\sinh(\pi^2\sqrt{g})}\left(1 + \frac{1}{2\pi^2 g}\right) \simeq 64\pi^2 g e^{-\pi^2\sqrt{g}}, \qquad (13.14)$$

i.e. E_{PN} vanishes when $g \gg 1$. A calculation based on the discrete SG model, such as the Taylor-Chirikov map, yields the following result which is asymptotically exact in the limit $g \gg 1$ (Lazutkin et al., 1989), $E_{PN} \simeq Age^{-\pi^2\sqrt{g}}$, where $A \simeq 712.26784...$, which differs from the result above only by a numerical factor of 1.13.

In a discrete lattice, the SG kink cannot move freely because of the PN potential, provided its initial kinetic energy is less than the PN barrier. In such a case, the kink gets trapped near a local minimum of the PN potential and oscillates with a frequency $\omega_{PN} \approx E_{PN}/2m$.

4.3 Kinks in external fields

In the presence of impurities or inhomogeneous stationary external fields the kink velocity varies, and the kink can be trapped by an impurity. Many features of soliton-impurity interactions have already been discussed by Kivshar and Malomed (1989) and Gredeskul and Kivshar (1992) in the framework of the SG model with local or extended inhomogeneities. For the discrete FK model, two new features of soliton-impurity interactions appear. First, in a discrete chain the kink moves in the presence of an effective PN potential. Thus, the kink parameters vary periodically, and this simple mechanism gives rise to the generation of phonons and a subsequent pinning of the kink by lattice discreteness. As a result, the discreteness effects may significantly modify the adiabatic scattering of kinks (see e.g. Braun and Kivshar, 1991). Second, an important feature of kink scattering by impurities in a discrete chain is the excitation of *impurity modes* during the scattering process. In fact, such an effect is also possible for continuous models provided one considers strong disorder (see below), but discreteness modifies the frequency of the impurity mode making is excitation easier (see Forinash et al., 1994).

To derive an effective equation of motion for a SG kink in a stationary external potential, we consider the simplest case of an inhomogeneous SG model,

$$u_{tt} - u_{xx} + \sin u = \epsilon f(x) \sin u. \tag{13.15}$$

Analyzing the kink dynamics in the framework of a collective co-ordinate approach, we obtain in a simple way an effective equation of motion for the kink co-ordinate (see e.g. Currie et al., 1977; McLaughlin and Scott, 1978). To derive such an equation, we note that the unperturbed SG system has an infinite number of conserved quantities (invariants) including the field momentum $P \equiv -\int_{-\infty}^{\infty} dx\ u_t u_x$. For the SG kink the momentum takes the form of the well-known relativistic expression $P = mv/\sqrt{1-v^2}$, where v is the kink velocity. In the presence of perturbations the momentum is no longer conserved; using Eq. (13.15) it is possible to show that it varies according to $dP/dt = \epsilon \int_{-\infty}^{\infty} dx f(x)(\cos u)_x$, provided the boundary conditions $u \to 0(2\pi)$ at $x \to \pm\infty$ hold. The adiabatic approach is now defined by the assumption that, for ϵ small enough, the kink shape is not affected, and only the kink co-ordinate X becomes a slowly varying function of time. Within this hypothesis it can be shown that in the non-relativistic limit the kink centre obeys Newton's equation of motion, $m\, d^2X/dt^2 = U'(X)$, where

$$U(X) \equiv -2\epsilon \int_{-\infty}^{\infty} dx \frac{f(x)}{\cosh^2(x-X)}. \tag{13.16}$$

The following two cases arise naturally from Eq. (13.16). If $f(x)$ varies rapidly over distances of the order of the kink length, then ϵ has to be small for our approximation to hold. For example, in the case $f(x) = \delta(x)$,

we have $U(X) = -2\epsilon \mathrm{sech}^2 X$ (McLaughlin and Scott, 1978). On the other hand, if $f(x)$ is a slowly varying function, i.e. its characteristic length (say L) is much larger than the kink width, it is not necessary for ϵ to be small, because all the parameters of the perturbation theory are of the order of L^{-1}, and we are left with $U(X) \approx 4\epsilon f(X/L)$.

The adiabatic theory is not valid when a localised impurity supports an impurity mode, i.e. an oscillating linear mode at the impurity site. In this case the kink position and the impurity mode amplitude are two effective collective co-ordinates. Different kinds of resonant interactions of solitons or kinks with impurities have recently been reviewed by Belova and Kudryavtsev (1997).

In a discrete lattice, the kink motion is affected by the PN relief. One of the ways to derive the effective equation of motion for the kink is based on the projection-technique approach developed by the group of C. Willis. Another approach is based on the Lagrangian formalism (see e.g. Braun and Kivshar, 1991).

4.4 Compacton kinks

When all linear interactions between particles in the chain vanish, so that their coupling becomes purely *nonlinear*, the kinks are localised on a finite interval and have no exponentially decaying tails. Such a kind of solitary wave is known as *compacton* and has first been introduced for the generalised Korteweg-de Vries equation with purely nonlinear dispersion (Rosenau and Hyman, 1993). Compacton kinks have been discussed by Dusuel et al. (1998) for the Klein-Gordon model with purely nonlinear dispersion. They found two types of the compaction kinks, a moving (unstable) and a static (stable) one, described by the analytical solutions $u_K(\xi) = \pm \sin\{(2V_0/3c_{\mathrm{nl}}^{1/4}(\xi - \xi_0)\}$, for $|\xi - \xi_0| < 1$, and $u_K(\xi) = \pm 1$ otherwise. Here $\xi = x - \sqrt{c_{\mathrm{l}}}t$, V_0 is the strength of the potential, c_{l} and c_{nl} are the coefficients of linear and nonlinear dispersion, respectively. When the linear dispersion vanishes, $c_{\mathrm{l}} = 0$, the compacton kink is stable. The mechanical analog of such a model was observed in a chain of pendulums with two degenerate ground states; see Dusuel et al. (1998).

5 Experimental verifications

5.1 Josephson junctions

A fluxon moving in a JJ of length L with velocity $v = u\bar{c}$ generates the dc voltage $V = \Phi_0 u\bar{c}/L$. In experiments with long JJs, the signatures of fluxon motion are the so-called zero-filed steps (ZFSs) at voltages $V_n = n\Phi_0\bar{c}/L$ which appear in the current-voltage characteristics (I-V curve) of the junction. These steps were first observed by Fulton and Dynes in 1973

(Fulton and Dynes, 1973) who suggested that the step index n is equal to the number of fluxons oscillating in the junction. The ZFS is given by the dependence of the average fluxon velocity ($u \propto V$) on the driving force ($\gamma \propto I$). The fluxon arriving at the junction boundary is reflected and, as an anti-fluxon, then driven back into the junction by the bias current. A discussion of this interesting shuttle-like process for different junction geometries can be found in several review papers (e.g. Pedersen, 1986). This is an indirect manifestation of the fluxon dynamics in experiments.

The influence of boundaries can be avoided in the special case of a ring-shaped (annular) junction geometry. An annular junction serves as an excellent model for studying soliton dynamics. Due to the magnetic flux quantization in a superconducting ring, the number of fluxons initially trapped in the annular junction is conserved. A circular motion of the fluxons under the influence of a current passing through the junction induces a dc voltage proportional to their average velocity. The soliton dynamics can be studied here under periodic boundary conditions. While the fabrication of annular junctions is rather easy, the trapping of fluxons inside them remains a difficult problem. Using various trapping techniques, both single-soliton (Davidson et al., 1985) and multi-soliton dynamics (Ustinov et al., 1992) have been investigated in annular junctions.

A interesting theoretical idea about the kink as a source of radiation, the so-called Cherenkov radiation, has been suggested by Kivshar and Malomed (1988). Recently, first experimental evidences for Cherenkov radiation have been reported. The experiments have been performed on two very different systems, but the mechanism is rather general: Cherenkov radiation can be generated if the kink velocity $v = u\bar{c}$ is equal to the phase velocity ω/k of linear electromagnetic waves. This condition can be satisfied if the kink velocity exceeds the lowest phase velocity of linear waves in the junction. Very clear evidence for Cherenkov radiation emitted by a moving kink has been found by Goldobin et al. (1998) in a system of two stacked annular JJs. As soon as the kink velocity exceeds the phase velocity, an oscillating wave corresponding to Cherenkov radiation arises behind the moving kink, and its image appears in the coupled junction and can be measured. When an integer number of Cherenkov radiation wavelengths fits to the corresponding eigenmodes of the junction, resonance occurs. The resonances result from the interaction of the Cherenkov wave with the fluxon.

Hechtfischer et al. (1997) observed very unusual broadband non-Josephson radiation which they attributed, due to the magnetic field dependence of radiation power, to the emission of Cherenkov radiation by Josephson vortices moving in a multi-layer JJ structure. For a stack with N junctions, there are N different linear mode velocities. For strong coupling, the lowest mode velocity is about $\bar{c}/\sqrt{2}$. As soon as the fluxon velocity v rises above $\bar{c}/\sqrt{2}$, Cherenkov emission should appear. This idea has recently been confirmed by numerical simulations.

In contrast to the continuum model, for the discrete SG model the phase

velocity $v_{ph} = \omega/k$ of the linear waves may coincide with the kink velocity v already in the case of one junction. Considering the kink as a quasi-particle, this radiation emission mechanism is equivalent to that of Cherenkov radiation. In a discrete JJ resonance between the emitted Cherenkov waves and the periodic chain of moving kinks (fluxons) can occur, if the wavelength and the fluxon spacing are commensurate (Ustinov et al., 1993). A clear example of this effect can be found for a fluxon rotating in the annular discrete JJ ring. The corresponding resonance values of the voltage V_m are given by the formula

$$V_m = \frac{V_0}{m}\sqrt{\Lambda_J^{-2} + 4\sin^2\left(\frac{\pi m}{N}\right)}, \tag{13.17}$$

where $\Lambda_J^2 = L_J/L_S$, L_J and L_S being the Josephson inductance and the self-inductance, respectively, and $V_0 = \Phi_0/2\pi\sqrt{L_S C}$, where C is the junction capacitance. Experimentally, the discreteness-induced single-fluxon resonances for $m_1 \neq 1$ and $m_2 = 1$ were observed by van der Zant et al. (1995) using an 8–junction ring. In these experiments they observed that the *I-V* curves are smooth only for high Λ_J^2 values; when Λ_J^2 is decreased by lowering the temperature, a fine structure became visible. Experiments show the fine structure in the current-voltage characteristic of a single vortex trapped in a ring for Λ_J^2=2.2. In total, six resonant steps are present, corresponding to local minima of the differential resistance dV/dI.

More recently, similar resonances were reported by Duwel et al. (1997) for a spatially-periodic fluxon chain moving in a long JJ. Another type of resonances with $m_1 = 1$ and $m_2 \neq 1$ correspond to very high fluxon oscillation frequencies. Such states were observed in experiments by Caputo et al. (1997).

5.2 Magnetic systems

As has been mentioned above, a kink soliton in a magnetic system represents a microscopic structure which cannot be directly observed like a solitary water wave but can only be probed by its influence on macroscopic properties. Magnetic solitons have a well-defined excitation energy, so one might speculate about the possibility to excite them from outside, like magnons, by resonant microwave absorption or inelastic neutron scattering. Such attempts, however, fail for the lack of an appropriate local excitation mechanism. On the other hand, to excite them externally by microwave pulses or field gradients would require to apply magnetic fields of some Teslas strength on a length scale of a few nanometers within a time scale of nanoseconds, which is still far beyond technical feasibility.

Hence, there remains only the chance to probe solitons which are thermally excited. To separate their contribution from coexisting thermal excitations, an experimental proof first requires that soliton excitations are

dominating within a certain range of field and temperature, and second that they differ from others by characteristic properties. In this sense, dynamic methods are generally more instructive than static ones.

Different standard techniques can be applied to study magnetic solitons. The first and most common technique was *inelastic neutron scattering* (Kjems and Steiner, 1978; Regnault et al., 1982; Steiner et al., 1983), which allows one to measure the soliton contribution to the *dynamic structure factor*, which is the spatial and temporal Fourier transform of the spin correlation function (see e.g. Mikeska, 1978). Complementary information can be obtained by nuclear magnetic resonance (NMR) (Boucher and Renard, 1980; Boucher et al., 1982; Benner et al., 1984; Seitz and Benner, 1987), which probes the dynamic structure factor integrated over all wavenumbers. Two other methods, Raman scattering (Cibert et al., 1981) and electron spin resonance (ESR) (Benner et al., 1984; 1987) were applied to probe the influence of solitons indirectly through the soliton-induced broadening of the magnon linewidth.

As for the physical systems, the prototype of a FM chain described by the SG equation is the well-known $CsNiF_3$ (e.g. Kjems and Steiner, 1978), whereas the prototype of an AFM chain is the compound $(CH_3)_4NMnCl_3$ (TMMC) (Regnault et al, 1982). To give an example for the experimental results on soliton properties in TMMC, we focus on NMR data on soliton pairing .

As was theoretically shown (Hołyst and Sukiennicki, 1985), the phase transition in weakly coupled AFM chains can be understood as being induced by the pairing of kink solitons. This means that π-solitons on the same chain, moving independently above the ordering temperature T_N, are bound into pairs below T_N. The pairing is effected through weak inter-chain interactions lifting the degeneracy of the ground-state and giving rise to AFM long-range order. The weak dipolar couplings between neighbouring chains tend to align facing spins parallel to each other. The occurrence of two separate π-solitons (or antisolitons) on one of the chains would reverse the AFM sublattices on a certain part of the chain and align the facing spins of neighbouring chains in an unfavourable anti-parallel orientation. The corresponding increase of energy can be minimized if both solitons move towards each other up to a certain minimum distance, thus forming a $+2\pi$ (-2π) twist of the chain, since they cannot annihilate for topological reasons. Describing the inter-chain interactions by a staggered mean field as the simplest approach, Hołyst (1985; 1989) was able to map the corresponding equation of motion to a *double*-SG equation,

$$\Phi_{zz} - \frac{1}{c^2}\Phi_{tt} = \frac{m^2}{2}\sin 2\Phi + \eta \sin \Phi, \tag{13.18}$$

where c and m are defined by $c = 4|J|S\sqrt{1 - A/4|J|}$ and $m = g\mu_B H/4|J|S$, while η describes the mean inter-chain coupling $\eta = 6J'/|J|$. Analytical solutions of this equation which, in fact, resemble two coupled π-kinks and the

corresponding thermodynamics have been treated in the literature (Leung, 1982; Condat et al, 1983).

Two important effects are related with π-soliton pairing: First, a change of the kink topology occurs, which drastically affects the dominating spin fluctuation rate. This rate is inversely proportional to the soliton density above T_N, while it is directly proportional to it below T_N. Second, there occurs an increase of the activation energy by roughly a factor of 2,

$$E_{2\pi} = 2E_\pi \left(\sqrt{1+\xi} + \xi \mathrm{sinh}^{-1} \sqrt{1/\xi} \right), \quad \xi = \frac{96 J' |J| S^2}{(g\mu_B H)^2} \quad , \tag{13.19}$$

approaching $2E_\pi = 2\alpha_\pi H k_B$ for $\xi \to 0$, i.e. in the limit of very weak interchain coupling. The corresponding NMR spin-lattice relaxation times read (Hołyst and Benner, 1989)

$$T_1 \sim H e^{-E_\pi / k_B T}, \quad T > T_N \quad (1D) \tag{13.20}$$

$$T_1 \sim \frac{T}{H^2} e^{+E_{2\pi} / k_B T}, \quad T < T_N \quad (3D) \tag{13.21}$$

for the case of unpaired and paired solitons, respectively.

Experimental evidence for soliton pairing in TMMC was obtained by extending field and temperature dependent T_1 measurements to the regime of 3D order, i.e. to $H > 40$kOe and $T < 2$K. This was, in fact, proved by the experimental data taken above T_N, which yield a negative slope with $\alpha_{\pi,\mathrm{exp}} \simeq 0.28$K/kOe. The 3D data taken below T_N show the opposite slope which reflects the expected change of topology. Since these data were measured at four different fields, they are no longer universal as expected from the different pre-factors in Eq. (13.21) and from the additional H dependence of $E_{2\pi}$ in Eq. (13.19). However, the corresponding theoretical curves are in very good agreement with these data, proving, in fact, the expected doubling of activation energy ($2\alpha_{\pi,\mathrm{exp}} \simeq 0.57$K/kOe, see Benner et al., 1991).

6 Concluding remarks

We have presented, from a rather general point of view, the basic concepts and properties of low-dimensional nonlinear models for describing the dynamics of solids where topological excitations play an important role in the system dynamics, modelling complex solid-state defects of different nature, such as domain walls, magnetic vortices, dislocations, charged defects, etc. We did not restrict ourselves to specific applications, but instead tried to summarize the properties that are common for many models. Special attention has been paid to magnetic systems and long Josephson junctions, where experimental results on the kink dynamics are available.

To conclude our brief presentation of the nonlinear dynamics of the FK model and its physically important generalizations, it is appropriate to mention some other examples of physical systems where the fundamental concepts and results based on the FK model and the analysis of its nonlinear excitations such as kinks and breathers are effectively applied.

In particular, a rather wide class of problems involves the continuum limit of the FK model, i.e. the perturbed SG equation. The number of novel physical phenomena, which appear in this model when external *ac* or *dc* forces are applied, is indeed too large to be reviewed in such a short paper. In particular, we mention the stabilization of breathers by direct or parametric forces and more complex nonlinear dynamics including period-doubling scenarios, spatio-temporal pattern formation, and chaos. Similar effects can also be found for discrete chains, but they have not been investigated in detail yet.

The over-damped discrete FK chain driven by an *ac* force has recently been investigated analytically and numerically (see e.g. a review paper by Floria and Mazo, 1996, and references therein). Besides, in view of possible applications of the FK model to the problems of tribology, a number of recent papers has been devoted to the FK model with dissipation (as well as its generalisation known as the FK-Tomlinson model) driven by a dc force (e.g. Weiss and Elmer, 1996, 1997; Braun et al., 1997).

Last but not least, during recent years an exponentially growing activity was directed towards understanding the properties of localised modes in discrete lattices with on-site and inter-site potential. Numerous results, including a rigorous proof of their existence and stability, have appeared in the literature which, however, are beyond the scope of this Chapter. We just mention the most recent review paper by Flach and Willis (1998) where a summary of recent progress in the theory of nonlinear localised modes (also called *discrete breathers*) is presented.

In conclusion, the understanding of the specific and unusual properties of solids on the basis of simplified nonlinear models is an active and attractive topic of research, which still offers many open problems. Since realistic physical models of solids are rather complicated, it is extremely important to develop the basic concepts by means of simple physical models.

7 References

[1] Ablowitz, M.J., D. Kaup, A.C. Newell, and H. Segur, 1973, Phys. Rev. Lett. **30**, 1262.

[2] Abraham, F.F., W.E. Rudge, D.J. Auerbach, and S.W. Koch, 1984, Phys. Rev. Lett. **52**, 445.

[3] Antonchenko, V.Ya., A.S. Davydov, and A.V. Zolotaryuk, 1983, phys. stat. solidi (b) **115**, 631.

[4] Atkinson, W., and N. Cabrera, 1965, Phys. Rev. **138**, A763.

[5] Bäcklund, A.V., 1882, *"Om ytor med konstant negative krökning"*, In: Lunds Universitets Ars-skrift XIX, IV, p. 1.

[6] Belova, T.I., and A.E. Kudryavtsev, 1997, Physics-Usphekhi **40**, 359.

[7] Benner, H., H. Seitz, J. Weise, J.P. Boucher, 1984, J. Magn. Magn. Mat., **45**, 354.

[8] Benner, H., J. Weise, R. Geick, and H. Sauer, 1987, Europhys. Lett. **3**, 1135.

[9] Benner, H., J.A. Holyst and J. Löw, 1991, Europhys. Lett., 14, 383.

[10] Braun, O.M., Yu.S. Kivshar, and A.M. Kosevich, 1988, J. Phys. C **21**, 3881.

[11] Braun, O.M. and Yu.S. Kivshar, 1990, J. Phys.: Cond. Matter **2**, 5961.

[12] Braun O.M. and Yu.S. Kivshar, 1991a, Phys. Rev. B **44**, 7694.

[13] Braun, O.M. and Yu.S. Kivshar, 1991b, Phys.Rev. B **43**, 1060.

[14] Braun, O.M. and V.K. Medvedev, 1989, Usp. Fiz. Nauk **157**, 631 [Sov. Phys.-Usp. **32**, 328 (1989)].

[15] Braun, O.M. and Yu.S. Kivshar, 1998, Phys. Rep., **306**, 1.

[16] Braun, O.M., Yu.S. Kivshar, and M. Peyrard, 1997a, Phys. Rev. E **56**, 6050.

[17] Braun, O.M., T. Dauxois, M.V. Paliy, and M. Peyrard, 1997b, Phys. Rev. Lett. **78**, 1295.

[18] Boucher,J.P., and J.P. Renard, 1980, Phys. Rev. Lett. **45**, 486.

[19] Boucher, J.P., H. Benner, F. Devreux, L.P. Regnault, J. Rossat-Mignod, C. Dupas, J.P. Renard, J. Bouillot, and W.G. Stirling, 1982, Phys. Rev. Lett., **48**, 431.

[20] Campbell, D.K., J.F. Schonfeld, and C.A. Wingate, 1983, Physica D **9**, 1.

[21] Caputo, P., A.V. Ustinov, N. Iosad, and H. Kohlstedt, 1997, J. Low Temp. Phys. bf 106, 353.

[22] Christiansen, P.L., A.V. Savin, and A.V. Zolotaryuk, 1997, J. Comput. Phys. **134**, 108.

[23] Christophorov, L.N., and Yu.B. Gaididei, 1992, Phys. Lett. A **167**, 367.

[24] J. Cibert, Y. Merle d'Aubigné, 1981, Phys. Rev. Lett. **46**, 1428.

[25] Condat, C.A., R.A. Guyer, and M.D. Miller, 1983, Phys. Rev. B **27**, 474.

[26] Currie, J.F., S.E. Trullinger, A.R. Bishop, and J.A. Krumhansl, 1977, Phys. Rev. B **15**, 5567.

[27] Dauxois, T., M. Peyrard, and A.R. Bishop, 1993a, Phys. Rev. E **47**, 684.

[28] Davidson, A,, B. Dueholm, B. Kryger, and N.F. Pedersen, 1985, Phys. Rev. Lett., **55**, 2059.

[29] Dehlinger, U., 1929, Ann. Phys. Lpz. **2**, 749.

[30] Döttling, R., J. Esslinger, W. Lay, and A. Seeger, 1990, in: *Nonlinear Coherent Structures*, Lecture Notes in Physics, Vol. 353, Eds. M. Barthes and J. Léon (Springer-Verlag, Berlin) p. 193.

[31] Dusuel, S., P. Michaux, and M. Remoissenet,1998, Phys. Rev. E, **57**, 2320.

[32] Duwel, A.E., S. Watanabe, E. Trias, T.P. Orlando, H.S.J. van der Zant, and S.H. Strogatz, 1997, J. Appl. Phys., **82**, 4661.

[33] Enneper, A., 1870, *Uber asymptotische Linien*, In: Nachr. Königl. Gesellsch. d. Wiss. Göttingen, p. 493.

[34] Flach, S., and C.R. Willis, 1998, Phys. Rep. **295**, 181

[35] Floria, L.M. and Mazo, J.J., 1996, Advances in Physics, **45**(6), 505.

[36] Forinash, K., M. Peyrard, and B. Malomed, 1994, Phys. Rev. E **49**, 3400.

[37] Frank, F.C. and J.H. van der Merwe, 1949a, Proc. Roy. Soc. (London) A **198**, 205; 216.

[38] Frank, F.C. and J.H. van der Merwe, 1949b, Proc. Roy. Soc.(London) A **200**, 125.

[39] Franzosi, P., G. Salviati, M. Seaffardi, F. Genova, S. Pellegrino, and A. Stano, 1988, J. Cryst. Growth, **88**, 135.

[40] Frenkel, Ya., and T. Kontorova, 1938, Phys. Z. Soviet Union **13**, 1.

[41] Frenkel, Ya., and T. Kontorova, 1939, Fiz. Zh. (Moscow) **1**, 137 (in Russian).

[42] Frenkel, Ya.I., 1972, *Introduction into the Theory of Metals* (Leningrad, Nauka) (in Russian).

[43] Fulton, T.A. and R.C. Dynes, Sol. St. Commun.,1973, **12**, 57.

[44] Gaeta, G., C. Reiss, M. Peyrard, and T. Dauxois, 1994, Rivista Nuovo Cim. **17**, 1.

[45] Goldobin, E., A.Wallraff, N. Thyssen, and A.V. Ustinov.,1998, Phys. Rev. B., **57**, 130.

[46] Gredeskul, S.A. and Yu. S. Kivshar, 1992, Phys. Rep. **216**, 1.

[47] Grønbech-Jensen, N., M.R. Samuelsen, P.S. Lomdahl, and J.A. Blacburn, 1990, Phys. Rev. B., **42**, 3976.

[48] Grønbech-Jensen, N., O.H. Olsen, and M.R. Samuelsen, 1993, Phys. Lett. A., **179A**, 27.

[49] Harten U., A.M. Lahee, J.P. Toennies, and Ch. Wöll, 1985, Phys. Rev. Lett. **54**, 2619.

[50] G. Hechtfischer, R. Kleiner, K. Schlenga, W. Walkenhorst, P. Müller, and H.L. Johnson, 1997a, Phys. Rev. B., bf 55, 14638.

[51] Hechtfischer, G., R. Kleiner, A.V. Ustinov, and P. Mueller, 1997b, Phys. Rev. Lett. **79**, 1365.

[52] Hochstrasser, D., H. Bütner, H. Desfontaines, and M. Peyrard, 1988, Phys. Rev. A **36**, 5332.

[53] Holyst, J.A. and A. Sukiennicki,1985, J. Phys. C **18**, 2411.

[54] Holyst, J.A., 1989, Z. Phys. B **74**, 341.

[55] Holyst, J.A. and H. Benner, 1989, Solid State Commun. **72**, 385.

[56] Homma, S., and S. Takeno, 1984, Progr. Theor. Phys. **72**, 679.

[57] Indenbom, V.L., 1958, Sov. Phys.- Cryst. **3**, 197.

[58] Josephson, B.D., 1965, Adv. Phys. **14**, 419.

[59] Kalosakas, G., A.V. Zolotaryuk, G.P. Tsironis, and E.N. Economou, 1997, Phys. Rev. E **56**, 1088.

[60] Kivshar, Yu.S. and B.A. Malomed, 1989, Rev. Mod. Phys. **61**, 763.

[61] Kivshar, Yu.S., 1991, Phys. Rev. A **43**, 3117.

[62] Kivshar, Yu.S. and B.A. Malomed, 1988, Phys. Rev. B., **37**, 9325.

[63] Kjems, J.K. and M. Steiner, 1978, Phys. Rev. Lett., **41**, 1137.

[64] Kleiner, R. F., Steinmeyer, G. Kunkel and P. Müller, 1992, Phys. Rev. Lett., **68**, 2394.

[65] Kleiner, R. and P.Müller, 1994, Phys. Rev. B., **49**, 1327.

[66] Kochendörfer, A. and A. Seeger, 1950, Z. Physik **127**, 533 (in German).

[67] Kosevich, A.M., B.A. Ivanov, and A.S. Kovalev, 1990, Phys. Rep. **194**, 117.

[68] Kovalev, A.S., A.D. Kondratyuk, A.M. Kosevich, and A.I. Landau, 1993, phys. stat. sol. (b) **177**, 117.

[69] Landau, A.I., A.S. Kovalev, and A.D. Kondratyuk, 1993, phys. stat. sol. (b) **179**, 373.

[70] Lazutkin, V.F., I.G. Schachmannski, and M.B. Tabanov, 1989, Physica D **40**, 235.

[71] Lee, J.U., J. E. Nordman, and G. Hohenwarter, Appl. Phys. Lett., 1995, **67**, 1471.

[72] Leung, K.M., 1982, Phys. Rev. B **26**, 226.

[73] Leung, K.M., 1983, Phys. Rev. **B** 27, 2877.

[74] Lomdahl, P.S. and D.J. Srolovitz, 1986, Phys. Rev. Lett. **57**, 2702.

[75] Mansfield, M. and R.J. Needs, 1990, J. Phys.: Cond. Matter **2**, 2361.

[76] Maugin, G.A. and A. Miled, 1986, Phys. Rev. B **33**, 4830.

[77] McLaughlin, D.W. and A.C. Scott, 1978, Phys. Rev. A **18**, 1652.

[78] Mikeska, H., 1978, J. Phys. C **11**, L 29.

[79] Mikeska, H.J. and M. Steiner, 1991, Adv. Phys. **40**, 191.

[80] Mineev, M.B., G.S. Mkrtchjan, and V.V. Schmidt, 1981, J. Low. Temp. Phys., **45**, 497.

[81] Nabarro, F.R.N., 1947, Proc. Roy. Soc. (London), **59**, 256.

[82] Paneth, H.R., 1950, Phys. Rev. **80**, 708.

[83] R.D. Parmentier, in: The New Superconducting Electronics, H. Weinstock, R.W. Ralston, eds. (Kluwer, Dordrecht, 1993), p.221.

[84] Pedersen, N.F., 1986, In: *Josephson Effect-Achievements and Trends*, Ed. A. Barone (World Scientific, Singapore).

[85] Pedersen, N.F.and A.V. Ustinov, 1995, Supercond. Sci. Technol., **8**, 389.

[86] Peierls, R., 1940, Proc. Roy. Soc. (London) **52**, 34.

[87] Perring, J.K., and T.H.R. Skyrme, 1962, Nucl. Phys. **31**, 550.

[88] Peyrard, M. and A.R. Bishop, 1989, Phys. Rev. Lett. **62**, 2755.

[89] Peyrard, M., St. Pnevmatikos, and N. Flytzanis, 1987, Phys. Rev. A, **36**, 903.

[90] Peyrard, M. and M. Remoissenet, 1982, Phys. Rev. B **26**, 2886.

[91] Pnevmatikos, St., 1988, Phys. Rev. Lett. **60**, 1534.

[92] Pnevmatikos, St., Yu.S. Kivshar, A.V. Savin, A.V. Zolotaryuk, and M.J. Velgakis, 1991, Phys. Rev. A **43**, 5518.

[93] Pouget, J. and G.A. Maugin, 1984, Phys. Rev. B **30**, 5306.

[94] Pouget, J. and G.A. Maugin, 1985, Phys. Lett. A **109A**, 389.

[95] Regnault, L.P., J.P. Boucher, J. Rossat-Mignod, J.P. Renard, J. Bouillot, and W.G. Stirling, 1982, J. Phys. C **15**, 1261.

[96] Rosenau, P. and J.M. Hyman, 1993, Phys. Rev. Lett., **70**, 564.

[97] Sakai, S., P. Bodin, and N.F. Pedersen, 1993, J. Appl. Phys. **73**, 2411.

[98] Savin, A.V. and A.V. Zolotaryuk, 1991, Phys. Rev. A **44**, 8167.

[99] Scott, A.C., 1969, Amer. J. Phys. **37**, 52.

[100] Seeger, A., 1980, In: *Continuum Models of Discrete Systems*, Eds. E. Kröner and K.H. Anthony, University of Waterloo, Waterloo (Ontario), p. 253.

[101] Seeger, A. and A. Kochendörfer, 1951, Z. Physik **130**, 321.

[102] Seeger, A., H. Donth, and A. Kochendörfer, 1953, Z. Physik **134**, 173.

[103] Segur, H., 1983, J. Math. Phys., **24**, 1439.

[104] Seitz, H. and H. Benner, 1987, Z. Phys. B **66**, 485.

[105] Snyman, J.A., and J.H. van der Merwe, 1974, Surface Sci. **45**, 619.

[106] Snyman, J.A., and H.C. Snyman, 1981, Surface Sci. **105**, 357.

[107] Steiner, M., K. Kakurai, and J.K. Kjems, 1983, Z. Phys. B, **53**, 117.

[108] Ustinov, A.V., M. Cirillo, and B.A. Malomed, Phys. Rev. B., 1993, **47**, 8357.

[109] Ustinov, A.V., T. Doderer, R.P. Huebener, N.F. Pedersen, B. Mayer, and V. A. Oboznov, 1992, Phys. Rev. Lett., **69**, 1815.

[110] Ustinov, A.V. and R.D. Parmentier, in: Nonlinear Physics: Theory and Experiment, E. Alfinito, M. Boiti, L. Martina, and F. Pempinelli, eds. (World Scientific, Singapore, 1996), p.582.

[111] van der Merwe, J.H., 1970, J. Appl. Phys. **41**, 4725.

[112] van der Zant, H.S.J., T.P. Orlando, S. Watanabe, and S.H. Strogatz, 1995, Phys. Rev. Lett., **74**, 174.

[113] Weiss, M., and F.-J. Elmer, 1996, Phys. Rev. B **53**, 7539.

[114] Weiss, M., and F.-J. Elmer, 1997, Z. Phys. B **104**, 55.

[115] Yakushevich, L.V., 1989, Phys. Lett. A **136**, 413.

[116] Yang, Y.N., B.M. Trafas, R.L. Siefert, and J.H. Weaver, 1991, Phys. Rev. B **44**, 3218.

[117] Yomosa, S., 1983, Phys. Rev. A **27**, 2120.

[118] Zabusky, N.J. and M.D. Kruskal, 1965, Phys. Rev. Lett. **15**, 240.

[119] Zakharov, V.E., S.V. Manakov, S.P. Novikov, and L.P. Pitaevsky, 1980, *Theory of Solitons* (Moscow, Nauka) [English Translation by Consultants Bureau, New York, 1984].

[120] Zhou, G.Z. and C.T. Zhang, 1991, Physica Scripta **43**, 347.

[121] Zolotaryuk, A.V., K.H. Spatschek and E.W. Ladke, 1984, Phys. Lett. A **101**, 517.

[122] Zolotaryuk, A.V., 1986, Teor. Mat. Fiz., **68**, 415 [Theor. Math. Phys. **68**, 916 (1986)].

[123] Zolotaryuk, A.V., and St. Pnevmatikos, 1990, Phys. Lett. A **143**, 233.

14
2-D Breathers and Applications

J. L. Marín
J. C. Eilbeck
F. M. Russell

ABSTRACT
In this chapter we show how a new type of nonlinear lattice excitation is helping to understand the long-standing puzzle of unexplained dark lines in crystals of muscovite mica. In fact, it was the conjecture that some kind of quasi-one-dimensional lattice soliton was responsible for those lines which led to the discovery of this new family of lattice excitations: mobile localized breathers of longitudinal type.
We explore several properties of these moving breathers, both by numerical methods and by experimenting with analogue models. The results suggest a much broader application than just the mica problem.

1 Introduction

There has been a long standing problem of understanding track formation in some mica minerals. Works by Russell and co-workers [25] had identified mysterious dark lines in natural crystals of muscovite mica, such as shown in Fig. 1. These are formed as a result of a local phase transition involving the precipitation of meta-stable dilutions of Fe and other impurity atoms. Track formation or recording processes of this type are normally thought to be triggered by energetic, charged particles (cosmic rays, particles from radioactive decay, etc.). However, most dark lines in mica cannot be explained in terms of charged particle tracks, although they closely resemble them. This prompted the conjecture of lattice solitons ("quodons") as the mechanisms forming the lines [20, 21].

Recent studies give strong support to this conjecture. In [14] we showed for the first time how localized excitations of longitudinal type can move in a 2-D lattice. These lattice modes appear to be the mobile version of intrinsic localized modes [24], more commonly known as *discrete breathers*. They are highly anharmonic and localized on just a few sites, and they propagate along lattice directions at sub-sonic speeds with a very small energy degradation. They are robust with respect to changes in the model, i.e., they appear in a wide range of lattices with different potentials. On the other hand they are not exact solitons, since they seem to interact non-elastically with other breathers. However, the loss of energy on collision

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 293−305, 2000.

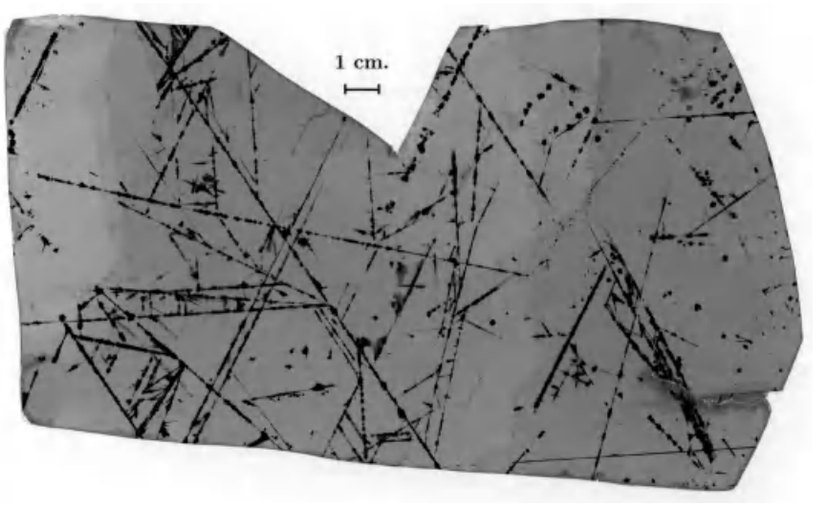

FIGURE 1. Photograph of a sheet of muscovite showing a wealth of tracks, most of them aligned along crystal directions.

frequently is small, so these pulses are reasonably robust.

We present here further studies of these moving breathers in systems with realistic potentials, obtaining a more accurate model for the problem of tracks in mica. It is now recognized that the phenomenon is highly generic, and we point to two possible applications in the areas of condensed matter.

2 Deciphering the lines in mica

Natural crystals of muscovite mica often are doped with iron and other trace elements such as Ca, Mn, Mg, and Ti. In a similar way to carbon in metallic iron, the iron (Fe) in mica is held in solid solution at the high temperature (about 800K) needed for crystal formation. Such crystals are formed in solidifying magmas deep underground. As the crystals slowly cool, the solubility of Fe in mica decreases and, depending on the initial concentration, it may reach saturation. If this occurs then further slight cooling results in a super-saturated state in which the crystal structure is metastable. Given the opportunity via some type of localized lattice perturbation, the structure will revert to the lower-energy stable saturated state, precipitating the excess Fe as an oxide. With further slow cooling, this solid-state phase transition continues until much of the Fe is eliminated from the mica lattice. During this later stage the precipitating Fe oxide accretes at the initial precipitation sites leading to massive growth or

"decoration" of those sites. The rate of this decoration process is limited by the slow rate of release of Fe as the crystals cool slowly within the surrounding rock. The Fe oxide precipitates in the form of magnetite, Fe_3O_4, a moderate electrical conductor that absorbs light. This is why it appears as black spots, plates or ribbons in the otherwise clear mica crystal. Although the initial precipitation site is on the atomic scale, the decoration process can magnify them sufficiently as to make such ribbons visible to the unaided eye.

Although muscovite mica is a common mineral, it can form only under conditions of high temperature and pressure such as exist in molten pegmatite 5 km or more underground. With mountain building it erodes slowly to the surface and can be mined. Large nearly perfect crystals grow very slowly and it is not practical nor feasible to grow them in the laboratory for study of the initial triggering of the precipitation process. Hence, studies can be made only on samples of natural crystals in which the metastable conditions cannot be controlled. So the study of decorated defects in doped mica is a deductive science, like astronomy or cosmology. Nevertheless, progress in understanding the defects has been made by studying a wide variety of mica samples.

In the first studies of these tracks in mica, one of the authors (FMR, [17]) identified long decorated ribbons in the (001) planes, but in random directions. Many measurements of these tracks led to the conclusion that they had been created by energetic charged particles, such as muons created by cosmic rays, electron-positron showers created from these muons, and from positron emission coming from the decay of isotope ^{40}K nuclei [19, 18]. It was shown that the tracks from the ^{40}K source were consistent with paths expected of charged particles following Rutherford scattering. The alternative explanation in terms of some type of crystal defect was highly improbable, given that the geometry of such defects would have to mimic the behavior of the charged particles paths.

But these were not the most interesting type of tracks observed in mica. In fact, most of the decorated lines could not be attributed either to the tracks of charged particles or to any other known type of crystal defect. These unexplained lines lie along crystal directions of low Miller indexes in the (001)-plane, although they show several features that resemble the tracks of charged particles. They can be scattered to branch off and create secondary and even tertiary tracks, and show decoration similar to that on charged particle tracks. A study of possible correlations between these unexplained lines and the tracks of charged particles showed that they did not occur in conjunction with positron tracks from ^{40}K nuclei, but could occur in association with tracks of high-energy shower particles and also energetic muons. This suggested a dependence on the momentum of the charged particle, and led to the idea that they might result from atomic scattering events in the mica involving significant motion of atomic nuclei [23]. It was conjectured that energetic particles, if they possessed sufficient momen-

tum, could create a new type of lattice excitation when scattered by one or more atoms of the lattice. Then this excitation would propagate through the crystal as a highly localized, particle-like entity. Atomic motions within this localized entity would influence the weak energy barrier that inhibits impurity precipitation, thereby triggering the recording process. The observations in mica showed that this entity had to possess a unique feature, namely, a remarkable localization or stability against lateral spreading as it propagated over great distances, even exceeding 500mm in length in rare large crystals (see Fig. 1).

It was already known from the charged particle tracks that the track recording process responds only to particles or disturbances moving in the potassium mono-atomic planes in mica. From these facts and a study of the mica crystal structure, it was concluded that the new kind of lattice excitation is associated with *quasi-one-dimensional* behavior in the K (001)-plane. To reflect this behavior the new entity or object was called a *quodon*.

3 Numerical and analogue studies

To investigate the properties of these hypotherical quodons and their behavior in the mica, lattice studies were made, using both numerical (molecular dynamics) techniques and analogue models of coupled pendulums. Our aim is to model the nonlinear behavior of lattice motions of the K atoms, therefore it is important to understand the structure and the nature of the bonds involved.

In Fig. 2 we can see the crystal structure of muscovite. This mineral (chemical formula $KAl_2(AlSi_3)O_{10}(OH)_2$) belongs to the family of micas, which in turn are a particular type of *layer silicates*. These are more commonly known as clays and clay-related minerals. The basic units forming these silicates are sheets of tetrahedra and octahedra (O atoms being the vertex and Si atoms being the centers of these). Fortunately, muscovite is not a very complicated silicate. It is a 2:1 silicate, with dioctahedral occupancy and with interlayer cation potassium. This means that the layers are composed of one sheet of octahedra sandwiched between two sheets of tetrahedra, and the octahedral sheet has only two out of three octahedra occupied with a cation (Si^{+4} or Al^{+3} in this case). The layer is negatively charged as a whole, due to the substitution of all Si atoms in the tetrahedral sheet by Al. Then the layers stack and bond through electrostatic forces, thanks to the presence of an interlayer cation sheet, composed of K^+ ions. The crystal structure of muscovite has been studied experimentally in detail [16], and recently it has been modeled accurately with molecular mechanical methods [4].

We can see how the silicate layer forms a fairly rigid structure, with bonds which are predominantly covalent. The K atoms, though held by Coulomb

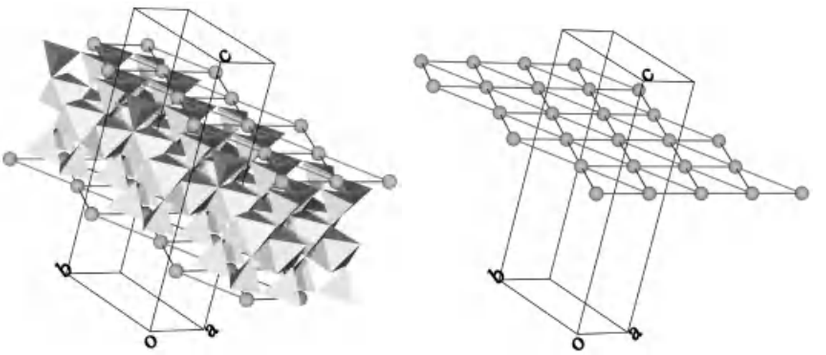

FIGURE 2. The structure of muscovite, emphasizing the potassium sheet.

forces in the direction perpendicular to the sheet, have easier motion on the sheet plane (the (001)-plane), only restrained by short-range forces. Therefore, a first approximation model consists of a 2D lattice of K atoms subject to the action of an external "on-site" potential, constructed as to mimic the presence of a totally rigid silicate layer above and below.

The molecular dynamic techniques available at the time could not easily be used to study the behavior of the K lattice following a simulated impact on an atom by a projectile. To explore this problem mechanical analogues were made (by FMR) to simulate a one-dimensional chain of atoms subject to both inter-particle and on-site forces. In particular, the nonlinearity of the forces in the analogues was made similar to those found from the molecular-mechanics studies. The analogue "atoms" were made from small permanent magnets that simulated the inter-particle forces. These magnets were suspended as rigid pendulums from fixed points and so used gravity to simulate the on-site forces. The strength of the inter-particle force could be varied by changing the spacing of the magnets along the chain. The strength of the on-site force could be changed by tilting the analogue away from a vertical position.

It was found that an impact on a particle in the chain quickly evolved into a traveling disturbance that retained its envelope shape. Within this envelope the particles executed anti-phase or optical-mode type oscillations, with a phase velocity that exceeded the group velocity of the envelope. The envelope was reflected if one particle in the chain was held fixed. Small perturbations of the chain, such as increasing the mass of one particle or the spacing between one pair, had little effect on the envelope. However, a large perturbation caused the envelope to separate into two of lower energy, one being reflected. By using a chain curved into a circle it was possible to follow such envelopes over many circuits round the circle. In summary, it was found that these localized excitations were remarkably robust and

closely simulated the properties expected of the postulated quodon. The results were confirmed and expanded in detail by a numerical model of the analogue model itself [22]. Everything indicated that these solutions were *moving breathers*, which have attracted a great deal of research among the condensed matter and dynamical systems community (for a review, see [9, 1]).

These one-dimensional models, however, could not account for the two-dimensional nature of the problem. For instance, it was already known that Toda-like solitons in a two-dimensional sheet spread rapidly sideways, failing to propagate more than about a hundred atomic spacings [15]. As for discrete breathers, they had been found to be mobile in many 1D models [10, 5], but not in higher-dimensional cases [2] (the solutions in Ref. [26] seem too extended to qualify as discrete breathers).

We then set out to study numerically a full-blown 2D model, in which the K ions could move in the plane. The silicate layer is a complicated structure of tetrahedra and octahedra, and both surfaces of this layer are formed by the basal oxygens of the tetrahedra. These triangular bases are connected with each other at the vertices, and form an hexagonal honeycomb structure. The interlayer, the K sheet, is therefore sandwiched between mirror images of this structure, the atoms occupying the dimples left at the center of such honeycomb cells. The result is that the K sheet forms a perfect triangular lattice, with six-fold symmetry (see again Fig. 2). The on-site potential created by the silicate layer preserves this symmetry: the projections of the basal oxygens on the K-plane fall right at the midpoints between K-K bonds, and the Al ions in the center of these tetrahedra project right on the center of the triangles formed by K ions[1]. In real mica crystals, there are always distortions to this ideal structure, like tetrahedral rotation. However, they normally preserve the six-fold symmetry for the environment of the K ions [4].

We studied several molecular-dynamics models of hexagonal 2D systems with different hexagonal symmetry, and eventually found solutions which perfectly fit with the quodon conjecture. More specifically, *quasi-one-dimensional* behavior is taken to mean that the excitation remains localized and moves along a favoured direction in the crystal, and that this phenomenon is assumed to be due to a particular type of crystal symmetry. This symmetry is such that if one of the atoms in the chain is disturbed from its equilibrium position *along this favoured direction*, then the restoring force is purely in the reverse direction to the displacement, i.e. there are no shear forces present [21]. The numerical 2D systems we studied have this property. The corresponding solitary waves that we found are *moving*

[1] Although the K atoms form a *triangular* lattice, we have used sometimes the term "hexgonal" instead. We hope the reader does not mistake it for a *honeycomb* structure, which is not the case.

discrete breathers of longitudinal type, and we describe in greater detail their properties in the next section.

4 Longitudinal moving breathers in 2D lattices

The numerical technique used in our models is purely classical molecular dynamics, although it is important to note that only first nearest neighbors are taken into account so far. The potentials used are those coming from molecular-mechanics studies, and although they are empirical, they are thought to be highly accurate for most modelling purposes.

However, for reasons to be shown shortly, it is not so important to get the detailed structure of the on-site potential to great accuracy, or even perfectly fitted interatomic potentials. We have settled for now with a model in which the K atoms interact through Lennard-Jones potentials,

$$W_{LJ}(r) = \frac{a^2}{72}\left(\left(\frac{a}{r}\right)^{12} - 2\left(\frac{a}{r}\right)^{6}\right); \qquad r \equiv |\vec{d}| \tag{14.1}$$

where $\vec{d} = \vec{u} - \vec{u}' - \vec{c}$ is the actual separation vector between the (i, j) and the (i', j')-th nearest neighbor sites. We used as coordinates the relative displacements from the lattice equilibrium positions, $\vec{u}$, and $\vec{c}$ are static vectors connecting nearest-neighbor equilibrium sites. The particular scaling of this potential, as well as that of all others, was chosen so as to have a frequency for low-amplitude oscillations of $\omega = 1$. The on-site potential is produced for each K atom by placing fixed Lennard-Jones atoms at positions above the plane, at the sites occupied by basal oxygens. For each K atom we took into account 12 oxygens: the 6 nearest neighbor ones (whose projections lie at midpoints of the K-K bonds) and 6 next-nearest neighbor ones (with projections at midpoints of bonds between neighboring K atoms). Note that these fixed oxygens are really off the plane, so that the distances to the K atoms are computed in full 3D Euclidean space (the z-coordinate being constant). This is in contrast with our initial model in [14], where we placed 6 fixed atoms in the plane, interacting through a Morse potential.

Using the intuition gained about discrete breathers in 1D models, it was judged that it was not the particular shape of the nonlinear potentials used which was of main importance. Rather, it was the *relative strength* between the inter-atomic and on-site potentials which was crucial for the existence of these solutions. In the simpler case of *stationary* breathers, it had been established rigorously how discrete breathers emerge from the limit of negligible inter-atomic potential (the so-called anti-continuum limit, see [11]). The case of moving breathers is much less understood, but it has been shown how one can tune the relative strengths of on-site and inter-particle potentials and eventually find regions of easy breather mobility [5]. These

regions appear to happen near bifurcations which connect the site-centered and bond-centered stationary breathers (see Ref. [12]), and the subject is being investigated in greater detail [6]. It is difficult to estimate how wide these regions are, other than by means of numerical simulation.

With these heuristics we found *moving longitudinal breathers* in our simulations. Just as in 1D models, it was observed that a too strong on-site potential hinders mobility, and favours the pinning of the excitations into stationary breathers (intrinsic localized modes). A too weak on-site potential usually destroys the breather by broadening and radiation into the background, due to resonances with the phonon band [12]. In between these extremes, there exists an ample region where it is easy to obtain moving breathers, with a more or less wide range of energies and velocities. The striking novelty here is that these breathers have a strong transversal focusing effect: the excitation is a sharply localized longitudinal wave, comprised of about 4–8 particles in the longitudinal direction, but practically only one particle in the transverse direction. Note that the focusing effect is clearly not an artifact of weak couplings: if one tries to create a moving breather under a certain energy threshold, the excitation dies out radiating in *all* crystal directions in a linear-like fashion. It all indicates that these breathers are the realization of the quodon conjecture.

Fig. 3 shows the profile of the longitudinal displacements of the K atoms along the breather path. These solutions travel very long distances along lattice directions, with negligible loss of energy by radiation. As in the analogue studies, the solutions are characterized by nearly out-of-phase motion of the atoms involved, with a frequency which seems to be independent of the envelope wave velocity. The internal vibration frequency is high (above the linear phonon band, due to the effective hardening nonlinearity), while the breather motion is slow, typically 20–60% of the sound velocity in the lattice. This indicates that they are truly moving breathers, with two dynamical degrees of freedom. Other types of lattice solitons, where carrier wave and soliton velocity go in unison, can be calculated with accurate procedures, such as continuation methods [8]. In some special cases, such as the discrete nonlinear Schrödinger equation, the same methods can be applied to calculate breathers [7]. But in general, moving breathers are much harder to obtain in such detail, and although some approaches have proved useful for very slow ones [5], we do not have accurate methods for most of the range of possible solutions observed in simulations.

As we mentioned above, our studies are based on heuristics and numerical simulation. However, the way moving breathers are produced in the simulation is very useful for the problem of mica. Instead of trying to isolate and "purify" a moving breather, it was interesting to know how they could be formed. It turned out that it is very easy to produce them, simply by giving some atoms certain amounts of energy. In practice, it is more controllable to give the particles an initial velocity impulse, as if an atomic collision took place. It was surprising to observe that giving an impulse to *just one*

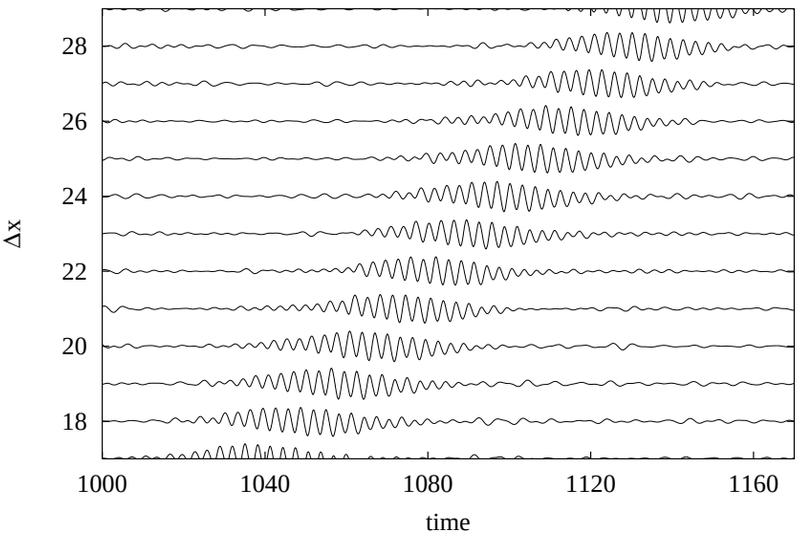

FIGURE 3. Longitudinal displacements vs. time for the atoms on the breather track. They have been scaled by a factor of five to make them more visible; actual displacements reach about 15–20% of the interatomic distance.

atom is enough to create the moving breather. There is initially a burst of radiation away from the disturbance, but once the breather emerges it sheds very little radiation and may travel tens of thousands of sites without appreciable degradation. Furthermore, there is robustness against deflections of this initial momentum: we found that deviations as great as 15° from the crystal axis would still produce a breather.

In summary, these longitudinal breathers are a very robust phenomenon appearing in 2D lattices, in the sense that they do not need a special shape for the potentials, or a very constrained range of parameters. We have found them in several types of Lennard-Jones, Morse and other lattices, within ample windows of parameter ranges. A common feature is the need for an on-site potential, although some recent results suggest that it is possible to reach the limit of null on-site terms in some particular models. After all, the exact extent of the parameter ranges in which these breathers exist will vary from one model to another, and the intrinsic nonlinear nature of the problem makes it difficult to predict it a priori. For the problem of mica, we are now trying to introduce the most realistic potentials we can find [4] and check whether the combination of on-site and inter-atomic forces for the K atoms allow for moving breathers. Our initial studies including only short range forces suggest so [13]. It is now needed to study the effect of

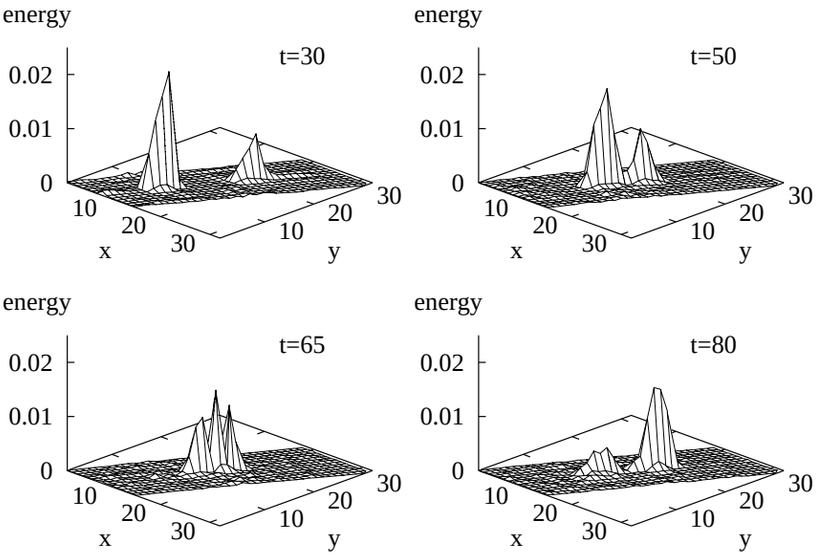

FIGURE 4. Head-on collision of two breathers with different energies and frequencies, in an hexagonal lattice. In this case they emerge out of the collision almost unaffected.

long-range Coulombic forces on these moving solutions, which is not well known even for 1D models.

5 Breather collisions

We have experimented with a number of different configurations to study breather-breather collisions. Two breathers can collide head-to-head in-line, or head-to-head but on parallel tracks, or at angles of 60° or 120°. In the latter case the collision can also be "head-on" or staggered (i.e. one breather just leaving a particular site as the other is arriving). The results will also depend on the relative phases of the underlying carrier waves and on the energies of each breather. The number of possible permutations are clearly large and will be described in more details elsewhere. We give here just one example, an in-line head-to-head collision between two breathers of equal amplitude, as shown in Fig. 4, a–d.

In this case we see that the two breathers are reflected almost elastically with very little radiation, either transverse or longitudinal. However this result is very dependent on the relative phases of the breathers, with a slightly different starting position (1 lattice spacing further apart), the two

breathers coalesce into a stationary "super-breather", two breathers bound together but with a time-periodic small separation which decreases slowly with time.

These results just confirm that discrete breathers are not exact solitons, and therefore they might not always have the expected robustness with respect to collisions. On the other hand, they have a different kind of robustness, that which is conferred by their high genericity. Exact solitons appear in integrable models, but discrete breathers appear generically in an extremely large class of lattices.

6 Conclusions and further applications

It does not pass unnoticed that such a generic phenomenon as these lattice breathers should have applications to many other fields, and not just mica structures. Here we sample some of our speculations with respect to the ubiquity of these new solutions.

6.1 Application to sputtering

These successes in simulating the predicted quodons and their behavior prompted application of the quodon concept, in particular, to sputtering at a crystal surface and to other crystals with a layered structure. Sputtering, the ejection of atoms from a crystal surface that is bombarded by energetic particles, has a long history and is of industrial importance. Pioneering work was done by Wehner in 1956 [27], who showed that sputtering was intimately related to the speed of sound in a material. He found also that sputtered atoms were ejected in atomic-chain directions and that particles bombarding a surface at normal incidence gave a higher energy threshold for sputtering than at oblique incidence. We believe that the new found moving breathers could bring many new insights into this old subject, which has never been given detailed atomistic treatment. Initial experiments carried out with our analogue models are finding agreement with the phenomenological theories of Wehner, and indicate the importance of understanding the nonlinear dynamics of lattices.

6.2 Application to layered HTSC materials

The discovery of high temperature superconductivity (HTSC) in layered cuprates raised the question of the possibility of quasi-one-dimensional behavior in those layers. A similar molecular dynamic study to that of the mica system was made of several known HTSC materials containing cuprate layers to find out if any of the many different layers might be capable of allowing quasi-one-dimensional behavior. The materials examined

were $YBa_2Cu_3O_7$, $Y_2Ba_4Cu_6O_{13}$, and La_2CuO_4. It was found that in all three materials movement of Cu and O atoms in chain directions in the non-puckered (001)-planes was one-dimensional like in that no shear forces were created in the surrounding lattice [21]. Some Y atoms also showed quasi-one-dimensional behavior. These findings suggest that quodons might be created and propagate in such materials.

The dominant mechanism of charge-pair coupling in HTSC is still not known. However, it has been shown recently that in the HTSC materials $Bi_2Sr_2Ca_{m-1}Cu_mO_y$, ($m = 1$ and 2) the T_c is independent of the separation distance between the cuprate layers [3] and is the same as that for a single layer. This strongly suggests that the pairing mechanism operates in a 2D layer. From study of the published structures of these two materials it seems both may show quasi-one-dimensional properties in certain layers. This needs to be examined in detail.

In order to investigate the possibilities of breathers mediating the coupling effects on charges in HTSC in tetragonal lattices, we have carried out a similar study to the present paper but using square 2D lattices instead of hexagonal lattices. We find localized breather solutions for a wide range of model parameters and initial disturbances. These results will be reported elsewhere.

Acknowledgments: JLM acknowledges a postdoctoral *Marie Curie* TMR fellowship from the EC (grant no. ERBFMBICT972761). We would also like to thank Paz Vaqueiro for providing us with accurate data for the crystal structures.

7 References

[1] S. Aubry. Breathers in nonlinear lattices: existence, linear stability and quantization. *Physica D*, 103:201–250, 1997.

[2] V. M. Burlakov, S. A. Kisilev, and V. N. Pyrkov. Computer simulation of intrinsic localized modes in one-dimensional and two-dimensional anharmonic lattices. *Phys. Rev. B*, 42:4921, 1990.

[3] Jin-Ho Choy, Soon-Jae Kwon, and Gyeong-Su Park. High T_c superconductors in the two-dimensional limit. *Science*, 280:1589–1592, 1998.

[4] D. R. Collins. *Computer simulations and netron scattering studies of layer silicates minerals*. PhD thesis, University of Keele, 1990.

[5] T. Cretegny and S. Aubry. Mobility and reactivity of discrete breathers. *Physica D*, 119:34–46, 1998.

[6] T. Cretegny, S. Aubry, and J. L. Marín. Exchange of stability and mobility of discrete breathers. *in preparation*.

[7] D. B. Duncan, J. C. Eilbeck, H. Feddersen, and J. A. D. Wattis. Solitons on lattices. *Physica D*, 68:1–11, 1993.

[8] J. C. Eilbeck and R. Flesch. Calculation of families of solitary waves on discrete lattices. *Phys. Lett. A*, 149(4):200–202, 1990.

[9] S. Flach and C. R. Willis. Discrete breathers. *Physics Reports*, 295:181–264, 1998.

[10] K. Hori and S. Takeno. Low-frequency and high-frequency moving anharmonic localized modes in a one-dimensional lattice with quartic anharmonicity. *J. Phys. Soc. Jpn.*, 61:4263, 1992.

[11] R. S. MacKay and S. Aubry. Proof of existence of breathers for time-reversible or Hamiltonian networks of weakly coupled oscillators. *Nonlinearity*, 7:1623–1643, 1994.

[12] J. L. Marín, S. Aubry, and L. M. Floría. Intrinsic localized modes: discrete breathers. existence and linear stability. *Physica D*, 113:283–292, 1998.

[13] J. L. Marín, J. C. Eilbeck, and F. M. Russell. Localized moving breathers in 2-d lattices. *in preparation.*

[14] J. L. Marín, J. C. Eilbeck, and F. M. Russell. Localized moving breathers in a 2-d hexagonal lattice. *Phys. Lett. A*, 1998. *to appear.*

[15] Yu V. Martynenko and P. G. Moscovkin. Solitons in radiation physics of crystals. *Radiation Effects and Defects in Solids*, 117:321–328, 1991.

[16] E. W. Radoslovich. The structure of muscovite, $KAl_2(Si_3Al)O_{10}(OH)_2$. *Acta Cryst.*, 13:919–930, 1960.

[17] F. M. Russell. The observation in mica of charged particles from neutrino interactions. *Phys. Lett. B*, 25:298–300, 1967.

[18] F. M. Russell. Identification and selection criteria for charged lepton tracks in mica. *Nucl. Tracks Radiat. Meas.*, 15:41–44, 1988.

[19] F. M. Russell. Positive charge transport in layered crystalline solids. *Phys. Lett. A*, 130:489–491, 1988.

[20] F. M. Russell and D. R. Collins. Lattice-solitons in radiation damage. *Nuclear instruments & methods in physics research B*, 105:1–4, 1995.

[21] F. M. Russell and D. R. Collins. Anharmonic excitations in high T_c materials. *Phys. Lett. A*, 216:197–202, 1996.

[22] F. M. Russell, Y. Zolotaryuk, and J. C. Eilbeck. Moving breathers in a chain of magnetic pendulums. *Phys. Rev. B*, 55:6304–6308, 1997.

[23] D. Schlosser, K. Kroneberger, M. Schosnig, F. M. Russell, and K. O. Groeneveld. Search for solitons in solids. *Radiat. Meas.*, 23:209–213, 1994.

[24] A. J. Sievers and S. Takeno. Intrinsic localized modes in anharmonic crystals. *Phys. Rev. Lett.*, 61:631–633, 1988.

[25] J. W. Steeds, F. M. Russell, and W. J. Vine. Formation of epidote fossil positron tracks in mica. *Optik*, 92:149–154, 1993.

[26] J. M. Tamga, M. Remoissenet, and J. Pouget. Breathing solitary waves in a sine-Gordon two-dimensional lattice. *Phys. Rev. Lett.*, 75:357–361, 1995.

[27] G. K. Wehner. Controlled sputtering of metals by low-energy Hg ions. *Phys. Rev.*, 102:690–704, 1956.

15

Scale Competition in Nonlinear Schrödinger Models

Yu.B. Gaididei
P.L. Christiansen
S.F. Mingaleev

ABSTRACT Three types of nonlinear Schrödinger models with multiple length scales are considered. It is shown that the length-scale competition universally gives rise to new localized stationary states. Multistability phenomena with a controlled switching between stable states become possible.

1 Introduction

The basic dynamics of deep water and plasma waves, light pulses in nonlinear optics and charge and energy transport in condensed matter and biophysics [3, 29, 32, 12] is described by the fundamental nonlinear Schrödinger (NLS) equation

$$i\frac{\partial}{\partial t}\psi + L^2\partial_x^2\psi + V|\psi|^2\psi + f(x)\psi = 0\ , \tag{15.1}$$

where $\psi(x,t)$ is the complex amplitude of quasi-monochromatic wave trains or the wave function of the carriers. The second term represents the dispersion and L is the dispersion length (e.g. in the theory of charge (energy) transfer $L^2 = \hbar^2/2m$ with m being an effective mass). The nonlinear term, $V|\psi|^2\psi$, describes a self-interaction of the quasiparticle caused either by its interaction with low-frequency excitations (phonons, plasmons, etc.) [7] or by the intensity dependent refractive index of the material (Kerr effect) [23]. The function $f(x)$ is a parametric perturbation which can be a localized impurity potential, a disorder potential, a periodic refractive index, an external electric field, etc. It is well known that as a result of competition between dispersion and nonlinearity nonlinear waves with properties of particles, solitons, arise. One may also say that this competition leads to the appearance of the new length-scale: the width of the soliton, $\zeta = L/\sqrt{V}$. The presence of the parametric perturbation $f(x)$ introduces additional interplays between nonlinearity, dispersion and perturbations. In the recent paper by Bishop *et al.* [1] the concept of competing length-scales and

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 307−321, 2000.

time-scales was emphasized. In particular, Scharf and Bishop [26, 27] have discussed the effects of a periodic potential ($f(x) = \epsilon \cos(2\pi x/\zeta_p)$) on the soliton of the NLS equation, and shown on the basis of an averaged NLS equation that for $\zeta_p/\zeta \lesssim 1$ or $\gtrsim 1$ that the periodic potential leads to a simple renormalization of the solitons and creates a 'dressing' of the soliton. But when $\zeta_p \sim \zeta$ there is a crucial length-scale competition which leads to the destruction of the soliton. Another interesting example of the length-scale competition was provided by Ref. [15] where the authors showed that the propagation of intense soliton-like pulses in systems described by the one-dimensional NLS equation may be left practically unaffected by the disorder (when $f(x)$ is a Gaussian δ-correlated process). This theoretical prediction has recently been confirmed experimentally using nonlinear surface waves on a superfluid helium film [13].

The goal of this paper is to extend the concept of the length-scale competition to the essentially non-integrable systems nonlocal dispersion and unstable stationary states.

2 Excitations in nonlinear Kronig-Penney models

Wave propagation in nonlinear photonic band-gap materials and in periodic nonlinear dielectric superlattices [30, 31] consisting of alternating layers of two dielectrics: nonlinear and linear, is governed by the NLS equation

$$\begin{aligned} i\partial_t \psi(x,z,t) &+ \zeta^2 \left(\partial_x^2 + \partial_z^2\right)\psi(x,z,t) \\ &+ w \sum_n \delta(x - x_n)|\psi(x,z,t)|^2 \psi(x,z,t) = 0 \,, \end{aligned} \tag{15.2}$$

where $x_n = n\,\ell$ is the coordinate of the n-th nonlinear layer (ℓ is the distance between the adjacent nonlinear layers), and it is assumed that the width w of the nonlinear layer is small compared to the soliton width ζ within the layer. In this case the problem can be described by the nonlinear Kronig-Penney model given by Eq. (15.2). It was shown in Ref. [9] that the field $\psi(x,z,t)$ can be expressed in terms of the complex amplitudes $\psi_n(z,t) \equiv \psi(x_n,z,t)$ at the nonlinear layers. The complex amplitudes $\psi_n(z,t)$ can be found from the set of pseudo-differential equations

$$\frac{\zeta^2\hat{\kappa}}{\sinh \ell\hat{\kappa}}(\psi_{n+1} + \psi_{n-1}) - \frac{2\zeta^2\hat{\kappa}}{\tanh \ell\hat{\kappa}}\,\psi_n + w|\psi_n|^2\,\psi_n = 0 \tag{15.3}$$

with periodic boundary conditions $\psi_{n+N} = \psi_n$, where N is the number of layers. In Eq. (15.3) the operator $\hat{\kappa}$ is defined as $\hat{\kappa}\psi = \zeta^{-1}\sqrt{-i\,\partial_t - \zeta^2\,\partial_z^2}\psi$. Eq. (15.2) has an integral of motion – the norm (in nonlinear optics this quantity is often called the power) $P = \int_{-\infty}^{\infty} |\psi|^2 dxdz$.

For the excitation pattern where the complex amplitudes are the same in all nonlinear layers, $\psi_n(z,t) = \Psi(z,t)$, we get

$$\sqrt{-i\partial_t - \zeta^2\,\partial_z^2}\,\tanh\left(\frac{\ell}{2\zeta}\sqrt{-i\partial_t - \zeta^2\,\partial_z^2}\right)\Psi - \frac{w}{2\zeta}|\Psi|^2\Psi = 0\,. \qquad (15.4)$$

Eq. (15.4) clearly shows the existence and competition of two characteristic length-scales: the interlayer spacing ℓ and the size of the soliton in the nonlinear layer ζ. When $\ell \ll \zeta$ one can expand the hyperbolic tangent and Eq. (15.4) takes the form of usual NLS equation

$$\left(i\partial_t + \zeta^2\,\partial_z^2\right)\Psi + \frac{w}{\ell}|\Psi|^2\Psi = 0\,. \qquad (15.5)$$

In the opposite limit, when $\ell \gg \zeta$ and $\tanh\left(\frac{\ell}{2\zeta}\sqrt{\cdots}\right) \simeq 1$, Eq. (15.4) takes the form

$$\sqrt{-i\partial_t - \zeta^2\,\partial_z^2}\Psi - \frac{w}{2\zeta}|\Psi|^2\Psi = 0\,, \qquad (15.6)$$

which, for static distributions ($\partial_t\Psi = 0$), reduces to the nonlinear Hilbert-NLS equation recently introduced in Ref. [11]. It is noteworthy that in contrast to usual NLS solitons, the localized solutions of the nonlinear Hilbert-NLS equation have algebraic tails [11].

It is worth to note the close relation of the problem under consideration to the theory of the long internal gravity waves in a stratified fluid with a finite depth h (see e.g. [16]) which are described by the equation

$$\partial_t u + \frac{1}{h}\partial_x u + 2u\partial_x u + T\partial_x^2 u = 0\,, \qquad (15.7)$$

where $T(\cdot)$ is the singular integral operator given by

$$(Tf)(x) = \frac{1}{2h}\,p.v.\int_{-\infty}^{\infty}\coth\left(\frac{\pi(y-x)}{2h}\right)f(y)dy \qquad (15.8)$$

($p.v.$ means the principal value integral). In the shallow water limit ($h \to 0$) the dynamics is described by the Korteweg-de-Vries equation, $\partial_t u + \frac{h}{3}\partial_x^3 u + 2u\partial_x u = 0$, while the Benjamin-Ono equation, $\partial_t u + H\partial_x^2 u + 2u\partial_x u = 0$, governs the water wave motion in the deep-water limit ($h \to \infty$). Here $(Hf)(x) = \frac{1}{\pi}\,p.v.\int_{-\infty}^{\infty} dy f(y)/(y-x)$ is the Hilbert transform.

Stationary states of the system are of interest. Thus we consider solutions of the form $\Psi(z,t) = \phi(z)\exp(i\Lambda t)$, where Λ is the nonlinear frequency and $\phi(z)$ is the real shape function. Since Eq. (15.3) is Galilean invariant, standing excitations can always be Galileo boosted to any velocity in z-direction. For the shape function $\phi(z)$, we obtain a nonlinear eigenvalue

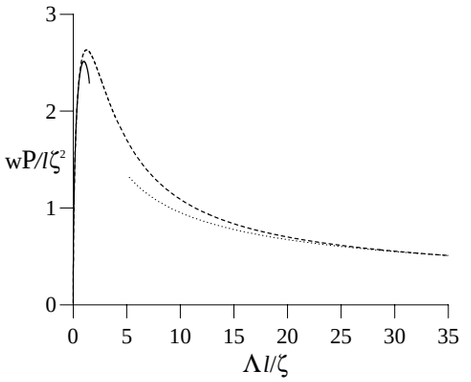

FIGURE 1. Power P of the stationary state $\psi_n(z,t) = e^{i\Lambda t}\phi(z)$ vs the nonlinear frequency Λ. Numerical results from Eq. (15.9) (dashed line), Padé approximation (full line) and the asymptotic relation $P \sim \Lambda^{-1/2}$ as $\Lambda \to \infty$ (dotted line) [9].

problem in the form

$$\sqrt{\Lambda - \zeta^2\,\partial_z^2}\,\tanh\left(\frac{\ell}{2\zeta}\sqrt{\Lambda - \zeta^2\,\partial_z^2}\right)\phi - \frac{w}{2\zeta}\phi^3 = 0\,. \tag{15.9}$$

Simple scaling arguments show that in the low-frequency limit ($\Lambda\ell^2/\zeta^2 \to 0$) the norm behaves in the same way as in the case of usual NLS equation (15.5): $P \sim \sqrt{\Lambda}$. When $\Lambda\ell^2/\zeta^2 \to \infty$ the norm P is a monotonically decreasing function: $P \sim 1/\sqrt{\Lambda}$. From the analysis of Ref. [9], it follows that the norm $P(\Lambda)$ is a non-monotonic function with a local maximum at $\Lambda_m \approx 1.25\zeta/\ell$ (see Fig. 1). Thus, the stationary states exist only in a finite interval, $0 \le P \le P(\Lambda_m)$, and for each value of norm in this interval there are two stationary states. This is an intrinsic property of the nonlinear Schrödinger superlattice system.

Discussing the stability of the stationary states satisfying Eq. (15.2), there are two sources of instability to be considered: longitudinal and transversal perturbations. The perturbations of the first type are of the same symmetry with respect to transversal degrees of freedom as the stationary states of Eq. (15.9), while the second type of perturbations breaks this symmetry. It was shown in [9] that stationary states which correspond to the branch with $dP/d\Lambda < 0$ are unstable due to the longitudinal perturbations. The states with $\Lambda > (4\zeta^2/3\ell^2)\sin^2(\pi/N)$ are, in their turn, unstable due to the transversal perturbations. Thus, one can expect stable stationary solutions for nonlinear frequencies satisfying the condition

$$\Lambda < \frac{\zeta^2}{\ell^2}\min\{\frac{4}{3}\sin^2(\frac{\pi}{N}), \Lambda_m\}\,. \tag{15.10}$$

In particular, this means that the stationary state $\psi_n(z,t) = e^{i\Lambda t}\phi(z)$ can neither exist in the case of only one nonlinear layer ($\ell \to \infty$) nor in the quasi-continuum limit ($N \to \infty$). But in the latter case the system supports stationary states which are localized in both spatial directions (see Ref. [9] for details).

3 Discrete NLS models with long-range dispersive interactions

In the main part of the previous studies of the discrete NLS models the dispersive interaction was assumed to be short-ranged and a nearest-neighbor approximation was used. However, there exist physical situations that definitely can not be described in the framework of this approximation. The DNA molecule contains charged groups, with long-range Coulomb interaction ($1/r$) between them. The excitation transfer in molecular crystals [6] and the vibron energy transport in biopolymers [28] are due to transition dipole-dipole interaction with $1/r^3$ dependence on the distance, r. The nonlocal (long-range) dispersive interaction in these systems provides the existence of additional length-scale: the radius of the dispersive interaction. We will show that it leads to the bifurcative properties of the system due to both the competition between nonlinearity and dispersion, and the interplay of long-range interactions and lattice discreteness.

In some approximation the equation of motion is the nonlocal discrete NLS equation of the form

$$i\frac{d}{dt}\psi_n + \sum_{m\neq n} J_{n-m}(\psi_m - \psi_n) + |\psi_n|^2\psi_n = 0\,, \tag{15.11}$$

where the long-range dispersive coupling is taken to be either exponentially, $J_n = J\,e^{-\beta|n|}$, or algebraically, $J_n = J\,|n|^{-s}$, decreasing with the distance n between lattice sites. In both cases the constant J is normalized such that $\sum\limits_{n=1}^{\infty} J_n = 1$ for all β or s. The parameters β and s are introduced to cover different physical situations from the nearest-neighbor approximation ($\beta \to \infty,\ s \to \infty$) to the quadrupole-quadrupole ($s = 5$) and dipole-dipole ($s = 3$) interactions. The Hamiltonian $H = \sum\limits_{n,m} J_{n-m}|\psi_n - \psi_m|^2 - \frac{1}{2}\sum\limits_{n}|\psi_n|^4$, which corresponds to the set of Eqs. (15.11), and the number of excitations $N = \sum\limits_{n}|\psi_n|^2$ are conserved quantities.

We are interested in stationary solutions of Eq. (15.11) of the form $\psi_n(t) = \phi_n \exp(i\Lambda t)$ with a real shape function ϕ_n and a frequency Λ. This gives the governing equation for ϕ_n

$$\Lambda\phi_n = \sum_{m\neq n} J_{n-m}(\phi_m - \phi_n) + \phi_n^3\,, \tag{15.12}$$

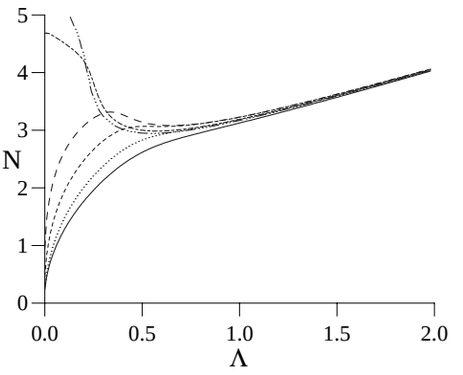

FIGURE 2. Number of excitations, N, versus frequency, Λ, found numerically from Eq. (15.12) for $s = \infty$ (full), 4 (dotted), 3 (short-dashed), 2.5 (long-dashed), 2 (short-long-dashed), 1.9 (dashed-dotted) [11].

which is the Euler-Lagrange equation for the problem of minimizing H under the constraint $N =$ constant.

Figure 2 shows the dependence $N(\Lambda)$ obtained from direct numerical solution of Eq. (15.12) for algebraically decaying J_{n-m}. A monotonic function is obtained only for $s > s_{cr}$. For $2 < s < s_{cr}$ the dependence becomes non-monotonic (of $\mathcal{N}$-type) with a local maximum and a local minimum. These extrema coalesce at $s = s_{cr} \simeq 3.03$. For $s < 2$ the local maximum disappears. The dependence $N(\Lambda)$ obtained analytically using the variational approach is in a good qualitative agreement with the dependence obtained numerically (see [11]). Thus the main features of all discrete NLS models with dispersive interaction J_{n-m} decreasing faster than $|n-m|^{-s_{cr}}$ coincide qualitatively with the features obtained in the nearest-neighbor approximation where only one stationary state exists for any number of excitations, N. However in the case of long-range nonlocal NLS equation (15.11), i.e. for $2 < s < s_{cr}$, there exist for each N in the interval $[N_l(s), N_u(s)]$ three stationary states with frequencies $\Lambda_1(N) < \Lambda_2(N) < \Lambda_3(N)$. In particular, this means that in the case of dipole-dipole interaction ($s = 3$) multiple solutions exist. It is noteworthy that similar results are also obtained for the dispersive interaction of the exponentially decaying form. In this case the bistability occurs for $\beta \leq 1.67$. According to a theorem which was proven in [21], the necessary and sufficient stability criterion for the stationary states is $dN/d\Lambda > 0$. Therefore, we can conclude that in the interval $[N_l(s), N_u(s)]$ there are only two linearly stable stationary states: $\Lambda_1(N)$ and $\Lambda_3(N)$. The intermediate state is unstable since $dN/d\Lambda < 0$ at $\Lambda = \Lambda_2$.

At the end points ($\Lambda(N_l)$ and $\Lambda(N_u)$) the stability condition is violated, since $(dN/d\Lambda)_s$ vanishes. Constructing the locus of the end points we obtain the curve that is presented in Fig. 3. This curve bounds the region of

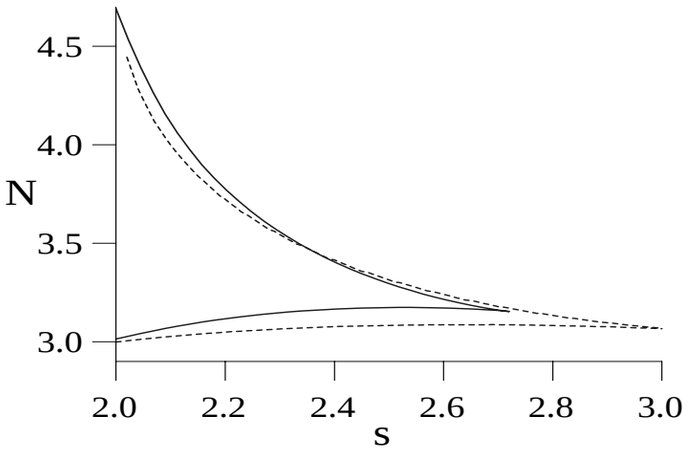

FIGURE 3. Shows endpoints of the bistability interval for N versus dispersion parameter s. For $s = s_{cr}$ the endpoints coalesce. Analytical dependence (full) gives $s_{cr} \simeq 2.72$. Numerical dependence (dashed) gives $s_{cr} \simeq 3.03$ [11].

bistability. It is analogous to the critical curve in the van der Waals' theory of liquid-vapor phase transition [19]. Thus in the present case we have a similar phase transition like behavior where two phases are the continuum states and the intrinsically localized states, respectively. The analog of the temperature is the dispersive parameter $s(\beta)$.

The shapes of three stationary states in the interval of bistability differ significantly (see Fig. 4). The low frequency states are wide and continuum-like while the high frequency solutions represent intrinsically localized states with a width of a few lattice spacings. It can be shown [11] that the inverse widths of these two states are $\alpha_1 \approx (N/8J)^{1/(s-2)} = (N/8J)^{\ln \xi/(1-2\ln \xi)}$, $\alpha_3 \approx \ln(N/J)$ with $\xi = \exp(1/s)$ being the characteristic length scale of the dispersive interaction which is defined as the half maximum interaction distance (expressed in lattice spacings). It is seen from these expressions that the existence of two so different soliton states for the same value of the excitation number, N, is due to the presence of two different length scales in the system: the usual scale of the NLS model which is related to the competition between nonlinearity and dispersion (expressed in terms of the ratio N/J) and the range of the dispersive interaction ξ.

Having established the existence of bistable stationary states in the non-local discrete NLS system, a naturally arising question concerns the role of these states in the full dynamics of the model. In particular, it is of interest

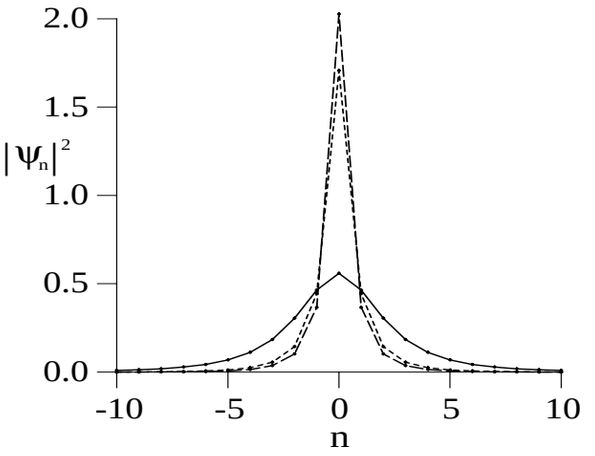

FIGURE 4. Shapes of three stationary states for $s = 2.5$ and $N = 3.1$. The stable: $\Lambda = 0.21$ (full), $\Lambda = 0.74$ (long-dashed). The unstable: $\Lambda = 0.57$ (short-dashed) [11].

to investigate the possibility of switching between the stable states under the influence of external perturbations, and to find out what type of perturbations can be used to control the switching. Switching of this type is important for example in the description of nonlinear transport and storage of energy in biomolecules like the DNA, since a mobile continuum-like excitation can provide action at distance while the switching to a discrete, pinned state can facilitate the structural changes of the DNA [8]. As it was shown recently in [14], switching will occur if the system is perturbed such that an internal, spatially localized and symmetrical mode ("breathing mode") of the stationary state is excited above a threshold value.

We will in the sequel mainly discuss the case when the matrix element of excitation transfer, J_{n-m}, decreases exponentially with the distance $|n-m|$. For $\beta = 1$ the multistability occurs in the interval $3.23 \leq N \leq 3.78$. It is worth noticing, however, that the scenario of switching described below remains qualitatively unchanged for all values of $\beta \leq 1.67$, and also for the algebraically decaying dispersive coupling with $2 \leq s \leq 3.03$.

An illustration of how the presence of an internal breathing mode can affect the dynamics of a slightly perturbed stable stationary state is given in Figs. 5 and 6. To excite the breathing mode, we apply a spatially symmetrical, localized perturbation, which we choose to conserve the number of excitations in order not to change the effective nonlinearity of the sys-

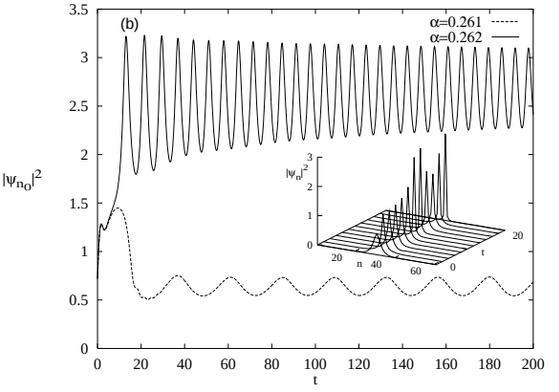

FIGURE 5. Switching from continuum-like to discrete state for $\beta = 1$. The initial state ϕ_n has the frequency $\Lambda \simeq 0.31$ and $N = 3.6$. The time evolution of $|\psi_{n_0}(t)|^2$ when a phase torsion is applied to the center site with $\theta = 0.261$ (lower curve) and $\theta = 0.262$ (upper curve), respectively; inset shows time evolution of $|\psi_n(t)|^2$ for $\theta = 0.262$ [14].

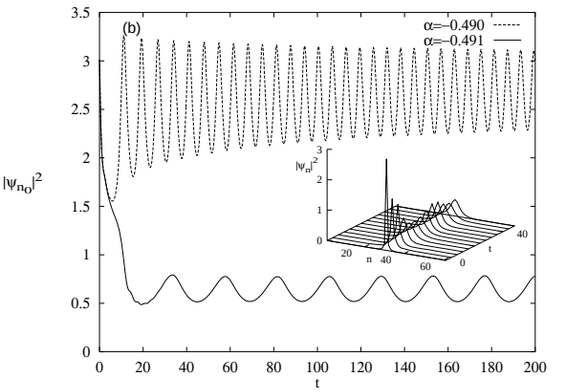

FIGURE 6. Switching from discrete to continuum-like state for $\beta = 1$. The initial state ϕ_n has the frequency $\Lambda \simeq 1.423$ and $N = 3.6$. Same as Fig.5 with $\theta = -0.490$ (upper curve) and $\theta = -0.491$ (lower curve), respectively; inset shows time evolution of $|\psi_n(t)|^2$ for $\theta = -0.491$ (only a part of a larger system is shown) [14].

tem. The simplest choice, which we have used in the simulations shown here, is to kick the central site n_0 of the system at $t = 0$ by adding a parametric force term of the form $\theta\delta_{n,n_0}\delta(t)\psi_n(t)$ to the left-hand-side of Eq. (15.11). As can easily be shown, this perturbation affects only the site n_0

at $t = 0$, and results in a "twist" of the stationary state at this site with an angle θ, i.e. $\psi_{n_0}(0) = \phi_{n_0}\, e^{i\theta}$. The immediate consequence of this kick is, as can been deduced from the form of Eq. (15.11), that $d\left(|\psi_{n_0}|^2\right)/dt$ will be positive (negative) when $\theta > 0$ ($\theta < 0$). Thus, we choose $\theta > 0$ to obtain switching from the continuum-like state to the discrete state, while we choose $\theta < 0$ when investigating switching in the opposite direction. We find that in a large part of the multistability regime there is a well-defined threshold value θ_{th}: when the initial phase torsion is smaller than θ_{th}, periodic, slowly decaying "breather" oscillations around the initial state will occur, while for strong enough kicks (phase torsions larger than θ_{th}) the state switches into the other stable stationary state.

It is worth remarking that the particular choice of perturbation is not important for the qualitative features of the switching, as long as there is a substantial overlap between the perturbation and the internal breathing mode. We also believe that the mechanism for switching described here applies for any multistable system where the instability is connected with a breathing mode.

4 Stabilization of nonlinear excitations by disorder

In this section we discuss disorder effects in NLS models. Usually investigations of disorder effects have been carried out on systems that are integrable - soliton bearing - in the absence of disorder. A common argument is that the equations, despite their exact integrability, provide a sufficient description of the physical systems to display the essential behavior. However, the more common physical situation is that integrability, and thus the exact soliton, is absent. A relevant example of such an equation is the two-dimensional (or higher-dimensional) NLS equation. The two-dimensional NLS equation is nonintegrable and possesses an unstable ground state solution which, in the presence of perturbations, either collapses or disperses (see e.g. [24, 25]).

We consider a quadratic two-dimensional lattice with the lattice spacing equal to unity. The model is given by the Lagrangian

$$L = \frac{i}{2} \sum_{n,m} \left(\psi^*_{n,m} \frac{d}{dt} \psi_{n,m} - c.c. \right) - H \ , \tag{15.13}$$

where

$$\begin{aligned} H &= \sum_{n,m} \left(|\psi_{n+1,m} - \psi_{n,m}|^2 + |\psi_{n,m+1} - \psi_{n,m}|^2 \right. \\ &- \left. \frac{1}{2} |\psi_{n,m}|^4 - \epsilon_{n,m} |\psi_{n,m}|^2 \right) \end{aligned} \tag{15.14}$$

is the Hamiltonian of the system. In Eqs. (15.13) and (15.14) (n, m) is the lattice vector (n and m are integers). The first two terms in Eq. (15.14) correspond to the dispersive energy of the excitation, the third term describes a self-interaction of the excitation and the fourth term represents diagonal disorder in the lattice. Here the random functions $\epsilon_{n,m}$ are assumed to have Gaussian distribution with the probability $p(\epsilon_{n,m}) = \exp[-(\epsilon_{n,m}/\eta)^2]/\eta\sqrt{\pi}$ with the autocorrelation function $\langle \epsilon_{n,m}\epsilon_{n',m'} \rangle = \eta^2 \delta_{n\,n'}\delta_{m\,m'}$, where the brackets $\langle ....\rangle$ denote averaging over all realizations of the disorder. From the Lagrangian (15.13) we obtain the equation of motion for the excitation function in the form

$$\begin{aligned} i\frac{d}{dt}\psi_{m,n} &+ (\psi_{m,n-1} + \psi_{m,n+1} + \psi_{m+1,n} + \psi_{m-1,n} - 4\psi_{m,n}) \\ &+ |\psi_{m,n}|^2\psi_{m,n} + \epsilon_{m,n}\psi_{m,n} = 0 \; . \end{aligned} \tag{15.15}$$

Eq. (15.15) conserves the norm $N = \sum_{n,m} |\psi_{n,m}|^2$ and the Hamiltonian H.

We are interested in the stationary solutions of Eq. (15.15) of the form

$$\psi_{n,m}(t) = \phi_{n,m} \exp(i\Lambda t) \; , \tag{15.16}$$

with a real shape function $\phi_{n,m}$ and a nonlinear frequency Λ.

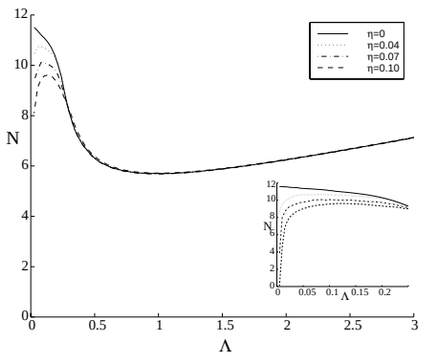

FIGURE 7. The norm N versus nonlinear frequency Λ for various disorder strengths η. Homogeneous case $\eta = 0$ (solid line), $\eta = 0.04$ (dotted line), $\eta = 0.07$ (dashed-dotted line) and $\eta = 0.1$ (dashed line) [10].

Eq. (15.15) together with Eq. (15.16) constitute a nonlinear eigenvalue problem which can be solved numerically using the techniques described in Ref. [5]. N versus Λ in the absence and in the presence of disorder is shown in Fig. 7. It has previously been shown [22, 20, 21, 18] that the linear

stability of the stationary states in the discrete case is determined by the condition $dN/d\Lambda > 0$. Thus, in the case without disorder (solid curve in Fig. 7) the low-frequency ($0 \leq \Lambda \leq \Lambda_{min} = 1.088$) nonlinear excitations in the discrete two-dimensional NLS model are unstable. It is important that in the continuum limit ($\Lambda \to 0$) the norm $N(\Lambda)$ tends to the non-zero value $N_c \simeq 11.7$.

The other curves in Fig. 7 show the dependence N on Λ for the stationary solutions of Eq. (15.15) in the presence of disorder. The results have been obtained as averages of 150 realizations of the disorder. Several new features arise as a consequence of the disorder. In the continuum limit ($\Lambda \to 0$) we no longer have $N = N_c$ with $dN/d\Lambda = 0$. Instead we have $N \to 0$ with $dN/d\Lambda > 0$ signifying that the disorder stabilizes the excitations in the low-frequency limit. The disorder creates a stability window such that a bistability phenomenon emerges. Consequently there is an interval of the excitation norm in which two stable excitations with significantly different widths have the same norm.

Furthermore, we see that the disorder creates a gap at small Λ in which no localized excitations can exist, and that the size of this gap apparently is increased as the variance of the disorder is increased. It is also clearly seen that as Λ increases (decreasing width) the effect of the disorder vanishes such that the very narrow excitations, in the average, are unaffected by the disorder. It is important to stress that this is an average effect, because for each realization of the disorder the narrow excitation will be affected. The narrow excitation will experience a shift in the nonlinear frequency equal to the amplitude of the disorder at the position of the excitation.

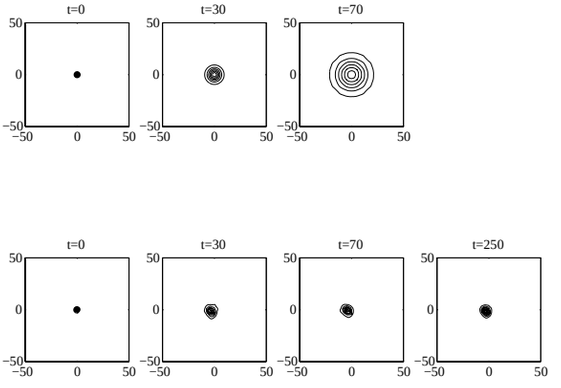

FIGURE 8. Evolution of an initial excitation of the norm $N = 10.4402$ without disorder (upper part) and with disorder strength $\eta = 0.1$ (lower part) [10].

The bistability we observe in Fig. 7 occurs due to the competition between two different length scales of the system: one length scale being defined by the relation between the nonlinearity and the dispersion, while the other length scale being defined by the disorder. A similar effect was observed in ref. [4] for the one-dimensional discrete NLS equation with a quintic nonlinearity. The latter is quite natural because as it is well known (see e.g. [2]) the properties of the two-dimensional NLS model with a cubic nonlinearity are similar to the properties the one-dimensional NLS equation with a quintic nonlinearity.

Having studied the stationary problem it is vital to compare the results to full dynamical simulations. Therefore we carry out a numerical experiment launching a pulse in a system governed by Eq. (15.15). Specifically, stationary solutions (15.16) of Eq. (15.15) with $\Lambda = 0.14$ (after reducing the amplitude of these solutions by 5%) were used as initial conditions of the dynamical simulations. Examples of this experiment are shown in Fig. 8. We see that the pulse behavior in the absence of disorder differs drastically from the pulse behavior in the presence of disorder. (In the latter case we present a realization corresponding to the disorder variance $\eta = 0.1$). While the pulse rapidly disperses in the ideal system (the contour plot for $t = 250$ is absent because the pulse width is of the system size), the process is arrested in the disordered system. After some transient behavior the excitation stabilizes and attains an approximately stationary width. The dynamical simulations thus support the conclusion that otherwise unstable excitations are stabilized by the presence of disorder in the low frequency limit.

An analytical theory of soliton states in disordered NLS models based on the collective coordinate approach and on the Rice's theorem from the theory of random processes [17] is presented in Refs. [4] and [10].

5 Summary

In summary we have shown that the presence of competing length scales leads to multistability phenomena in nonlinear Schrödinger models. We have analyzed three types of the NLS models.

The nonlinear Schrödinger-Kronig-Penney model presents an example where two competing length scales exist: the width ζ of the soliton in the nonlinear Schrödinger equation and the interlayer spacing ℓ. Due to the interplay between these two length scales the localized stationary states exist only in a finite interval of the excitation power. Two branches of stationary states exist but only the low-frequency branch is stable.

In discrete nonlinear Schrödinger models with long-range dispersive interactions, there exist three types of length scales: the soliton width, the lattice spacing and the radius of the dispersive interaction. Here the com-

petition of the length scales provides the existence of three branches of stationary states. Two of them: the low-frequency branch with continuum-like excitations and the high-frequency branch with intrinsically localized excitations, are stable. It is shown that a controlled switching between narrow, pinned states and broad, mobile states is possible. The particular choice of perturbation is not important for the qualitative features of the switching, as long as there is a substantial overlap between the perturbation and the internal breathing mode. The switching phenomenon could be important for controlling energy storage and transport in DNA molecules.

Considering nonlinear excitations in two-dimensional discrete nonlinear Schrödinger models with disorder, it was found that otherwise unstable continuum-like excitations can be stabilized by the presence of the disorder. For the very narrow excitations the disorder has no effect on the averaged behavior. The bistability that was observed in this case is very similar to the bistability that occurs in nonlocal NLS models. Here the bistability arises on similar grounds because of competition between the solitonic length scale and the length scale defined by the disorder.

Acknowledgments: Yu.B.G. thanks MIDIT, and the Department of Mathematical Modelling, the Technical University of Denmark for the hospitality. Yu.B.G. and S.F.M. acknowledge support from the Ukrainian Fundamental Research Fund under grant 2.4/355.

6 References

[1] A. R. Bishop, D. Cai, N. Grønbech-Jensen, and M. I. Salkola. In A. R. Bishop, S. Jiménez, and L. Vázquez, editors, *Fluctuation phenomena: disorder and nonlinearity*, page 316, Singapore, 1995. World Scientific.

[2] O. Bang, J. J. Rasmussen, and P. L. Christiansen. *Nonlinearity*, **7**: 205, 1994.

[3] P. L. Christiansen, J. C. Eilbeck, and R. D. Parmentier, editors. *Future Directions of Nonlinear Dynamics in Physical and Biological Systems.* Plenum Press, New York, 1993.

[4] P. L. Christiansen, Yu. B. Gaididei, M. Johansson, K. Ø. Rasmussen, D. Usero, and L. Vázquez. *Phys. Rev. B*, **56**: 14407, 1997.

[5] P. L. Christiansen, Yu. B. Gaididei, K. Ø. Rasmussen, V. K. Mezentsev, and J. J. Rasmussen. *Phys. Rev. B*, **54**: 900, 1996.

[6] A. S. Davydov. *Theory of Molecular Excitons.* Plenum, New York, 1971.

[7] J. C. Eilbeck, P. S. Lomdahl, and A. C. Scott. *Physica D*, **16**: 318, 1985.

[8] S. Georghiou, T. D. Bradrick, A. Philippetis, and J. M. Beechem. *Biophysical J.*, **70**: 1909, 1996.

[9] Yu. B. Gaididei, P. L. Christiansen, K. Ø. Rasmussen, and M. Johansson. *Phys. Rev. B*, **55**: 13365R, 1997.

[10] Yu. B. Gaididei, D. Hendriksen, P. L. Christiansen, and K. Ø. Rasmussen. *Phys. Rev. B*, **58**: 3075, 1998.

[11] Yu. B. Gaididei, S. F. Mingaleev, P. L. Christiansen, and K. Ø. Rasmussen. *Phys. Rev. E*, **55**: 6141, 1997.

[12] H. Hasegawa and Y. Kodama. *Solitons in Optical Communications*. Claredon Press, Oxford, 1995.

[13] V. A. Hopkins, J. Keat, G. D. Meegan, T. Zhang, and J. D. Maynard. *Phys. Rev. Lett.*, **76**: 1102, 1996.

[14] M. Johansson, Yu. Gaididei, P. L. Christiansen, and K. Ø. Rasmussen. *Phys. Rev. E*, **57**: 4739, 1998.

[15] Yu. S. Kivshar, S. A. Gredeskul, A. Sánchez, and L. Vázquez. *Phys. Rev. Lett.*, **64**: 1693, 1990.

[16] T. Kubota, D. R. Ko, and D. Dobbs. *J. Hydronaut.*, **12**: 157, 1978.

[17] O. Krée and C. Soize. *Mathematics of random phenomena*. Reidel, Dordrecht, 1986.

[18] E. W. Laedke, O. Kluth, and K. H. Spatschek. *Phys. Rev. E*, **54**: 4299, 1996.

[19] L. D. Landau and E. M. Lifshitz. *Statistical Physics*. Pergamon Press, London, 1959.

[20] E. W. Laedke, K. H. Spatschek, V. K. Mezentsev, S. L. Musher, I. V. Ryzhenkova, and S. K. Turitsyn. *Pis'ma Zh. Eksp. Teor. Fiz.*, **62**: 652, 1995. [*JETP Lett.*, **62**: 677, 1995].

[21] E. W. Laedke, K. H. Spatschek, and S. K. Turitsyn. *Phys. Rev. Lett.*, **73**: 1055, 1994.

[22] V. K. Mezentsev, S. L. Musher, I. V. Ryzhenkova, and S. K. Turitsyn. *Pis'ma Zh. Eksp. Teor. Fiz.*, **60**: 815, 1994. [*JETP Lett.*, **60**: 829, 1994].

[23] A. C. Newell and J. V. Moloney. *Nonlinear Optics*. Addison-Wiley, Amsterdam, 1992.

[24] J. J. Rasmussen and K. Rypdal. *Phys. Scr.*, **33**: 481, 1986.

[25] K. Rypdal and J. J. Rasmussen. *Phys. Scr.*, **33**: 498, 1986.

[26] R. Scharf and A. R. Bishop. *Phys. Rev. E*, **47**: 1375, 1993.

[27] R. Scharf. *Chaos, Solitons and Fractals*, **5**: 2527, 1995.

[28] A. C. Scott. *Phys. Rep.*, **217**: 1, 1992.

[29] K. H. Spatschek and F. G. Mertens, editors. *Nonlinear Coherent Structures in Physics and Biology*. Plenum Press, New York, 1994.

[30] C. M. Soukoulis, editor. *Photonic Band Gaps and Localization*. Plenum Press, New York, 1993.

[31] C. M. Soukoulis, editor. *Photonic Band Gap Materials*. Kluwer Academic Publishers, Dordrecht / Boston / London, 1996.

[32] L. Vázquez, L. Streit, and V. M. Pérez-Garcia, editors. *Nonlinear Klein-Gordon and Schrödinger Systems: Theory and Applications*. World Scientific, Singapore, 1996.

16

Demonstration Systems for Kink-Solitons

M. Remoissenet

ABSTRACT We consider a mechanical lattice where the basic oscillating units experience a double-well on-site potential, and are linearly and nonlinearly coupled. In the continuum limit the lattice equations can be approximated by a nonlinear partial differential equation. With nonlinear coupling only, this equation exhibits a static kink solution with a compact shape or compacton. When both linear and nonlinear coupling exist, one can obtain a dynamic compacton solution propagating at the characteristic velocity of linear waves. Contrary to the static compacton solution, which is stable, the dynamic compacton is unstable: it looses progressively its compact shape as it propagates, and evolves into a kink waveform. A nice feature is that mechanical analogs of this lattice can be constructed that allow one to observe kinks. We constructed two mechanical lattices: one with torsion and gravity pendulums and another one with flexion and gravity oscillating units. Both experimental systems allow one to illustrate and study qualitatively the dynamical properties of the propagating kinks. In the strong nonlinear coupling limit static compactons can be created. In the weak coupling limit, lattice effects and the observation of discrete kinks are briefly considered.

1 Introduction

Localized waves with permanent profile, known as solitons, are widely accepted as a structural basis for viewing and understanding the dynamical behavior of complex nonlinear systems, and they play a significant role in different physical problems. Although historically solitary waves were discovered experimentally the number of experimental investigations on solitonic waves is rather small compared to the considerable body of literature devoted to the theoretical and numerical studies. Thus, it is necessary to perform experiments on complex systems in the real world, but it is also important to develop and construct simple experimental models [15] or laboratory experiments that allow us to create, observe and manipulate localized waves. For example, electrical nonlinear transmission lines are simple and versatile devices that enable to observe and study quan-

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 323−337, 2000.

titatively the propagation and properties of solitons and nonlinear lattice modes, in real continuous and discrete systems. Water-tank experiments, in shallow or deep water, allow one to study and illustrate the important features of pulse or envelope solitons, and also kink-solitons [3]. Analog mechanical systems, such as the mechanical transmission line first introduced by Scott [17], play an important role in the study of kink-solitons and their remarkable properties. Contrary to non-topological pulse solitons and envelope solitons, where the system returns to its state after the passage of the wave, kink-solitons belong to a particular class of nonlinear excitations called topological solitons because in some cases the structure of the system is modified after the passage of the wave. Kink-solitons have been used to describe various phenomena in one-dimensional systems, such as ferromagnetic [1] or ferroelectric domain walls [9], dislocations [7], dynamics of base pairs in DNA molecules [11], polymer chain twistings [4] and Josephson junctions [14]. The basic models are generalized Klein-Gordon (KG) models where the particles may be considered as coupled to nearest neighbors only, via an interaction potential $U(\theta_{n+1} - \theta_n)$ and subjected to a nonlinear on-site or substrate potential $V(\theta_n)$, where $\theta_n(t)$ is the on-site degree of freedom, which represents the influence of the surrounding lattice atoms and external effects. The lattice Hamiltonian is

$$H = \sum_n \left[\frac{1}{2} \left(\frac{d\theta_n}{dt} \right)^2 + U(\theta_{n+1} - \theta_n) + V(\theta_n) \right]. \tag{16.1}$$

The corresponding equations of motion can be written in the standard form

$$\frac{d^2\theta_n}{dt^2} - [U'(\theta_{n+1} - \theta_n) - U'(\theta_n - \theta_{n-1}) + V'(\theta_n)] = 0. \tag{16.2}$$

Depending on the shape of the on-site potential a nonlinear lattice with the Hamiltonian (16.1) may sustain different kinds of nonlinear excitations. If $V(\theta_n)$ has two degenerate minima (double-well shape like in the Φ -four model) or multiple degenerate minima (periodic shape like in the Frenkel-Kontorova model) topological kink excitations, which connect two equivalent ground states, can exist. If interparticle interactions U' are linear the kink solutions can be calculated exactly, in the continuum limit [9]. For the discrete equations (16.2) the kink solutions can be obtained either by perturbation approaches [5] or by numerical techniques. If U' also include anharmonic interactions specific kink internal modes may be created [18]. The case where U' is nonlinear only, is interesting because of the presence of nonlinear dispersion. Recently it was shown by Rosenau and Hyman [16] that solitary-wave solutions may compactify under the influence of nonlinear dispersion which is capable of causing deep qualitative changes in the nature of genuinely nonlinear phenomena. Such robust soliton-like solutions, characterized by the absence of the infinite tail, have been called

compactons [16]. They have been obtained for a special class of the Korteweg de Vries (KdV) type equations with nonlinear dispersion [10].
Part of the motivation of this work finds its origin in the possibility of observing kinks in "real systems" with a double-well potential. In this regard, in Section 2 we show that compacton-like kinks, or compactons for short, can exist for specific velocities in physical systems modeled by a nonlinear Klein-Gordon equation with strong anharmonic coupling. Then, in Section 3 we present analog experiments: a nice feature is that mechanical chains can be constructed allowing to observe kinks and compactons. The mechanical chains are also useful to oberve discrete kinks as briefly considered in section 4. Section 5 is devoted to concluding remarks.

2 Mechanical chains with double-well potential

2.1 Chain with torsion and gravity

In order to observe kinks we have constructed [2] a mechanical analog which consists of a new experimental chain of identical pendulums. Each basic unit is similar to the pendulum recently studied by Peters [12]: it can oscillate with a motion whose character is determined by the forces of torsion and gravity acting in opposition. Depending on the length d (control parameter) of each pendulum, the on-site potential V can present one or two minima. For the configuration presently considered it possesses two equilibrium positions (two wells). Each pendulum is connected to its neighbors by springs, as sketched in Fig. 1. When the dissipation is neglected and the difference between angular displacement of neighboring pendulums are small enough, the equation of motion of the nth chain unit is given (see Appendix) by

$$I\frac{d^2\theta_n}{dt^2} = -K\theta_n + Mgd\sin\theta_n + C_{0,l}(\theta_{n+1} + \theta_{n-1} - 2\theta_n)$$
$$-C_{0,nl}(\theta_n - \theta_{n+1})^3 - C_{0,nl}(\theta_n - \theta_{n-1})^3, \qquad (16.3)$$

where the terms on right hand side represent the restoring torque owing to the torsion, the gravitational torque and the restoring torque owing to the coupling with the neighboring pendulums (see Appendix). $\theta_n(t)$ is the angular displacement as a function of time t of the nth pendulum, $I = Md^2$ is the moment of inertia of a single pendulum of mass M and length d, g is the gravitation, K is the torsion constant: to a first approximation we assume that K is independent of θ_n. $C_{0,l}$ and $C_{0,nl}$ are the linear and nonlinear torque constant of a spring between two pendulums: they are given by

$$C_{0,l} = kd^2(1 - \frac{l_0}{l_1}), \qquad (16.4)$$

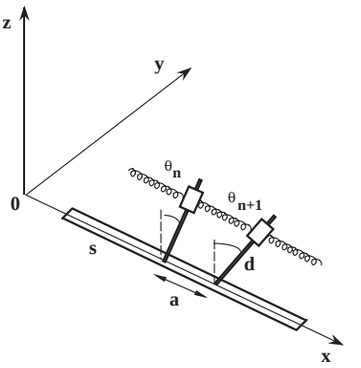

FIGURE 1. Sketch of the mechanical chain. Here, only two pendulums, n and $n+1$, coupled to each other by spring and attached to the steel ribbon s (parallel to x axis), are represented with their respective angular displacements θ_n and θ_{n+1}. The pendulums are at equilibrium in one of the two equivalent potential wells. The motion can occur in a plane perpendicular to the chain (x axis). See also Fig. 5.

$$C_{0,nl} = kd^2\left(\frac{l_0 d^2}{2l_1^3} - \frac{C_{0,l}}{6}\right), \tag{16.5}$$

where k is the spring stiffness, l_0 the natural length of a spring at rest and l_1 the length of this spring when it is stretched between two adjacent pendulums at equilibrium (bottom of one well). Note that the nonlinear coupling term must be fully taken into account because the linear term is especially small when l_1 is not very different from l_0. Moreover, when $l_0 = l_1$ we have $C_{0,l} = 0$ and $C_{0,nl} = kd^4/2Kl_1^2$, as we will see in the following. Setting

$$\tau = \frac{K}{I},\ \Gamma = \frac{Mgd}{K},\ C_{1,l} = \frac{C_{0,l}}{K},\ C_{nl} = \frac{C_{0,nl}}{K}, \tag{16.6}$$

we transform Eq. (16.3) into

$$\frac{d^2\theta_n}{d\tau^2} = -\theta_n + \Gamma\sin\theta_n + C_{1,l}(\theta_{n+1} + \theta_{n-1} - 2\theta_n) \\ -C_{nl}[(\theta_n - \theta_{n+1})^3 + (\theta_n - \theta_{n-1})^3]. \tag{16.7}$$

In Eq. (16.7), the quantity $(-\theta_n + \Gamma\sin\theta_n)$ represents the "on-site" (zero coupling limit) torque. In the continuum approximation one obtains

$$\frac{\partial^2\theta_n}{\partial\tau^2} - \left[C_{1,l} + 3C_{nl}\left(\frac{\partial\theta}{\partial X}\right)^2\right]\ \frac{\partial^2\theta}{\partial X^2} + \theta - \Gamma\sin\theta = 0 \tag{16.8}$$

with $X = x/a$. The interaction and on-site potential corresponding to Eq. (16.7) are

$$U(\theta_{n+1} - \theta_n) = \frac{C_{1,l}}{2}(\theta_{n+1} - \theta_n)^2 + \frac{C_{nl}}{4}(\theta_{n+1} - \theta_n)^4, \tag{16.9}$$

and

$$V(\theta_n) = \frac{1}{2}(\theta_n^2 - \theta_m^2) + \Gamma(\cos\theta_n - \cos\theta_n). \tag{16.10}$$

Here, the parameter Γ plays the role of a control parameter. For $\Gamma > 1$ it determines the depth and separation of the two wells [12] and θ_m correspond to the two equilibrium positions. Equation (16.7) and its continuum approximation (16.8) cannot be solved analytically. Nevertheless, in order to get some approximate solution one can replace the potential $V(\theta_n)$ by the standard Φ -four potential given by

$$V(\Theta_n) = \frac{V_0}{2}(1 - \Theta_n^2)^2, \tag{16.11}$$

with $\Theta_n = \theta_n/\theta_m$ and $V_0 = -\theta_m^2 + 2\Gamma(1 - \cos\theta_m)$. As depicted in Fig. 2 the

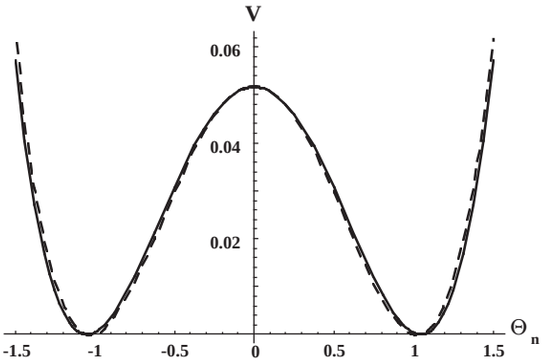

FIGURE 2. Fitting of the double well potential (16.10): dotted line, by a Φ-four potential of form (16.11): continuous line. For $-1.5 < \Theta_n = \theta_n/\theta_m < 1.5$ the two curves are practically superimposed and the approximation of potential (16.10) of the real system by Φ-four potential is reasonable.

fitting is good for $-1.5 < \theta_n/\theta_m < 1.5$. Under these conditions eq (16.7) reduces to

$$\frac{d^2\Theta_n}{dT^2} - C_{1,l}(\Theta_{n+1} + \Theta_{n-1} - 2\Theta_n)$$
$$+C_{nl}[(\Theta_n - \Theta_{n+1})^3 + (\Theta_n - \Theta_{n-1})^3] - 2V_0'(\Theta_n - \Theta_n^3) = 0, \tag{16.12}$$

where $\Theta_n = \theta_n/\theta_m$, $T = \theta_m\tau$, $C_l = C_{1,l}/\theta_m^2$ and $V_0' = 2V_0/\theta_m^4$. In the continuum limit, Eq. (16.12) is approximated by

$$\frac{\partial^2\Theta}{\partial T^2} - \left[C_l + 3C_{nl}(\frac{\partial\Theta}{\partial X})^2\right]\frac{\partial^2\Theta}{\partial X^2} - 2V_0'(\Theta - \Theta^3) = 0, \tag{16.13}$$

where we have assumed that $3C_{nl}(\partial^2\Theta/\partial X^2) \gg (C_l/12)(\partial^4\Theta/\partial X^4)$, as it will be the case in the following. Note that C_l represents the square of the velocity of linear waves in the chain. For $C_{nl} = 0$, Eq. (16.13) reduces to the standard continuous Φ-four model with linear coupling, which admits tanh-shaped kink solutions. We then look for localized waves of permanent profile of the form $\Theta(s) = \Theta(X - uT)$, such as $\Theta \to \pm 1$ and $d\Theta/ds \to 0$, when $s \to \pm\infty$, where s is a single independent variable depending on u which is an arbitrary velocity of propagation. Integrating (16.13) and taking account of these conditions we get

$$2(u^2 - C_l)\Theta_s^2 - 3C_{nl}\Theta_s^4 + 2V_0'(1 - \Theta^2)^2 = 0. \tag{16.14}$$

For $u^2 - C_l \neq 0$ this equation cannot be integrated analytically, but numerically, only. One gets a kink solution which connects the two equivalent ground states (potential minima). For $u^2 - C_l = 0$ that is, for the two particular cases: $C_l = 0$ (zero linear coupling: linear waves cannot exist) and u=0, and $u = \pm\sqrt{C_l}$ which correspond to kinks with a compact support or compactons (see Appendix), Eq. (16.14) can be integrated. One obtains

$$\Theta_c(X) = \pm\sin[(2V_0'/3C_{nl})^{1/4}(X - X_0)], \tag{16.15}$$

when $\mid X - X_0 \mid < (\pi/2)(2V_0'/3C_{nl})^{1/4}$, and $\Theta = \pm 1$ otherwise. As usual, the constant of integration (X_0) defines the position of the center of the compacton. For the second case we have

$$\Theta(X,T) = \pm\sin[(2V_0'/3C_{nl})^{1/4}(s - s_0)], \tag{16.16}$$

when $\mid X - \sqrt{C_l}t - s_0 \mid = \mid s - s_0 \mid < (\pi/2)(2V_0'/3C_{nl})^{1/4}$ and $\Theta = \pm 1$ otherwise.

From (16.15) and (16.16) we can calculate the full width of the compactons which in both cases is equal to: $L = \pi(3C_{nl}/2V_0')^{1/4}$. Consequently, when there is no linear coupling one has a static compacton (anti-compacton) solution and when both linear and nonlinear coupling are present, a dynamic compacton (anti-compacton) solution traveling at particular velocity $\sqrt{C_l}$ (or $-\sqrt{C_l}$) may exist. The shape of the dynamic compacton is identical to the shape of the static one, it is represented in Fig. 3 for $s_0 = 0$. From the continuum approximation of (16.1), using (16.11) and (16.9), one can calculate the total (kinetic plus potential) energy E_{tot} localized in the compacton traveling at velocity $-\sqrt{C_l}$. One has

$$E_{tot} = \int_{-\pi/2\gamma}^{\pi/2\gamma}\left[\frac{1}{2}(\frac{\partial\Theta}{\partial T})^2 + \frac{1}{2}C_l(\frac{\partial\Theta}{\partial X})^2 + \frac{1}{4}C_{nl}(\frac{\partial\Theta}{\partial X})^4 + \frac{1}{2}(V_0(1 - \Theta^2)^2\right]dX. \tag{16.17}$$

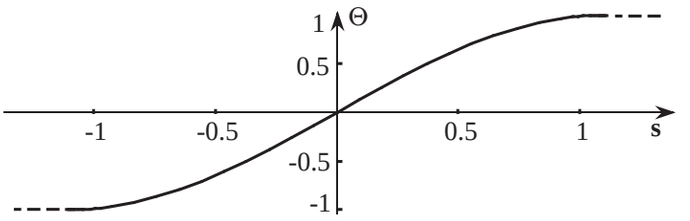

FIGURE 3. Compacton waveform as given by (16.16), with $s_0 = 0$ (compacton center) and $2V_0'/3C_{nl} = 4$.The S-shaped waveform (continuous line) corresponds to the compact part which connects the two constant parts ($\Theta = \pm 1$: horizontal dotted lines) of the solution.

where Θ is given by (16.16), and we have set $\gamma = (2V_0/3C_{nl})^{1/4}$. After simple calculations we obtain

$$E_{tot} = a\gamma \frac{\pi}{2} \left(C_l + \frac{V_0'}{2} \right). \tag{16.18}$$

In the case $u = \sqrt{C_l} = 0$, one has

$$E_{tot}^{stat} = a\gamma V_0' \frac{\pi}{4}, \tag{16.19}$$

which represents the "mass" of a static compacton. Note that the energy is strictly localized and contrary to a standard (tanh-shaped) kink which possesses (exponential) wings and can interact with an antikink, a compacton and an anticompacton will not interact unless they come into contact in a way similar to the contact between two hard spheres. Such a result should be interesting for the modeling of static domain walls in condensed matter physics.

2.2 Chain with flexion and gravity

Another mechanical chain can be constructed, where each basic oscillating units is an Euler strut [13]. An Euler strut consists [8] of a light elastic (metal) strip with its bottom edge attached to a support and two masses on opposite sides of the strip (to maintain symmetry) at a distance d (control parameter) from the bottom. Like for the previous chain we consider the configuration with two equilibrium positions and each unit is coupled to its nearest neighbors by springs. To a first approximation we assume that distance d (and consequently the moment of inertia $I = Md^2$) does not depend on θ. Second, we assume that the torque generated by the strip about the bottom axis, when it is displaced by an angle θ, is simply proportional to this angle: the restoring torque due to flexion is $-K\theta$. It

acts in opposition with the torque due to gravity. Under these assumptions, neglecting dissipation, we obtain an equation of motion similar to (16.3) which can be reduced to (16.13) if the coupling between oscillators is large enough compared to the potential barrier height.

2.3 Numerical simulations

In order to check the validity of our analytical results and stability of our solutions, we have performed numerical simulations [2] of the equation of motion (16.3)(or Eq. (16.12)) which, in the strong coupling (continuous) limit, reduces to (16.8). It has been integrated with a fourth-order Runge-Kutta scheme with a time step chosen to preserve the total energy of the system to an accuracy better than 10^{-5} over a complete run.
For $C_l = 0$, we have first verified the validity of the static compacton solution Θ_c given by Eq. (16.15). If this solution (with parameters: $V_0 = 2$, $L = 32a$ and $C_{nl} = 14500$) is chosen as an initial condition of the system and let to evolve in the presence of a weak additional dissipation, it relaxes to: $\Theta = \Theta_c + \Delta\Theta$, where $\Delta\Theta$ are weak spatial sinusoidal deviations from the exact solution with amplitude 4.10^{-4}. Moreover, an arbitrary tanh-shaped initial kink also relaxes toward the same profile Θ proving that Θ_c is a good solution to order 10^{-4}. This deviation $\Delta\Theta$ from the exact solution can be reduced if we choose a compacton with larger width L, which indicates that the closer to the continuum limit we are, the better the solution is. Actually, this result points out that the static compacton is an exact solution of the continuous system.

For $C_l \neq 0$, a dynamic compacton (with parameters: $C_l = 208$, $V_0 = 2$, $L = 32a$ and $C_{nl} = 14500$) launched at initial velocity $\sqrt{C_l}$ emits small radiations. Consequently its velocity decreases and we no more have a dynamic compacton as described by solution (16.16), but rather a kink (whose waveform is unknown analytically). It turns out that as soon as they are launched and propagate, dynamic compactons, as described by solution (16.16), loose their compact shape, they cannot exist. Then, with the same parameters as above, we have analyzed the head-on collision of a compacton (initial velocity $v_{1c} = \pm\sqrt{C_l}$) and an anticompacton (initial velocity $v_{2c} = -v_{1c}$). Our results [2] show that the collision is inelastic: the two kinks that emerge from the collision are deformed, they radiate oscillations and propagate at velocities lower than v_1. The collision between a dynamic compacton (initial velocity $v_{1c} = \pm\sqrt{C_l}$) and a kink (initial velocity $v_{1k} = -0.3\sqrt{C_l}$), was also investigated. Again, two deformed kinks emerge from the collision. Thus the moving compacton solution is unstable.

3 Experiments

3.1 Chain with torsion and gravity

The apparatus is a lattice of 20 pendulums attached vertically to the center of a horizontal steel ribbon (2 m long, 6 mm wide and 0.1 mm thick) supported by vertical metallic plates which are equidistant ($a = 10$ cm) (see Figs. 1 and 4). A basic pendulum consists of a thin rod (diameter 3mm) along which a cylinder (mass $M = 67$ g) can be displaced and fixed. Depending on the vertical position d of the mass along the rod, the system can oscillate with a motion which depends on the potential shape and is determined, as mentioned earlier, by the forces of gravity and torsion in opposition. Here, with $d = 87$ mm and $K = 0.03$, the control parameter is $\Gamma = 1.9$, thus the on-site potential is a symmetric double well potential. Once its tension is adjusted, the ribbon is held tight on the top of each plate. With this precaution, the torsion constant is the same for each pendulum and the weak residual torsional coupling between pendulums can be neglected. Each pendulum (cylinder) is attached to its neighbor with a spring. Springs connecting pendulums that are at equilibrium, in one of the bottom of a potential well, are horizontal (see Fig. 4). The stiffness of the springs is large enough to ensure the validity of the continuum approximation. With the physical parameters: $C_l = 0$ ($l_1 = l_0 = 68$ mm, see (16.4)), $C_{nl} = 25$ and $k = 120$, a static compacton can be observed as represented in Fig. 4. The experimental shape fits approximatively (see Fig. 5) the theoretical shape calculated from (16.15).

When $C_l \neq 0$ and $C_{nl} \neq 0$, solution (16.16) predicts a compacton moving at velocity $\sqrt{C_l}$ of the linear waves. In this case, we cannot conclude that the moving S-shaped entity we observe has a compact shape for the following reasons. First, we cannot control with sufficient precision the initial velocity of the kink. Second, even if we could launch a kink with exact velocity $\sqrt{C_l}$ it would gradually slow down owing to dissipative effects (that are important compared to small radiation effects predicted by our numerical simulations, see 2.3) ; thus, we can never observe a moving compacton. Nevertheless, with this mechanical line we can easily observe the dynamical properties of the kinks. For example, if one launches a moving kink (unknown analytical shape) with arbitrary initial velocity at one end of the line, after reflection at the opposite free end this kink becomes an anti-kink moving freely in the opposite direction and so on. Depending on its initial velocity a kink can reflect three or four times before gradually slowing down owing to dissipative effects which inevitably occur for a real mechanical line. With the above physical parameters no radiation of waves due to discreteness effects are observed. Thus, the continuum approximation is valid. Nevertheless, lattice effects and also the pinning of kinks can be observed by simply decreasing the stiffness of the springs, such experiments will be briefly discussed in section 4.

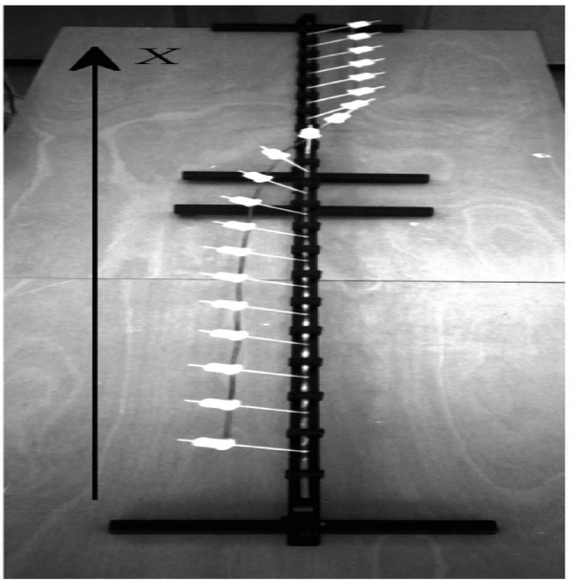

FIGURE 4. Picture showing a static compacton as observed on the experimental chain, with torsion and gravity, lying horizontally on a table.The X axis is directed along the lengthy black support. The compacton connects pendulums directed to the left of x axis, at equilibrium in one potential well($\Theta = -1$), to the pendulums directed to the right of x axis, at equilibrium in the other potential well ($\Theta = +1$).

3.2 Chain with flexion and gravity

In this subsection we describe briefly the chain and qualitative observations. The chain is constructed with 40 Euler struts attached vertically to a horizontal support. The lattice spacing is a=4 cm. A basic unit [8] consists of a flexible metal strip (width =10 mm, thickness=0.1mm and d=97 mm for $\theta_m = 50$) along which, two rectangular pieces (total mass, M=18.5 g) of metal are fixed. In the present case, the on-site potential determined by the distance of the masses to the support is a double well potential. Each unit is connected to its neighbors by springs.

As for the chain with torsion and gravity, moving and static kinks can be created. For example, if the condition $C_l = 0(l_0 = l_1)$ is satisfied, one can observe a compacton (with spatial extension 12a) and an anti-compacton as represented in Fig. 6. In this chain the moving kinks slow down rapidly because the dissipation effects are more important than for the chain considered in section 3.1. It means that the dynamics of the system will be correctly described by adding a dissipative term to (16.3) and consequently

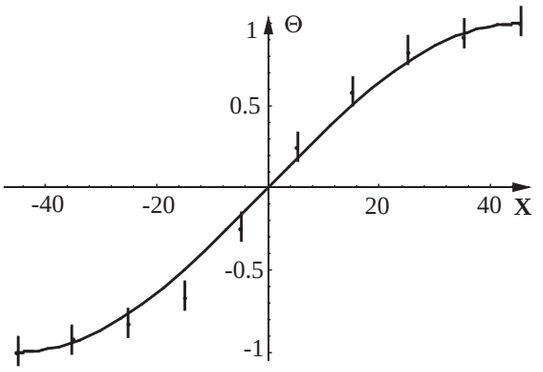

FIGURE 5. Comparison of the experimental compacton shape, observed in Fig. 4, to the theoretical shape calculated from (16.15). Here, $\Theta = \theta/\theta_m = \pi/3$ and $X = x/a$ (with a=10cm is dimensionless). The vertical lines represent the experimental precision.

to (16.13).

4 Lattice effects

In section 3.1 we have considered a mechanical chain whose dynamics can be described by (16.13) in the continuum approximation which is valid when the coupling is strong (stiff springs). Nevertheless, lattice effects, which play an important role in microscopic systems, may induce the pinning of an initially moving kink (whose width is a few lattice spacings) [15]. This effect, which we briefly discuss here, can be observed by simply decreasing the stiffness of the springs for the chain with torsion and gravity. If the pinning is important enough the kinks cannot propagate. For example, in Fig. 7 one can observe (chain with torsion and gravity) an arbitrary sequence of highly discrete kinks and antikinks, whose width is one lattice spacing, that are completely pinned on the lattice and cannot propagate. With such a simple mechanical chain, whose springs can be modified very rapidly, one can do more. For example, one can observe nonlinear localized modes (see in this book the chapter by Marin, Eilbeck, and Russell and also the recent review [6]), such experiments will be discussed elsewhere.

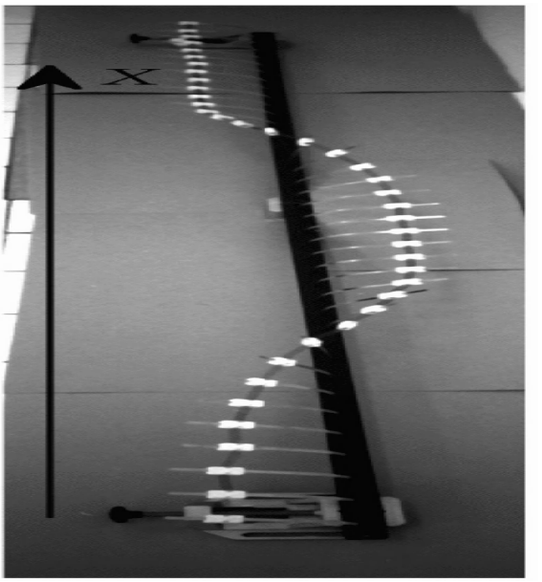

FIGURE 6. Photograph showing (with X axis directed along the black support) a static compacton and anticompacton as observed on the experimental chain, with flexion and gravity.

5 Conclusion

We have considered a chain of identical pendulums that experience a double-well on-site potential and are nonlinearly coupled. In the continuum limit this chain can be modeled by a nonlinear partial differential equation. With nonlinear coupling only, this equation exhibits a static kink solution with a compact shape or compacton. When both linear and nonlinear coupling exist, one can obtain a dynamic compacton solution propagating at the characteristic velocity of linear waves. Our numerical solutions point out that, contrary to the static compacton solution, which is stable, the dynamic compacton is unstable: it looses progressively its compact shape as it propagates, and evolves into a kink waveform which is unknown analytically.

In order to observe kinks and compactons we constructed two mechanical chains. One chain with torsion and gravity pendulums and one chain with flexion and gravity oscillating units. Both experimental systems allow us to illustrate and study qualitatively the dynamical properties of the kinks that can move back and forth. In the strong nonlinear coupling limit static compactons can be created. Such strictly localized entities, that do not in-

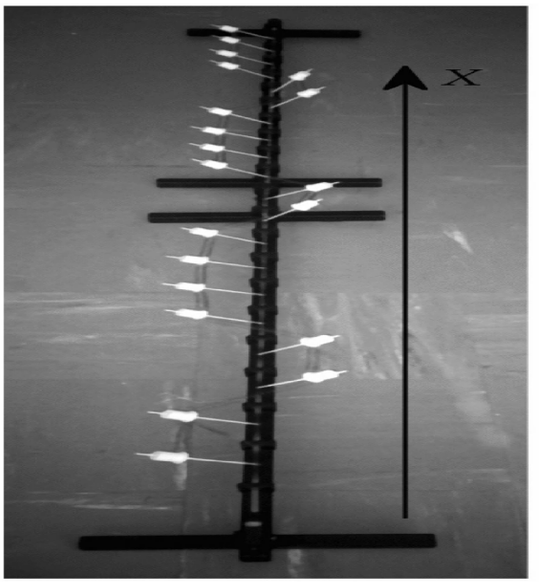

FIGURE 7. Photograph showing (with X axis directed along the black support) highly discrete (respective widths: one lattice spacing) kinks and antikinks, when the coupling between pendulums is weak (soft springs) compared to the potential barrier. They cannot propagate, as observed on the experimental chain with torsion and gravity.

teract unless they come into contact, should play a role in the modeling of domain walls in real physical systems. In the weak coupling limit, highly discrete kinks, pinned on the lattice, can be oberved.

Acknowledgments: Thanks go to S. Dusuel who performed some of this research and P. Michaux who constructed the experimental systems. I also acknowledge G. Pierre, V. Boudon and R. Chaux for technical assistance and support.

6 Appendix

In this appendix we derive Eq. (16.3). The general equation of motion of the *nth* pendulum of the chain sketched in Fig. 1 is

$$I\frac{d^2\theta_n}{dt^2} = -K\theta_n + mgd\sin\theta_n + M_{n-1,n} - M_{n,n+1}, \tag{16.20}$$

where $M_{n-1,n}$ and $M_{n,n+1}$are the torque exerted by pendulum n-1 on pendulum n and pendulum n on pendulum $n+1$, respectively. In terms of the components $y_n = -d\sin\theta_n$ and $z_n = d\cos\theta_n$ of the displacement, the elongation of the spring between pendulums n and $n+1$ is

$$\Delta l = \sqrt{{l_1}^2 + (y_{n+1} - y_n)^2 + (z_{n+1} - z_n)^2} - l_0,$$

where l_0 is the length of the spring at rest, and l_1 the minimal length of the stretched spring between two pendulums. Thus we have

$$M_{n,n+1} = \frac{k\Delta l}{V}(y_{n+1}z_n - z_{n+1}y_n).$$

where

$$y_{n+1}z_n - z_{n+1}y_n = d^2\sin(\theta_n - \theta_{n+1})$$

and

$$V = \sqrt{{l_1}^2 + 4d^2\sin^2\left(\frac{\theta_n - \theta_{n+1}}{2}\right)}$$

One gets

$$M_{n,n+1} = kd^2\left[1 - \frac{l_0}{l_1}\frac{1}{\sqrt{1 + \frac{4d^2}{l_1^2}\sin^2\left(\frac{\theta_n - \theta_{n+1}}{2}\right)}}\right]\sin(\theta_n - \theta_{n+1})$$

The torque $M_{n-1,n}$ is obtained by replacing n by $n-1$ in the above expression. When the difference between the angular displacement of neighboring pendulums is small enough (weak discrete limit), the torques can be replaced by their expansion in terms of these angular differences, and we obtain Eq. (16.3.

7 References

[1] A.R. Bishop and T.R. Lewis, *J. Phys.C*, 11:3811, 1979.

[2] S. Dusuel, P. Michaux, and M. Remoissenet, *Phys.Rev.E*, 57:2320, 1998.

[3] B. Denardo, W. Wright, S. Putterman, and A. Larraza, *Phys.Rev.Lett*, 64:1518, 1990.

[4] F. Fillaux and F. Carlile, *Phys.Rev.B*, 42:5990, 1990.

[5] S. Flach and K. Kladko, *Phys.Rev.E*, 54:2912, 1996.

[6] S. Flach and C.R. Willis, *Phys.Rep*, 295:181, 1998.

[7] R. Hobart, *J.Appl.Phys*, 36:1944, 1965.

[8] E.A. Jackson, *Perspectives in Nonlinear Dynamics, Chap 3 Tome 1*, Cambridge University Press, 1991.

[9] J.A. Krumhansl and J.R. Schrieffer, *Phys.Rev.B*, 11:3535, 1975.

[10] P.J. Olver and P. Rosenau, *Phys.Rev.E*, 53:1900, 1996.

[11] M. Peyrard and A.R. Bishop, *Phys.Rev.Lett*, 62:2755, 1989.

[12] R.D. Peters, *Am.J.Phys*, 63:1128, 1995.

[13] A.B. Pippard, *The Physics of Vibrations, Chap 9*, Cambridge University Press, 1979.

[14] R.D. Parmentier and N.F. Pedersen, *Nonlinear Superconducting devices and high Tc materials*, Proceedings of the international Conference, Capry Italy, World Scientic, 1994.

[15] M. Remoissenet, *Waves Called Solitons.Concepts and Experiments*, Springer, New York, 2nd Ed 1996.

[16] P. Rosenau and J.M. Hyman, *Phys.Rev.Lett*, 70:564, 1993.

[17] A.C. Scott, *Am.J.Phys*, 37:52, 1969.

[18] F. Zhang, *Phys.Rev.E*, 54:4325, 1996.

17
Quantum Lattice Solitons

Alwyn Scott

ABSTRACT
A brief introduction to the theory of quantum solitons on a lattice is presented through three examples: (i) local modes of vibration on a small molecule, (ii) a quantum discrete nonlinear Schrödinger soliton on a one-dimensional lattice, and (iii) local modes on a molecular crystal.

1 Introduction

The aim of this chapter is to introduce the theory of quantum lattice solitons by sketching applications to some basic problems arising in the theory of molecular vibrations. Two types of nonlinearity are considered: (i) *intrinsic*, which is caused by a readjustment of the electronic cloud as a molecular bond is stretched, and (ii) *extrinsic*, arising from interactions of a molecular vibration with the nearby structure of a lattice. The first of these is of interest in the context of local modes in small molecules, and the second arises in relation to the formation of local modes on molecular crystals. Further details are available in reference [12].

2 Local modes in the dihalomethanes

Methane is a biologically interesting molecule with the formula CH_4, corresponding to a structure of four hydrogens attached at equal angles to a tetravalent carbon atom. In the *dihalomethanes*—CF_2H_2, CCl_2H_2, CBr_2H_2, and CI_2H_2—two of these hydrogens are replaced by the halides: fluorine, chlorine, bromine, or iodine. These molecules form little laboratories in which interactions between the two CH oscillators can be investigated.

For dichloromethane, the picture is like this:

$$\begin{array}{c} \mathbf{Cl} \\ \mathbf{H}\longleftrightarrow\mathbf{C}\leftrightarrow\mathbf{H} \\ \mathbf{Cl} \end{array}$$

where the two CH stretching oscillations are each anharmonic (nonlinear), and they can interact with each other through both mechanical and electro-

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 339−355, 2000.

magnetic forces. For a sufficiently large value of the (intrinsic) anharmonic parameter, it is possible for the energy to be concentrated on one or the other of the CH bonds as is indicated in the above diagram.

2.1 Classical analysis

In a rotating wave approximation, the classical Hamiltonian (energy) that governs our two interacting CH oscillators is

$$H = \omega_0(|A_1|^2 + |A_2|^2) - \frac{\gamma}{2}(|A_1|^4 + |A_2|^4) - \varepsilon\,(A_1 A_2^* + A_2 A_1^*)\,, \quad (17.1)$$

implying the dynamic equations

$$\begin{aligned} \mathrm{i}\frac{dA_1}{dt} &= \omega_0 A_1 - \gamma|A_1|^2 A_1 - \varepsilon A_2 \\ \mathrm{i}\frac{dA_2}{dt} &= \omega_0 A_2 - \gamma|A_2|^2 A_2 - \varepsilon A_1\,. \end{aligned} \quad (17.2)$$

For notational convenience, it is here assumed that A_1 and A_2 are unitless, while d/dt, ω_0 (site frequency), γ (anharmonic parameter), and ε (dispersive parameter) are all expressed in units of cm^{-1}.

With $\varepsilon = 0$, A_1 and A_2 are complex mode amplitudes of two uncoupled anharmonic oscillators. The aim is to understand the relationship between the classical and quantum behaviors of the system when $\varepsilon \neq 0$.

Under the classical dynamics of Equations (17.2), the energy of Equation (17.1) is conserved. It is easily verified that another conserved quantity is the *norm*

$$N = |A_1|^2 + |A_2|^2\,. \quad (17.3)$$

Assume that the system of Equations (17.2) has a stationary solution of the form $A_1 = \phi_1 e^{-\mathrm{i}\omega t}$ and $A_2 = \phi_2 e^{-\mathrm{i}\omega t}$, where ϕ_1 and ϕ_2 are independent of time. Then for a fixed value of N, three classes of solutions are readily demonstrated [5].

There is a *symmetric mode* for which $\phi_1 = \phi_2 = \sqrt{N/2}$, and an *antisymmetric mode* with $\phi_1 = -\phi_2 = \sqrt{N/2}$. Both of these modes are found in the linear limit $\gamma = 0$. As a result of the nonlinearity, there is also a *local mode*, for which the amplitude on one mode is larger than the other; thus

$$\begin{aligned} \phi_1 &= \left\{\frac{N}{2}\left[1 \pm \left(1 - \frac{4\varepsilon^2}{N^2\gamma^2}\right)^{1/2}\right]\right\}^{1/2} \\ \phi_2 &= \left\{\frac{N}{2}\left[1 \mp \left(1 - \frac{4\varepsilon^2}{N^2\gamma^2}\right)^{1/2}\right]\right\}^{1/2}\,, \end{aligned}$$

and one of the oscillator amplitudes goes to zero as γ becomes much greater than ε.

The local mode requires that

$$\gamma \geq \frac{2\varepsilon}{N}\,, \tag{17.4}$$

and the concentration of vibrational energy on one of the CH bonds at a sufficiently large value of the anharmonic parameter γ is an elementary example of a *lattice soliton.* How is it represented by the linear structure of quantum theory?

2.2 Quantum analysis

To quantize the problem we let the Hamiltonian and norm, defined in Equations (17.1) and (17.3), become operators

$$N \to \hat{N} \quad \text{and} \quad H \to \hat{H}\,,$$

where

$$\hat{N} = b_1^\dagger b_1 + b_2^\dagger b_2$$

and

$$\hat{H} = (\omega_0 - \gamma/2)\hat{N} - \frac{\gamma}{2}[(b_1^\dagger b_1)^2 + (b_2^\dagger b_2)^2] - \varepsilon(b_1^\dagger b_2 + b_2^\dagger b_1)\,.$$

In constructing these operators, we have set $A_1 \to b_1$, $A_2 \to b_2$, $A_1^* \to b_1^\dagger$ and $A_2^* \to b_2^\dagger$, so b_1 and b_2 are bosonic lowering operators for the first and second oscillators, respectively. Similarly, $b_1^\dagger$ and $b_2^\dagger$ are raising operators for the two oscillators. In these expressions for $\hat{N}$ and $\hat{H}$, the constant terms have been dropped; thus they measure number and energy with respect to the lowest (ground state) eigenvalues.

If the two oscillators are uncoupled ($\varepsilon = 0$), Dirac's expression for an eigenfunction is $|n_1\rangle|n_2\rangle$, where n_1 and n_2 are the bosonic number levels of the two oscillators. For convenience, this ket product is written as

$$|n_1\rangle|n_2\rangle \equiv |n_1, n_2\rangle\,,$$

whereupon

$$\begin{aligned}
b_1\,|n_1, n_2\rangle &= \sqrt{n_1}\,|n_1 - 1, n_2\rangle \\
b_2\,|n_1, n_2\rangle &= \sqrt{n_2}\,|n_1, n_2 - 1\rangle \\
b_1^\dagger\,|n_1, n_2\rangle &= \sqrt{n_1 + 1}\,|n_1 + 1, n_2\rangle \\
b_2^\dagger\,|n_1, n_2\rangle &= \sqrt{n_2 + 1}\,|n_1, n_2 + 1\rangle\,.
\end{aligned}$$

If $\varepsilon \neq 0$, $|n_1, n_2\rangle$ is no longer an eigenfunction of the energy operator because the two CH stretching modes interact through mechanical and electromagnetic forces. Thus it seems reasonable to try a combination of such expressions, but what combination?

To answer this question, note that the operators $\hat{N}$ and $\hat{H}$ commute, implying that a nondegenerate eigenfunction of $\hat{H}$ will also be an eigenfunction of $\hat{N}$. For a fixed value of

$$n = n_1 + n_2 \,,$$

the most general eigenfunction of $\hat{N}$ is

$$|\psi_n\rangle = c_1|n,0\rangle + c_2|n-1,1\rangle + \cdots + c_n|1,n-1\rangle + c_{n+1}|0,n\rangle \,, \qquad (17.5)$$

where it is evident that

$$\hat{N}|\psi_n\rangle = n|\psi_n\rangle \,.$$

In this expression for $|\psi_n\rangle$, the c_j are $n+1$ arbitrary complex constants that can be determined by demanding that $|\psi_n\rangle$ be also an eigenfunction of $\hat{H}$. That is to say, the requirement

$$\hat{H}|\psi_n\rangle = E|\psi_n\rangle \qquad (17.6)$$

fixes the values of the c_j.

Thus Equation (17.6) can be viewed as $n+1$ equations for the coefficients of ket products: $|n_1, n_2\rangle$. All such equations can be written in matrix form as

$$\mathcal{H}\mathbf{c} = E\mathbf{c} \,,$$

where $\mathbf{c}$ is the column vector $\mathrm{col}(c_1, c_2, \ldots, c_{n+1})$ and

$$\mathcal{H} \equiv H_0 - \varepsilon V \,. \qquad (17.7)$$

The eigenvalues of the matrix $\mathcal{H}$ are the energy eigenvalues, and the corresponding eigenvectors—$\mathbf{c}$—together with Equation (17.5) determine the energy eigenfunctions. This approach to the solution of problems in quantum soliton theory is called the *number state method* [6].

The matrix $\mathcal{H}$ can be described as follows.

- With $\varepsilon = 0$, $\mathcal{H} = H_0$, a diagonal matrix of the form

  $$H_0 = \mathrm{diag}[\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_2, \alpha_1, \alpha_0] \,,$$

  where

  $$\alpha_j \equiv \omega_0 n - \frac{\gamma}{2}(n + n^2 - 2nj + 2j^2) \,.$$

- With $\varepsilon \neq 0$, off-diagonal elements are introduced into $\mathcal{H}$ through the additional matrix $-\varepsilon V$, where

$$V = \begin{bmatrix} 0 & \sqrt{n} & 0 & 0 & \cdot & 0 \\ \sqrt{n} & 0 & \sqrt{2(n-1)} & 0 & \cdot & 0 \\ 0 & \sqrt{2(n-1)} & 0 & \sqrt{3(n-2)} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \sqrt{3(n-2)} & 0 & \sqrt{2(n-1)} & 0 \\ 0 & \cdot & 0 & \sqrt{2(n-1)} & 0 & \sqrt{n} \\ 0 & \cdot & 0 & 0 & \sqrt{n} & 0 \end{bmatrix}.$$

Thus $\mathcal{H} = H_0 - \varepsilon V$ is an $(n+1) \times (n+1)$, real, nonsingular matrix with symmetry about both diagonals. From a perturbation theory in small ε/γ, Bernstein has shown that its lowest two eigenvalues differ by [1]

$$\Delta E = \frac{2n\varepsilon}{(n-1)!} \left(\frac{\varepsilon}{\gamma}\right)^{n-1} + \mathrm{O}(\varepsilon^{n+1}/\gamma^n)\,, \tag{17.8}$$

corresponding to eigenfunctions of $\hat{H}$ of the form

$$|\psi_n\rangle^{\pm} = \frac{1}{\sqrt{2}}(|n,0\rangle \pm |0,n\rangle) + \mathrm{O}(\varepsilon/\gamma)\,.$$

With a wave packet constructed as

$$|\Psi_n(t)\rangle = \frac{1}{\sqrt{2}}|\psi_n\rangle^{+} e^{-\mathrm{i}E^{+}t} + \frac{1}{\sqrt{2}}|\psi_n\rangle^{-} e^{-\mathrm{i}E^{-}t} + \mathrm{O}(\varepsilon/\gamma)\,,$$

then

$$|\Psi_n(0)\rangle = |n,0\rangle + \mathrm{O}(\varepsilon/\gamma)\,,$$

indicating that most of the energy is localized in oscillator #1 at time $t = 0$. How long will it remain there? Until the two main components in the wave packet $|\Psi_n(t)\rangle$ change their phase. This occurs in a *tunneling time* of order

$$\tau \sim \frac{\pi\hbar}{\Delta E}\,,$$

where $\Delta E \equiv |E^{+} - E^{-}|$ is measured in joules.

From numerical computations of the eigenvalues of $\mathcal{H}$, it is observed that the $\mathrm{O}(\varepsilon^{n+1}/\gamma^n)$ correction in Equation (17.8) is always negative so

$$\Delta E < \frac{2n\varepsilon^n}{(n-1)!\gamma^{n-1}}\,,$$

and the tunneling time τ for the initially localized energy to move from one oscillator to the other is

$$\tau > \frac{\pi\hbar(n-1)!\gamma^{n-1}}{2n\varepsilon^n}\,,$$

where γ is large compared with ε. As $n \to \infty$ (the classical limit), the tunneling time becomes arbitrarily long.

2.3 Comparison with experiments

In the previous subsection, we have derived several formulas; now the moment of truth is at hand. Does our quantum analysis bear any relation to experimental reality? Fortunately, measurements on the dihalomethanes—CCl_2H_2, CBr_2H_2, and CI_2H_2 —provide relevant data [10]. For each molecule, 15 transition energies between CH stretching vibrations have been measured, and our formulation contains only three parameters: ω_0, γ, and ε. A least squares fit of the eigenvalues of $\mathcal{H}$ to these spectral lines yields the optimum values for the parameters (in cm^{-1}) of the three molecules given in the following table.

Molecule:	CCl_2H_2	CBr_2H_2	CI_2H_2
ε	29.54	32.80	33.69
γ	127.44	125.45	124.25
$\omega_0 - \gamma/2$	3083.79	3089.53	3068.74
RMS errors	9.0	6.9	8.0

How good is the agreement? The last row of entries records the root mean square (RMS) errors in matching the spectral transitions for each molecule. By setting the three parameters ω_0, γ, and ε to the values indicated, our theory matches the 15 observed lines to better than 0.3%. It should be noted, however, that this analysis does not yield intensities for overtone lines, and it greatly underestimates the linewidths of the transitions.

3 A lattice nonlinear Schrödinger equation

Consider next a lattice nonlinear Schrödinger equation of the form

$$\left(\mathrm{i}\frac{d}{dt} - \omega_0\right) A_j + \varepsilon(A_{j+1} + A_{j-1}) + \gamma|A_j|^2 A_j = 0\,,$$

with $j = 1, 2, 3, \ldots, f$ and periodic boundary conditions so $A_{j+f} = A_j$. In this equation, γ is a parameter of intrinsic anharmonicity and ε introduces the dispersive effects of nearest neighbor interactions.

Introducing the transformation $A_j = \phi_j \exp[-\mathrm{i}(\omega_0 - \gamma)t]$, and quantizing by letting $\phi_j \to b_j$ and $\phi_j^* \to b_j^\dagger$, leads to a quantum discrete nonlinear Schrödinger (QDNLS) equation, which is related to the number operator

$$\hat{N} = \sum_{j=1}^{f} b_j^\dagger b$$

and the energy operator

$$\hat{H} = -\sum_{j=1}^{f} [\varepsilon(b_j^\dagger b_{j+1} + b_j^\dagger b_{j-1}) + \frac{\gamma}{2} b_j^\dagger b_j^\dagger b_j b_j] \,.$$

Since $\hat{H}$ commutes with both $\hat{N}$ and the translation operator $\hat{T}$, where

$$\hat{T}\,|n_1, n_2, \ldots, n_f\rangle = |n_2, \ldots, n_f, n_1\rangle \,,$$

a quantum analysis is begun by constructing a general eigenfunction of both $\hat{N}$ and $\hat{T}$. For $n = 1$

$$|\psi_{1,\tau}\rangle = c_1 \sum_{j=1}^{f} (\hat{T}/\tau)^{j-1} |100\cdots00\rangle$$

so

$$E_1 = -\varepsilon(\tau + \tau^{-1}) = -2\varepsilon \cos k \,.$$

Here $\tau = e^{\mathrm{i}k}$ is the eigenvalue of $\hat{T}$ that corresponds to the propagation number $k = 2\pi\nu/f$, where for f odd, $\nu = 0, \pm1, \pm2, \ldots, \pm(f-1)/2$.

This expression for the energy eigenvalues can be written as

$$E_1(k) = E_1(0) + \frac{k^2}{2m^*} + \mathrm{O}(k^4) \,,$$

where $m^* \equiv 1/2\varepsilon$ is called the *effective mass*.

With f odd and $n = 2$, a general eigenfunction of both $\hat{N}$ and $\hat{T}$ is

$$\begin{aligned} |\psi_{2,\tau}\rangle \quad &= \quad c_1 \sum_{j=1}^{f} (\hat{T}/\tau)^{j-1} |20\cdots0\rangle + c_2 \sum_{j=1}^{f} (\hat{T}/\tau)^{j-1} |110\cdots0\rangle \\ &+ \quad \cdots \quad + c_{(f+1)/2} \sum_{j=1}^{f} (\hat{T}/\tau)^{j-1} |10\cdots010\cdots00\rangle \,. \end{aligned} \tag{17.9}$$

Requiring that

$$\hat{H}|\psi_{2,\tau}\rangle = E|\psi_{2,\tau}\rangle$$

and equating the coefficients of kets leads directly to the block-diagonalized matrix

$$\mathcal{H}_2 = \mathrm{diag}[\mathcal{Q}(2,\tau_1), \mathcal{Q}(2,\tau_2), \ldots, \mathcal{Q}(2,\tau_f)]\,.$$

For a particular value of τ, energy eigenvalues are obtained from the matrix equation

$$\mathcal{Q}(2,\tau)\mathbf{c} = E\mathbf{c}\,,$$

where $\mathbf{c} \equiv \mathrm{col}(c_1, c_2, \ldots, c_{(f+1)/2})$, $\mathcal{Q}(2,\tau)$ is the $[(f+1)/2] \times [(f+1)/2]$ tridiagonal matrix

$$\mathcal{Q}(2,\tau) = -\begin{bmatrix} \gamma & q^*\sqrt{2} & 0 & \cdot & \cdot & \cdot \\ q\sqrt{2} & 0 & q^* & 0 & \cdot & 0 \\ \cdot & q & 0 & q^* & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & q & 0 & q^* \\ \cdot & \cdot & \cdot & \cdot & q & p \end{bmatrix},$$

and $q \equiv 1 + \tau$ and $p \equiv \tau^{(f+1)/2} + \tau^{(f-1)/2}$.

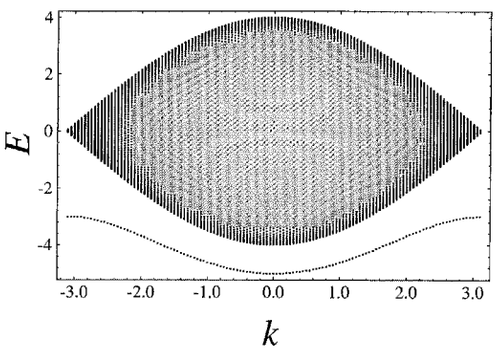

FIGURE 1. Energy eigenvalues for the QDNLS equation calculated from $\mathcal{Q}(2\tau)$ with $\tau = \exp(\mathrm{i}k)$, $\gamma = 3\epsilon$, and $f = 125$. The soliton band lies below a quasicontinuum. (Courtesy of J.C. Eilbeck.)

In Figure 3, are displayed the energy eigenvalues of $\mathcal{Q}(2,\tau)$ with $\varepsilon = 1$, $\gamma = 3$, and $f = 125$, indicating the existence of a quasicontinuum and a lower discrete band. Examination of the corresponding eigenfunctions shows that for the lower band the coefficient c_1 in Equation (17.9) is large compared to the other coefficients; thus this band indicates that the system prefers to be in a state where the two quanta sit on the same site.

Near $k = 0$, it is clear from Figure 3 that the energies of states in the soliton band are given by an expression of the form

$$E_2(k) = E_2(0) + \frac{k^2}{2m^*} + \mathrm{O}(k^4)\,.$$

In the limit $f \to \infty$, the soliton band has energy [14]

$$E_2(k) = -\sqrt{\gamma^2 + 16\varepsilon^2 \cos^2(k/2)}\,.$$

Defining the *soliton binding energy*, E_b, as the difference between the energy of the soliton band at $k = 0$ and the bottom of the continuum band implies that $E_\mathrm{b} = \sqrt{\gamma^2 + 16\varepsilon^2} - 4\varepsilon$ and $m^* = \sqrt{\gamma^2 + 16\varepsilon^2}/4\varepsilon^2$. Proceeding in this manner, it is possible—in principle—to construct block-diagonalized Hamiltonian matrices for any value of the quantum number n.

We are now in a position to consider the quantum description of a lattice soliton. For each value of the principal quantum number n and propagation number k, there is a lowest energy eigenvalue. These f lowest eigenvalues lie on a band $E = E_n(k)$, as is indicated in Figure 3. Each energy eigenvalue corresponds to a pure eigenstate, normalized as $\langle\psi_n(k)|\psi_n(k)\rangle = 1$. Wave packet solutions of the time dependent Schrödinger equation can be constructed as sums over the principal quantum number (n) and the propagation number (k); thus

$$|\Psi(t)\rangle = \sum_{n=1}^{\infty} a_n \sum_k G_n(k)|\psi_n(k)\rangle \exp\left(-\mathrm{i}E_n(k)t\right), \tag{17.10}$$

where k takes the f values between $-\pi$ and π that are indicated in Figure 3.

The wave function of Equation (17.10) is a *soliton wave packet*, which does not represent the most general wave function because it is constructed only from eigenfunctions with eigenvalues on the soliton bands.Thus the soliton wave packet is characterized by two interdependent properties: (i) for given values of k and n, $|\Psi(t)\rangle$ has the lowest energy, and (ii) under the same conditions, $|\Psi(t)\rangle$ has the highest probability of quanta being located near each other.

For larger values of n, it is convenient to calculate the $|\psi_n(k)\rangle$ in a Hartree approximation [16] and choose the a_n to correspond to a coherent state [7]. In the Hartree approximation, the binding energy is

$$E_\mathrm{b} \doteq \frac{\gamma^2}{48\varepsilon} n(n-1)^2\,,$$

whereas the exact binding energy is

$$E_\mathrm{b} = \frac{\gamma^2}{48\varepsilon} n(n^2-1)\,.$$

4 Local modes in crystalline acetanilide

In the early 1970s, Careri attempted to circumvent some of the problems arising in the study of natural protein by looking at hydrogen bonded molecular crystals that can be regarded as 'model proteins,' such as crystalline acetanilide ($CH_3CONHC_6H_5$) or ACN. The similarity of bond lengths and angles for the peptide group (HNCO) suggested that the dynamic behavior of ACN might provide clues to the corresponding behavior of natural protein. This comparison becomes more striking if one notes that both crystalline ACN and alpha-helix feature hydrogen bonded *peptide chains* with the atomic structure

$$\cdots \mathbf{H}-\mathbf{N}-\mathbf{C}=\mathbf{O}\cdots\mathbf{H}-\mathbf{N}-\mathbf{C}=\mathbf{O}\cdots\mathbf{H}-\mathbf{N}-\mathbf{C}=\mathbf{O}\cdots .$$

Careri's intuition was rewarded by the discovery of the anomalous resonance peak at 1650 cm^{-1} in crystalline ACN (shown in Figure 4), but a decade of effort established only that this peak was an Amide-I (CO stretching) resonance that could not be conventionally assigned. It was after all possibilities for conventional assignment of the band had been exhausted that dynamical self-trapping was considered [2].

The basic idea is that the 1650 cm^{-1} band is evidence of a self-trapped Amide-I state (or local mode) arising through extrinsic interactions with the lattice. This picture is related to—and it was inspired by—a soliton theory previously proposed by Davydov for the alpha-helix structures of natural protein [3, 11]. In both pictures the anharmonicity stems from lattice interactions, but Davydov invoked *acoustic* mode interactions, whereas Careri et al. considered *optical* mode interactions.

A theoretical model, therefore, is a one-dimensional lattice with a single phonon degree of freedom (sometimes called an *Einstein oscillator*) of frequency ω_0, coupled to each Amide-I mode through a phonon coupling parameter λ_0. An energy operator is defined as

$$\hat{H} = \hat{H}_{\rm CO} + \hat{H}_{\rm ph} + \hat{H}_{\rm int} \ ,$$

where

$$\begin{aligned} \hat{H}_{\rm CO} &= \sum_{j=1}^{f} \left[E_0 B_j^\dagger B_j - J(B_j^\dagger B_{j+1} + B_j^\dagger B_{j-1}) \right] \\ \hat{H}_{\rm ph} &= \hbar\omega_0 \sum_{j=1}^{f} \left(b_j^\dagger b_j + \frac{1}{2} \right) \\ \hat{H}_{\rm int} &= \lambda_0 \sum_{j=1}^{f} (b_j + b_j^\dagger) B_j^\dagger B_j \ , \end{aligned} \tag{17.11}$$

and j $(= 1, 2, \ldots, f)$ counts ACN molecules along the peptide chain. Thus $B_j^\dagger$ (B_j) are boson raising (lowering) operators for the Amide-I (CO stretch-

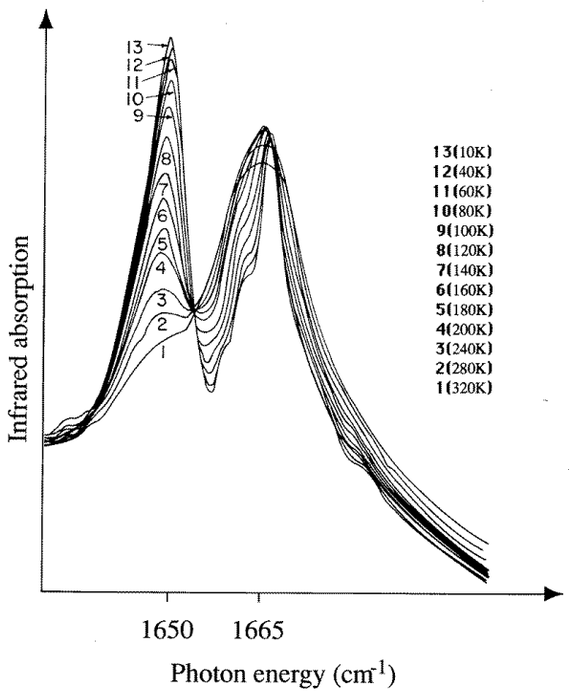

FIGURE 2. An infrared absorption spectrum of crystalline ACN in the Amide-I region. Notice both the anomalous peak at 1650 cm^{-1}, indicating a self-trapped (soliton) state, and the standard (unbound) peak at 1665 cm^{-1}. (Courtesy of G. Careri.)

ing) oscillators at site j, and $b_j^\dagger$ (b_j) are boson raising (lowering) operators for the corresponding Einstein oscillators.

Let us assume a single Amide-I quantum ($n = 1$) and a temperature of zero kelvin and, following Davydov, introduce a product approximation to the solution.[1] Thus

$$|\psi\rangle = |\Psi\rangle|\Phi\rangle\,,$$

where

[1]Although criticisms of the validity of this assumption have been advanced [11], one should recall that it is equivalent to corresponding formulations of the Born-Oppenheimer and Hartree approximations [12]. It is shown here that Davydov's approximation can be tested.

$$|\Psi\rangle = \sum_{j=1}^{f} a_j(t) B_j^{\dagger} |0\rangle_{\rm CO} \,,$$

and $|\Phi\rangle$ is a sum of coherent states for the Einstein oscillators for which $\langle\Phi|b_j|\Phi\rangle = \beta_j$. Treating $\langle\psi|\hat{H}|\psi\rangle$ as a classical Hamiltonian implies the dynamic equations

$$\begin{aligned} {\rm i}\hbar \frac{da_j}{dt} &= \left[E_0 + W + \lambda_0(\beta_j + \beta_j^*)\right] a_j - J(a_{j+1} + a_{j-1}) \,, \\ {\rm i}\hbar \frac{d\beta_j}{dt} &= \hbar\omega_0 \beta_j + \lambda_0 |a_j|^2 \,, \end{aligned}$$

where W is the total energy of the Einstein oscillators. Assuming that $\dot{\beta}_j = 0$ (the adiabatic assumption) and introducing the rotating transformation $a_j = \phi_j \exp[-{\rm i}t(E_0 + W)/\hbar]$ leads to the lattice nonlinear Schrödinger equation

$${\rm i}\hbar \frac{d\phi_j}{dt} + \frac{2\lambda_0^2}{\hbar\omega_0} |\phi_j|^2 \phi_j + J(\phi_{j+1} + \phi_{j-1}) = 0 \,, \tag{17.12}$$

where it turns out that the anharmonic parameter $(2\lambda_0^2/\hbar\omega_0)$ is about ten times greater than the dispersive parameter (J).

To establish this inequality, note first that from detailed electromagnetic calculations $J = 4\,{\rm cm}^{-1}$ [4]. Second, if the inequality is satisfied, then it is shown below that the binding energy of a self-trapped state (soliton) with respect to an unbound Amide I mode (at $k = 0$) is

$$E_{\rm b} \doteq \frac{\lambda_0^2}{\hbar\omega_0} - 2J \,. \tag{17.13}$$

Finally, in Figure 4, the 1650 ${\rm cm}^{-1}$ peak is assigned to a soliton and the peak at 1665 ${\rm cm}^{-1}$ to an unbound mode. This implies that $E_{\rm b} = 15\,{\rm cm}^{-1}$ so $(\lambda_0^2/\hbar\omega_0) = 23\,{\rm cm}^{-1}$, confirming the original assumption.

Because J is small compared with the extrinsic anharmonic parameter $(2\lambda_0^2/\hbar\omega_0)$, it is feasible to use a perturbation expansion in small J. This approach has two advantages: (i) it provides quantitative estimates of the accuracy of Davydov's product assumption, and (ii) solutions are obtained when the number of Amide-I quanta n is greater than unity. To this end, it is convenient to write $\hat{H}$ in the form

$$\hat{H} = \sum_{j=1}^{f} \hat{h}_j - J\hat{V} \,,$$

where

$$\hat{h}_j = \left[E_0 + \lambda_0(b_j + b_j^\dagger)\right] B_j^\dagger B_j + \hbar\omega_0 b_j^\dagger b_j$$

$$\hat{V} = \sum_{j=1}^{f} (B_j^\dagger B_{j+1} + B_j^\dagger B_{j-1}) \,,$$

and the ground state phonon energy is ignored.

The operator $\hat{h}_j$ is the energy operator of a *displaced harmonic oscillator* for which exact eigenfunctions are known [9]. Dropping the molecule index j $(= 1, 2, \ldots, f)$, the eigenfunctions of

$$\hat{h}|u\rangle = E|u\rangle$$

can be written as

$$|u\rangle = |n\rangle|\phi(n,\tilde{n})\rangle \,,$$

where

$$|\phi(n,\tilde{n})\rangle = \sqrt{\tilde{n}!}\exp\left[-\frac{1}{2}\left(\frac{n\lambda_0}{\hbar\omega_0}\right)^2\right] \times \sum_{m=0}^{\infty} \frac{(-n\lambda_0/\hbar\omega_0)^{m-\tilde{n}}}{\sqrt{m!}} \mathrm{L}_{\tilde{n}}^{m-\tilde{n}}\left[\left(\frac{n\lambda_0}{\hbar\omega_0}\right)^2\right] |m\rangle \,, \quad (17.14)$$

and

$$E = nE_0 - n^2\frac{\lambda_0^2}{\hbar\omega_0} \,.$$

In these equations, $B^\dagger B|n\rangle = n|n\rangle$, $b^\dagger b|m\rangle = m|m\rangle$, and $\mathrm{L}_n^m[\,\cdot\,]$ is an associated Laguerre polynomial.

Assuming periodic boundary conditions, a zero-order estimate (in powers of J) of the wave function is

$$|\psi\rangle = \frac{1}{\sqrt{f}}\sum_{j=1}^{f} e^{\mathrm{i}kj}|u_j\rangle \,, \quad (17.15)$$

where $k = 2\pi\nu/f$ and $\nu = 0, \pm 1, \ldots, f/2$ (or $\pm(f-1)/2$) for f even (or odd). This wave function is an eigenfunction of the translational symmetry operator.

Next make two assumptions. The first is $n = 1$, implying that only a single quantum of Amide-I oscillation is considered. The second is $\tilde{n} = 0$, implying that the phonon system is in the ground state, as one expects at

low temperatures. Neglecting terms of order $(\hbar\omega_0 J/2\lambda_0^2)^2 \sim 1\%$, the energy is

$$E(k) \doteq E_0 - \frac{\lambda_0^2}{\hbar\omega_0} - 2J \exp\left[-\left(\frac{\lambda_0}{\hbar\omega_0}\right)^2\right] \cos k\,. \tag{17.16}$$

Since the wavelength of infrared light is large compared with a lattice spacing, only $k = 0$ unbound states will be observed experimentally. From the first of Equations (17.11), the energy of such a state is

$$E_{\rm CO}(k=0) = E_0 - 2J.$$

If a wave packet of the energy eigenfunctions is constructed to localize a quantum of Amide-I energy near a single molecule, then all values of k must be equally represented, and from Equation (17.16) the average energy of the wave packet will be

$$E_{\rm sol} = E_0 - \frac{\lambda_0^2}{\hbar\omega_0}\,.$$

Assigning $E_{\rm CO}(k=0)$ to the 1665 $\rm cm^{-1}$ peak in Figure 4 and $E_{\rm sol}$ to the 1650 $\rm cm^{-1}$ peak implies that $E_0 - 2J = 1665$ and $E_0 - \lambda_0^2/\hbar\omega_0 = 1650$. Since $J = 4$ $\rm cm^{-1}$, we find that $\lambda_0^2/\hbar\omega_0 = 23$ $\rm cm^{-1}$, in agreement with our previous estimate of the anharmonic parameter, which was based upon Davydov's product assumption.

From Equation (17.16), the binding energy of a $k = 0$ soliton is

$$E_{\rm b} = E_{\rm CO}(k=0) - E(0) = \frac{\lambda_0^2}{\hbar\omega_0} - 2J\left\{1 - \exp\left[-\left(\frac{\lambda_0}{\hbar\omega_0}\right)^2\right]\right\}\,,$$

which is larger than that of Equation (17.13). Also from Equation (17.16) the effective mass of a $k = 0$ soliton is $M_{\rm sol} = m^* \exp(\lambda_0/\hbar\omega_0)^2$, where $m^* = \hbar^2/2a^2J$ is the effective mass of an unbound mode. From numerical analysis of Equation (17.12), on the other hand, it turns out that the soliton is pinned to the lattice, implying that $M_{\rm sol} = \infty$. These results provide quantitative estimates for errors arising from the product wave function.

In addition to the translational invariance of Equation (17.15), the exact wave function has two additional properties that are not shared by the product wave function:

(i) *Higher levels of Amide-I excitation.* One expects to observe an overtone series at the frequencies $\nu(n) = E(n) - E(0) \approx 1673n - 23n^2$ $\rm cm^{-1}$, where the coefficient of n has been chosen so $\nu(1) = 1650\rm cm^{-1}$. Measured overtone frequencies are as follows [15].

n	$\nu(n)$ measured	predicted
1	1650.0 ± 0.5	1649.3
2	3250 ± 1	3249.2
3	4803 ± 3	4799.7
4	6304 ± 5	6300.8

If the coefficients are adjusted to $\nu(n) = 1674n - 24.7n^2$, the measured values are predicted to an RMS error of 2.4 cm^{-1}. Thus a slightly better value for the extrinsic anharmonic parameter is

$$\frac{2\lambda_0^2}{\hbar\omega_0} = 49.4 \text{ cm}^{-1} .$$

(ii) *Higher levels of phonon excitation.* With $n = 1$ and $\tilde{n} = 0$, Equation (17.14) shows that $|\phi(n, \tilde{n})\rangle$ reduces to a coherent state. What about $\tilde{n} > 0$? For $n = 0$

$$|\phi(0, \tilde{n})\rangle = |\tilde{n}\rangle ,$$

a phonon number state with $b^\dagger b|\tilde{n}\rangle = \tilde{n}|\tilde{n}\rangle$, which will be thermally populated when $kT \geq \hbar\omega_0$. Taking this into consideration leads to a quantitative understanding of the temperature dependence of the 1650 cm^{-1} band that is shown in Figure 4.

To calculate the temperature dependent intensity of the 1650 cm^{-1} band, one must find the sum of all transitions from the ground states $|\tilde{n}\rangle$ to first excited states $|\phi(1, \tilde{n})\rangle$ with $m = \tilde{n}$. Since the ground states are thermally populated with probabilities

$$P(\tilde{n}) = \left[1 - \exp\left(-\frac{\hbar\omega_0}{kT}\right)\right] \exp\left(-\tilde{n}\frac{\hbar\omega_0}{kT}\right) ,$$

the temperature dependence of the 1650 cm^{-1} band is given by

$$W(T) = \sum_{\tilde{n}=0}^{\infty} P(\tilde{n}) |\langle \tilde{n}|\phi(1, m)\rangle|^2 .$$

Using identity number 9.976 from Gradshteyn and Ryzhik [8], this sum is computed to be

$$W(T) = \exp\left[-\left(\frac{\lambda_0}{\hbar\omega_0}\right)^2 \coth\left(\frac{\hbar\omega_0}{2kT}\right)\right] I_0\left[\left(\frac{\lambda_0}{\hbar\omega_0}\right)^2 \text{csch}\left(\frac{\hbar\omega_0}{2kT}\right)\right] , \tag{17.17}$$

with $I_0[\,\cdot\,]$ a modified Bessel function of the first kind.

Experimental values for $W(T)$ are as follows [13],

Temperature (K)	σ	Relative intensity	σ
21	4	1.000	0.016
21	4	0.996	0.016
53	4	0.892	0.019
53	4	0.895	0.019
100	10	0.710	0.022
100	10	0.726	0.022
149	6	0.496	0.025
149	6	0.507	0.025
227	12	0.356	0.028
227	12	0.346	0.028
305	8	0.170	0.032
305	8	0.169	0.032

where the σ are RMS errors.

Using the value $\lambda_0^2/\hbar\omega_0 = 24.7\,\mathrm{cm}^{-1}$, the only free parameter in Equation (17.17) is $\hbar\omega_0$. Fitting to the experimental data requires the optical mode frequency to be $\hbar\omega_0 \sim 75\,\mathrm{cm}^{-1}$, a value that is in accord with the far infrared lattice phonon spectrum of ACN [2].

5 Conclusions

Although this survey of quantum lattice solitons has skipped over many details, we have managed to uncover a number of experimentally useful results. Since quantum effects appear primarily at the atomic level, emphasis has been upon understanding the structures of states with few quanta, as is expected in spectral observations of solitons on molecules and molecular crystals.

In addition to such problems, we have considered soliton systems with translational symmetry, a property that must be shared by the eigenfunctions of the corresponding quantum system. Wave packets describing quantum corrections to the classical behavior of a soliton are recognized as generalized Fourier transforms composed of extended eigenfunctions, and the quantum uncertainty principle stems from the fact that components oscillating at different frequencies will eventually get out of step. Thus the wave packet describing the location of a nonlinear lattice solitary wave or soliton disperses with time just as does that for an electron. From the perspective of quantum theory, one is no more or less a 'particle' than the other.

Acknowledgments: My understanding of quantum lattice solitons has developed over the past two decades through fruitful collaborations and discussions with many colleagues. Those who have made substantial contributions to this chapter include: Mariette Barthes, Lisa Bernstein, Irving Bigio,

Giorgio Careri, Alexander Davydov, Chris Eilbeck, Victor Enol'ski, Henriette Gilhøj, Enrico Gratton, Jim Krumhansl, Mario Salerno, and Ewan Wright. It is a pleasure to thank them for the many ideas and insights they have shared.

6 References

[1] L Bernstein, J C Eilbeck, and A C Scott. The quantum theory of local modes in a coupled system of nonlinear oscillators. *Nonlinearity* 3 (1990) 293–323.

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

[3] A S Davydov. *Solitons in Molecular Systems.* 2nd edition. Reidel, Dordrecht, 1991.

[4] J C Eilbeck, P S Lomdahl and A C Scott. Soliton structure in crystalline acetanilide. *Phys. Rev. B* 30 (1984) 4703–4712.

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

[6] 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.

[7] R J Glauber. Coherent and incoherent states of the radiation field. *Phys. Rev.* 131 (1963) 2766–2788.

[8] I S Gradshteyn and I M Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, New York, 1980.

[9] H Haken. *Quantum Field Theory of Solids.* North-Holland, Amsterdam, 1976.

[10] O S Mortensen, B R Henry, and M A Mohammadi. The effects of symmetry within the local mode picture: A reanalysis of the overtone spectra of the dihalomethanes. *J. Chem. Phys.* 75 (1981) 4800–4808.

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

[12] A C Scott. *Nonlinear Science: Emergence and Dynamics of Coherent Structures.* Oxford University Press, Oxford, 1999.

[13] A C Scott, I J Bigio and C T Johnston. Polarons in acetanilide. *Phys. Rev. B* 39 (1989) 12883–12887.

[14] A C Scott, J C Eilbeck, and H Gilhøj. Quantum lattice solitons. *Physica D* 78 (1994) 194–213.

[15] A C Scott, E Gratton, E Shyamsunder and G Careri. The IR overtone spectrum of the vibrational soliton in crystalline acetanilide. *Phys. Rev. B* 32 (1985) 5551–5553.

[16] E Wright, J C Eilbeck, M H Hays, P D Miller, and A C Scott. The quantum discrete self-trapping equation in the Hartree approximation. *Physica D* 69 (1993) 18–32.

18
Noise in Molecular Systems

G. P. Tsironis

ABSTRACT Molecular and macromolecular systems in contact with fluctuating environments close to equilibrium present several interesting transport features. We will focus on simple systems constituted of particles in periodic non-symmetric potentials under the influence of non-equilibrium noise. The resulting ratchet or motor motion is manifested through non-zero persistent particle currents. These model systems are not dissimilar to motor proteins executing directed motion on cytoskeletal filaments.

1 Introduction

The simultaneous presence of nonlinearity and non-equilibrium noise in a molecular system results many times in novel, interesting physical properties. The stochastic motion of a particle in a periodic not symmetric potential and the resulting *ratchet effect* has been the focus of attention recently [1, 2]. The original motivation of this effect was in fact biological: When a microtubular associated protein (MAP) executes motion on a microtubule, its diffusive dynamics has a specified direction. This directionality in the protein motion is associated with the non-symmetric form of the periodic potential of the microtubule and is thought to be induced by the correlated character of the ATP hydrolysis mechanism. Since the original work on the correlated ratchet effect there has been an increased interest in the phenomenon, both experimental and theoretical. On the experimental front, optical, electrical and mechanical systems have been shown to have the ratchet property. On the theoretical front, there have been several extensions of the ratchet effect involving compound objects [3, 4] and solitons [5].

The understanding of the exact mechanism of motor protein motion in the cytoskeleton [6]-[18] has led physicists in the study of several stochastic ratchet models the simplest of which involves one overdamped particle, representing, for instance, the motor protein kinesin, in a periodic but not symmetric force field, driven by different types of correlated noises. The periodic forces are exerted by the $8nm$ long α-β partially asymmetric tubulin dimers on kinesin while the noise terms represent the fluctuating environment. This model leads to macroscopic particle current in a specific

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 357−370, 2000.

direction determined by the potential asymmetry and the properties of the noises. In the additive colored noise case, the finite correlation time τ corresponds to ATP kinesin binding event time and subsequent energy release through hydrolysis. The ATP hydrolysis rate is approximately $50s^{-1}$ and results on the average in one $8nm$ kinesin step per ATP cycle, i.e. to a typical kinesin speed of $400nm/s$ [14, 15]. The single particle colored noise model gives in the high correlation noise regime a natural interpretation for the kinesin steps observed in experiments. In this regime the Brownian particle simply waits in the potential minimum of a tubulin unit until the right fluctuation arrives that making it "escape" to the next tubuline dimer [19, 20]. The noisy environment is ambient liquid containing a variety of molecules including ATP molecules at μM concentration levels. A large number of unsuccessful binding attempts of ATP molecules contribute to medium fluctuations while the successful critical noise fluctuation can be interpreted as an ATP successful binding on kinesin. The critical binding fluctuation determines the average exit time $< T >$, or average step time, leading to a distance x after n steps, i.e. $x \approx n < T >$.

2 Additive correlated ratchets

Consider the basic "engine" equations in the Langevin picture,

$$\begin{aligned} \frac{dx}{dt} &= f(x) + \xi(t) \\ \frac{d\xi}{dt} &= -\frac{1}{\tau}\xi + \frac{1}{\tau}\eta(t). \end{aligned} \tag{18.1}$$

Here x denotes the position of the Brownian particle in the overdamped limit, and $f(x) \equiv -V'(x)$ where $V(x)$ is a periodic non-symmetric potential. Assuming a piece-wise linear potential with barrier height Q and lengths d_1, d_2 ($d_1 < d_2$) we have in each periodic unit cell ($\alpha - \beta$ tubulin units) of length $d_1 + d_2$ the positive force Q/d_1 and the negative force $-Q/d_2$, where the latter is smaller in absolute value than the former. The auxiliary variable ξ represents the coupling of the particle to the environment; if the noise variable $\eta(t)$ is Gaussian and delta-correlated, $< \eta(t)\eta(t') >= D\delta(t-t')$, then $\xi(t)$ is an Ornstein-Uhlenbeck process, that is, $\xi(t)$ is Gaussian and exponentially correlated,

$$< \xi(t)\xi(t') >= \frac{D}{2\tau} e^{-\frac{|t-t'|}{\tau}}, \tag{18.2}$$

with white noise strength D being proportional to the ambient temperature. When there are time-correlations in the noise the system is "open" in the thermodynamic sense, and a non-zero current can arise due to the

asymmetry of the periodic potential. As a result, the particle can move in a specified direction even when the driving fluctuations are completely symmetric. The particle effectively acts as an engine that operates in the Brownian regime. The engine consumes energy extracted from the nonequilibrium fluctuations of the environment and transforms it into mechanical work, manifested by its average net velocity in a given direction. The efficiency of this engine is determined mainly by two characteristics of the system: the asymmetry of the ratchet potential, and the "environmental" features such as the correlation properties of the non-equilibrium fluctuations and the ambient temperature. The induced particle motion is a finite temperature phenomenon that disappears at very high temperatures since then the details of the asymmetric ratchet potential are washed out by the noise. The presence of environmental correlations is included in the model through additive colored noise.

When τ is exactly zero there is no current; in the limit $D \to 0$ and for τ very small the induced current also essentially vanishes. In the other limit, that is, when $\tau \gg 1$, we can evaluate an asymptotic expression for the current using the following argument. Let us initially place the Brownian particle at the bottom (minimum) of one of the cells of the potential. For extremely correlated noise, i.e., when $\tau \gg 1$, the effect of the white noise $\eta(t)$ in the evolution of ξ, given in Eq. (18.1), is negligible. Thus in this limit we can set $\dot{\xi} \approx 0$. The net force acting on the particle is the quasi-static force $f + \xi$, i.e. the force ξ fluctuates but very slowly in the time-scale of particle motion. The particle escapes to the next well, left or right, when a fluctuation leads to a value of ξ of the appropriate size to cancel the force due to the potential. The particle is thus brought to the "top of the barrier." The value of ξ must remain essentially constant for the time that it takes the particle to reach this position. Once there, the particle can immediately "roll down" to the next minimum in a time that is small compared to the time it has waited for the appropriate fluctuation. In this picture, the average time that the particle waits before passing from one well to another is given by the mean first passage time for the noise ξ to reach the appropriate critical value ξ_c to cancel the effects of the potential [19, 20]. For the particle to escape to the right, the noise must reach the critical value $\xi_1^c = Q/d_2$, and to escape to the left the critical value of the noise is $\xi_2^c = -Q/d_1$. The mean first passage time for the Ornstein-Uhlenbeck process $\xi(t)$ to reach a value ξ_i^c is [21]

$$T_i(\xi_i^c) \;=\; \frac{\sqrt{2\pi D\tau}}{|\xi^c{}_i|}\;\exp\left(\frac{\xi^c{}_i{}^2\tau}{2D}\right). \tag{18.3}$$

The color-induced current is proportional to the net rate R of escape from a well. Inserting the appropriate critical values of ξ, this rate is then given

by

$$R = \frac{1}{2}\left[\frac{1}{T_1(\xi_1^c)} - \frac{1}{T_2(\xi_2^c)}\right] = \frac{Q}{2\sqrt{2\pi D\tau}}\left[\frac{1}{d_1}\exp\left(-\frac{Q^2\tau}{2D{d_1}^2}\right) - \frac{1}{d_2}\exp\left(-\frac{Q^2\tau}{2D{d_2}^2}\right)\right]. \tag{18.4}$$

The rate (18.4) exhibits a maximum at a finite value of τ (whose specific value depends on the other parameters) even though it is an asymptotic expression strictly valid only in the large-τ regime. The formula thus qualitatively correctly captures the finite-τ dynamics: it is known, mainly from numerical simulations, that there is indeed a maximum in the current at intermediate correlations [19] (Figure 1).

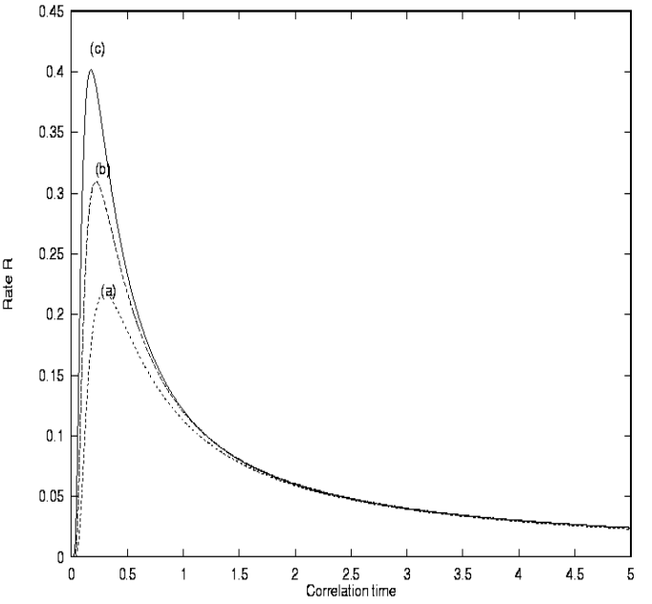

FIGURE 1. Color-induced rate R, proportional to the net current, as a function of the correlation time τ. The parameter values are $Q = 0.5$, $d_1 = 2.5$, $d_2 = 0.5$, and (a) $D = 0.3$, (b) $D = 0.4$, (c) $D = 0.5$.

In order to apply these ideas to true MAP's we use for simplicity an "upper bound" estimate for the mean first passage time from one tubu-

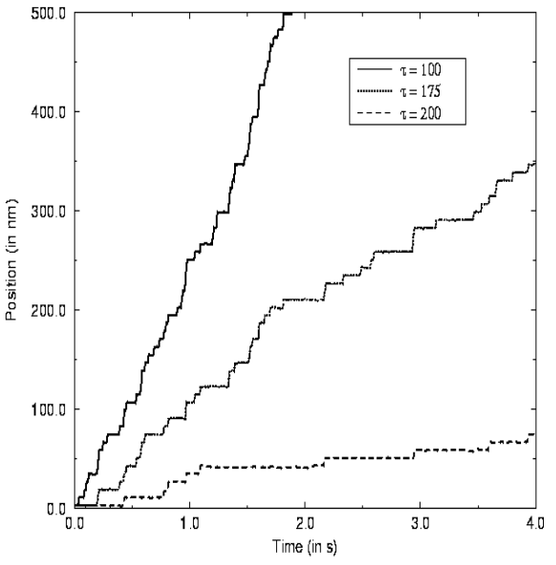

FIGURE 2. Position-time traces for highly correlated, quasi-deterministic Brownian engines showing the characteristic, step-like forward motion of molecular motor proteins. The engine correlation time τ is the unique adjustable parameter and relative large values of which denote a quasi-deterministic process. Protein motor position in nanometers as a function of time in seconds for different correlation times τ in nanoseconds. For the trajectories we used a potential barrier $Q = 10k_BT$ and α-tubulin binding site $d_t = 2.4nm$ leading to forces of approximately $16.6pN$ and $-7.1pN$. Higher correlation times lead to step-like protein motion with long quiescent times while shorter correlation times result in biased stochastic motion with completely different features.

lin dimer to the next. We assume that kinesin binds primarily to one of the tubulin molecules leading to relatively asymmetric hypercell tubulin potential ignoring thus exit times towards the high force direction as very improbable. The mean exit time to the forward direction then (including explicitly the damping coefficient γ) is

$$< T > = \frac{\sqrt{2\pi D\gamma\tau}}{|\xi_c|} \exp\left(\frac{{\xi_c}^2\tau}{2D\gamma}\right). \tag{18.5}$$

which is nothing but the Kramers time for the fluctuating force ξ to reach for the first time the critical value ξ_c. The critical force ξ_c necessary for canceling the tubulin force $f(x)$ and rendering kinesin free to move to the next tubulin hypercell is equal to $\xi_c = Q/d_t + \gamma d_t/\tau$. The first "dominant" term is the total force need to be overcome and equal to the potential maximum Q over the distance d_t traveled while an additional "correction" term (see ref. [20]) permits kinesin to move to the next cell before the highly correlated force ξ acquires a new value and interrupts its exit flight.

The average upper bound residence time expression depends critically on the correlation time τ as its only adjustable parameter and this dependence is in the form $< T >= A(\tau)exp[S(\tau)]$. To find an estimate of this time we consider a potential barrier is $Q \approx 8k_BT$ while taking $d_t \approx 5.3nm$ leads to $Q/d_t \approx 6pN$. We also use $D = k_BT \approx 4 \times 10^{-21}J$ at room temperature and $\gamma = 6 \times 10^{-11}kg/s$ and obtain $\xi^c \approx 6pN + 318pN/\tau[ns]$ where $\tau[ns]$ is the correlation time in nanosecond units. The final expression in "engineering units" is [20]

$$< T > = \ 6.5 \ ns \ \sqrt{\tau[ns]} \ e^{8+0.075\tau[ns]+210\tau} \tag{18.6}$$

We note the dramatic exponential dependence of the average step time on the correlation time τ. For these numbers a correlation time of $\tau \approx 100ns$ leads to average exit time $< T >\approx 2sec$ or a kinesin speed of $4nm/s$ which is in general expected range but substantially smaller than the observed ones due to the asymptotic character of the expression. Numerical simulations, on the other hand, for the same range of parameters give smaller mean exit times, leading to kinesin speeds one to two orders of magnitude larger, results that are compatible with several experimental data. The simple Brownian ratchet model for kinesin in the high correlation regime due to its quasi-deterministic character (Figure 2), gives reduced position variance compatible with the experiments. In the additive ratchet case, the variance becomes substantially reduced with the increase of the correlation time while backward slips are extremely improbable even for small potential asymmetries.

From the compatibility of the simple correlated Brownian ratchet model with the qualitative and quantitative features of kinesin motion a simplified picture emerges for the movement of motor proteins on microtubles. We find that the main features of this picture are the asymmetric periodic tubulin potential and the coupling to a stochastic environment that is necessarily highly correlated and quasi-deterministic. The single adjustable parameter of the model is the noise correlation time that physically is related to the event times for ATP binding at the active kinesin head as well as the hydrolysis time. For reasonable kinesin parameters this time is of the order of $100ns$ to $1\mu sec$, a reasonable range for the phenomena involved. It is not the aim of this simple Brownian model to account fully for the specifics of the kinesin walk by taking into account higher dimensional features.

It is however possible to extend it in this direction for a more complete quantitative agreement with the data.

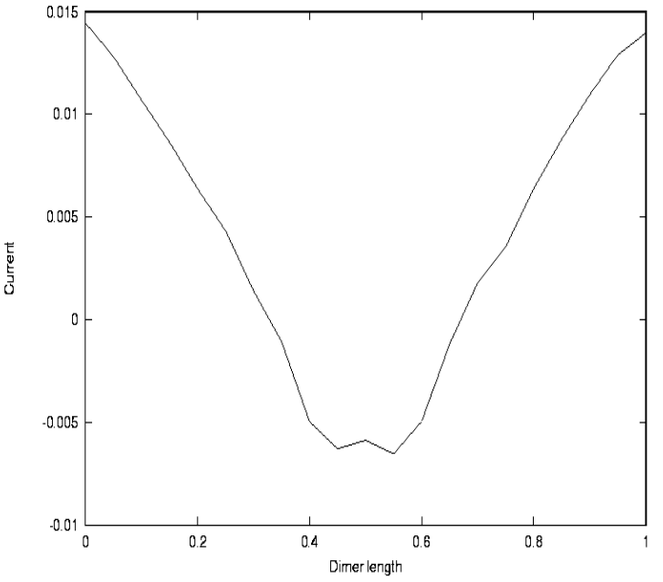

FIGURE 3. Simulation of the current as a function of the length of the rigid Brownian dimer on the sawtooth ratchet. Parameter values are $Q = 1.0$, $d_1 = 0.55$, $d_2 = 0.45$, $\tau = 0.1$, and $D = 0.5$.

3 Current reversal

Current reversal on ratchet systems that arise as a consequence of a number of modifications that can be made to the problem as stated above have also been of interest. We mention in particular the current inversion found when in place of a point particle the Brownian particle is an extended rod (a rigid dimer). The dimer consists of a Brownian particle of length l that experiences a net potential $\tilde{V}(x)$ that is a superposition of the periodic ratchet potential $V(x)$ evaluated at x and at $x + l$, i.e., $\tilde{V}(x) = V(x) + V(x+l)$ where $V(x)$ is an asymmetric periodic potential. Current inversion

has been calculated analytically and observed numerically for a dimer on the ratchet when the driving noise is dichotomous [3]. The results shown in Figure 3 confirm this behavior when the driving noise is Gaussian colored noise. For very short rods (not unlike point particles) and for long rods the current is to the right, but for rods of intermediate length there is a regime where the ratchet current, driven by the same Gaussian noise, is actually to the left. The simulations for the same rod lengths used in [3] demonstrate that the current inversion range is approximately the same as in the case of dichotomous noise [19]. An increase in the number of simulations and/or of the lengths of the trajectories would again smooth out the results and would more clearly display the fact that the current is symmetric about a rod length of 0.5 for the potential parameters used in the figure. The rigid rod or rigid dimer model can be used to describe very simply in the context of the simple ratchet model the possibility of current reversals, and as a result the differential motion of similar motor proteins along the microtubule. Nevertheless, more involved models are necessary if one wants to address additional problems, such as the dynamics of artificial motor proteins.

4 Synthetic motor protein motion

The molecular motors kinesin and non-claret disjunctional (ncd) belong to the same superfamily of motor proteins, and while they are *almost* identical they nevertheless move towards opposite ends of microtubules [6, 7, 11]. While the protein catalytic domain seems to be responsible for the pocessivity of the motor on the microtubule, the "neck" region adjacent to the motor heads was found recently to control the directionality of movement [12, 13]. A simple Newtonian model of two motor head particles connected through a neck coiled-coil spring (non-rigid dimer) whose rest length changes with each ATP hydrolysis event captures the essential motor dynamics features. In particular, the observed directionality reversal in synthetic "chimaeras" with different coiled-coil regions results in the model from a change in the stiffness of the spring coefficient. We find that motor speed is determined by the average ATP absorption rate while the effect of ambient temperature is small, leading to essentially non-Brownian, deterministic motor motion [22].

In our synthetic motor protein model x_1, x_2 denote the positions of the two dimer heads interacting with the microtubule surface through a one dimensional periodic non-symmetric potential with a unit cell equal to $8nm$ and with each-other through an internal harmonic potential $V_s(x_1, x_2) = \frac{1}{2}\kappa(|x_1 - x_2| - l(t))^2$ where $l(t)$ is the time-dependent rest length of the spring that takes two values corresponding to two different states of the dimer. In state (a) $l(t) = l_1$, the two heads are close to each other while

after an ATP hydrolysis event a protein conformational change takes place and the dimer transits to state (b) characterized by an opened α-helical coiled-coiled with length $l(t) = l_2$. In the following we take $l_1 = 0$ and $l_2 = 8nm$. Once ADP is released state (b) transits to state (a) and the conformational change cycle repeats with an ATP hydrolysis rate taken here to be approximately $50s^{-1}$. We assume initially that $l(t)$ varies periodically in time with the dimer spending equal time at each state leading to a period of the conformational change cycle $t_c = 0.02sec$ while $t_a = t_b = t_c/2$ where t_a, t_b are the time the dimer stays at states (a) and (b) respectively. This assumption for the specific time-dependence of $l(t)$ can be easily relaxed. Each protein head is in contact with a heat bath at temperature K_BT simulated through a Gaussian white noise term. In the overdamped limit the Langevin equations of motion that describe fully the motion of each of the protein heads read

$$\begin{aligned} \gamma \dot{x}_1 &= f_p(x_1) + f_s^1(x_1) + \xi_1(t) \\ \gamma \dot{x}_2 &= f_p(x_2) + f_s^2(x_2) + \xi_2(t). \end{aligned} \tag{18.7}$$

The forces due to the head interaction with the microtubule are $f_p(x) = -\partial_x V_p$ where $V_p(x)$ is the periodic non-symmetric microtubule potential taken to be piecewise linear with period $8nm$ given explicitly over one period through $V_p(x) = Q\frac{x}{d_1}$ for $0 \leq x \leq d_1$ and $V_p(x) = Q[1 + \frac{d_1 - x}{d_2}]$ while $d_1 \leq x \leq d_1 + d_2$. The coiled-coil spring force is $f_s^i(x) = -\partial V_s/\partial x_i$, γ is a dissipation constant and $\xi_i(t)$ is the Gaussian white noise where $< \xi_i(t_1)\, \xi_j(t_2) >= k_BT\gamma\, \delta_{ij}\, \delta(t_1 - t_2)$. For the calculations we take $\gamma = 6 \times 10^{-11} kg/sec$, $Q \approx 10 k_BT$ and consider the asymmetry in the binding site within the α-β tubulin unit to be represented reasonably through the selection $d_1 = 2nm$, $d_2 = 6nm$. Finally, we keep the coiled-coil spring coefficient κ an unknown adjustable parameter that has different values depending on the specific neck domain of different motor proteins. In order to demonstrate the fundamentally deterministic aspect of the motion, we will ignore at first the noise terms and show qualitatively that the motor dimer moves due to spring conformational changes and that stiff springs support motion to the left, while soft springs support motion to the right. In this fully newtonian case, there are only two forces acting upon the heads, the spring force f_s and the microtubule potential force f_p taking the two values f_p^r (steep potential side) and f_p^l (less steep slope).

Let us first examine the case when the spring connecting the protein heads is soft, as shown in the motion snapshots of Figure 4. Initially $l(t) = l_1 = 0$ and the two heads relax to the bottom of the lower microtubule well. When ATP binds and the coiled-coil unwinds, the rest length of the spring changes to $l(t) = l_2$ while the spring tension tends to move the two heads apart in order to relax the spring. For ultra soft springs with $f_p >> f_s$, the spring tension cannot surmount the microtubule potential force and both heads remain trapped in the bottom of the well. As the

spring tension increases the head which is pushed by the spring upwards ("lower head") still cannot surmount the steep potential slope and remains at the bottom of the well. On the contrary, the "upper" head is pushed up towards the weaker force ($f_p^l \leq f_s$), and the head moves upwards until the spring relaxes. At that stage the distance between the two heads is exactly l_2, i.e. equal to the period of the potential and the spring is relaxed, while at the same time each single head relaxes at the bottom of two adjoining wells. Implicit in this analysis is the assumption that the spring remains at its expanded state sufficiently long time for the right head to reach the neighboring well. Subsequently the motor spring transits to the shrinked state and thus the two heads tend to come closer to each other; they do this using the easier path and thus the lower head moves forward completing the one ATP-8nm step.

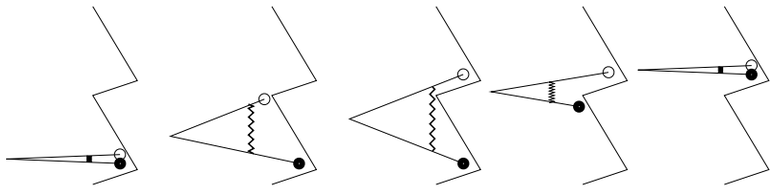

FIGURE 4. Quasi-deterministic protein motor motion on a microtubule, depicted vertically as a saw-tooth potential. One ATP molecule is consumed per step of the walk. The coiled-coil spring connecting the two motor heads is soft and the motor moves forward towards the "easy" microtubule potential slope. Filled circle designates the attached, lagging head. From left to right: **a.** Attached state. **b.** ATP hydrolysis event and coiled-coil spring opening. **c.** Short-lived open spring relaxed state. **d.** Motor steps forward through lagging head detachment. **e.** ADP release and motor relaxation in attached state in the next tubilin unit ready for repetition of the cycle.

We now consider the stiff spring case i.e. $f_s >> f_p^r > f_p^l$ (Figure 5). The two heads are initially relaxed at the top well, i.e. $l(t) = l_1$. When the spring transits from state (a) to state (b), the spring tension pushes the two heads apart. Since the potential force is negligible in comparison to the spring tension the two heads move equidistantly apart and, due to geometrical reasons, the lower (black) head moves "backwards". When the spring closes again the two heads move towards each other and for the same reasons as previously they meet in the lower cell. Thus, in this case, after one ATP cycle the motor protein movers one step "backwards".

We found qualitatively that the specific value of the spring coefficient connecting the two protein dimers plays a decisive role in directionality selection. In order to connect this simple deterministic motor mechanism with the kinesin experimental data we perform numerical simu-

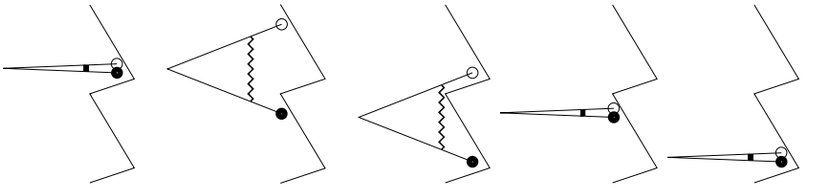

FIGURE 5. Quasi-deterministic motor dimer walk when the coiled-coil spring is hard and the motor moves forward towards the "difficult" microtubule potential slope. The microtubule is depicted vertically. Filled circle designates the attached, lagging head. From left to right: **a.** Attached state. **b.** ATP hydrolysis event and coiled-coil spring opening. **c.** Short-lived open spring relaxed state. **d.** Motor steps backward through lagging head detachment. **e.** ADP release and motor relaxation in attached state in the next tubilin unit ready for repetition of the cycle.

lations evaluating the protein velocities for different spring coefficient κ, corresponding to different possible motor proteins and chimaeras. The results for the mechanical protein model are shown in Figure 5 where we assumed that the process of opening and closing the coiled-coil spring is stochastic. We observe directionality dependent on the actual calculation of the spring coefficient κ, compatible with the synthetic protein experiments. In addition to the spring dependent directionality reversal, the model gives motor protein velocities in the expected experimentally range for an ATP burning rate of $50s^{-1}$, or equivalently averaged residence times $< t_a > + < t_b > = < t_c > = 0.02sec$. These velocities increase linearly with ATP concentration (or $< t_a >^{-1}$) until saturation, a feature obtained here in the very fast switching regime. We note that exact residence portion the dimer remains at its shrinked ($< t_a >$) or strained ($< t_b >$) state is not important while $min(< t_a >, < t_b >) >> t_s$ where t_s is the time the heads need to move to the new ground state of the spring after some transition from state (a) to (b) (or from (b) to (a)) and is typically of the order of hundreds nanoseconds. The results do not depend critically on the asymmetric potential details since for $d_1 = 1-3nm$ and $d_2 = 5-7nm$, we obtain similar qualitative results but with slightly modified κ-regions. Also, we could use different values -within a certain region- of l_1, and l_2, without altering the qualitative features of the dynamics of the system. For instance, numerical simulations with $l_1 = 1$ and $l_2 = 9$ yield similar results. Finite temperature Monte Carlo simulation results demonstrate that the basic features of the zero-temperature model are preserved in the finite temperature case [22].

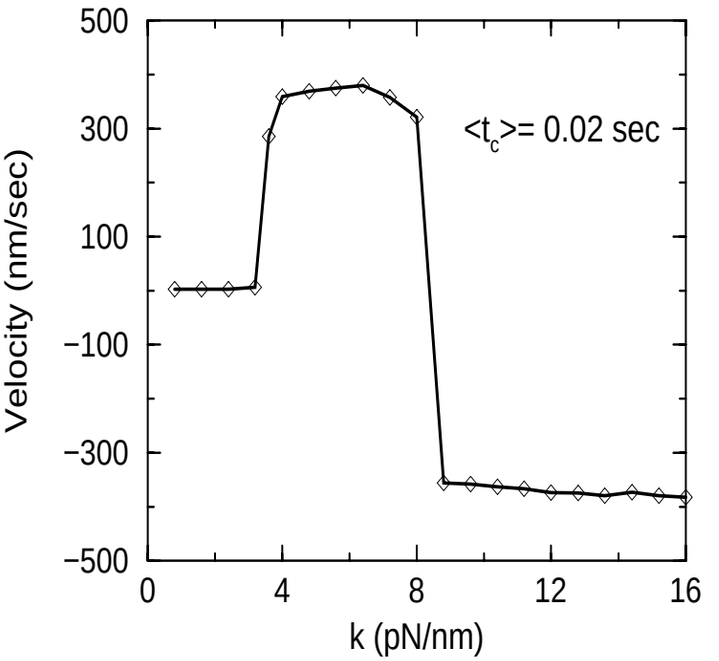

FIGURE 6. Synthetic motor velocity as a function of the spring coefficient κ for an averaged ATP absorption rate $50s^{-1}$. The direction of motion depends on the protein neck spring.

5 Targeted energy transfer and nonequilibrium fluctuations in bioenergetics

The simple mechanical and stochastic models used in order to understand and analyze active transport within a cell have been so far phenomenological. In a more realistic physical model, one needs to worry about the specifics of the energy transduction, viz. what are the mechanisms for the energy transfer from the ATP binding site to relatively long distances (more than one nanometer) and how the conformational changes occur that enable motor motion. Additionally, one should investigate the role of the nonequilibrium fluctuations in this process of energy transfer as well as the role "landscapes" play in these processes [23]. Work in nonlinear lattice systems demonstrates that intrinsic localized modes, or discrete breathers, might be responsible for energy storage and transfer in nonlinear networks that re-

semble somehow proteins [24]. In particular, we now know that breathers induce multiple time-scales in nonlinear networks as well as are able to transfer energy in discrete packages from site to site. What also seems to be possible is that under some general conditions breathers can induce *targeted energy transfer* in the lattice, that is both very efficient and very site specific [24]. These latter features provide additional evidence that breathers can provide a new paradigm for energy storage as well as transfer in complex biological units. Even though specifics have to be worked out in detail in the future, the already established properties of breathers, viz. discreteness, stability, essentially lossless transport and robustness to thermal fluctuations turn them into appealing agents for efficient energy transduction in macromolecular systems. In the motor protein dynamics case, for instance, one could invoke a targeted breather mechanism for the transfer of energy from the ATP binding site to the neck region that controls the motion. These an other related issues at the interface of physics with biology will be certainly addressed in the near future.

Acknowledgments: I wish to thank S. Aubry, H. Frauenfelder and K. Lindenberg for very stimulating discussions on the topics addressed in this chapter.

6 References

[1] F. Julicher, A. Ajdari and J. Prost, Modeling molecular motors, Rev. Mod. Phys. **69**, 1269 (1997) and references therein.

[2] R. D. Astumian, Thermodynamics and Kinetics of a Brownian Motor, Science **276**, 917 (1997), ans references therein.

[3] T. E. Dialynas and G. P. Tsironis, Phys. Lett. A **218**, 292 (1996).

[4] I. Derényi and T. Vicsek, Proc. Natl. Acad. Sci. USA **93**, 6775-6779 (1996).

[5] A. Savin, G. P. Tsironis and A. V. Zolotaryuk, Phys. Lett. A, **229**, 279 (1997).

[6] K. Svoboda et al., Direct observation of kinesin stepping by optical trapping interferometry, Nature **365** 721-727 (1993).

[7] K. Svoboda and S. M. Block, Force and velocity measured for single kinesin molecules, Cell **77**, 773-784 (1994).

[8] K. Svoboda, P. P. Mitra and S. M. Block, Fluctuation analysis of motor protein movement and single enzyme kinetics, Proc. Natl. Acad. Sci. USA **91**, 11782-11786 (1994).

[9] R. D. Vale et al., Direct observation of single kinesin molecules moving along microtubules, Nature **380**, 451-453 (1996).

[10] C. M. Coppin et al., Detection of sub-8-nm movements of kinesin by high-resolution optical-trap microscopy, Proc. Natl. Acad. Sci. USA **93**, 1913-1917 (1996).

[11] J. Howard, The movement of kinesin along microtubules, An. Rev. Physiol. **58**, 703-729 (1996).

[12] R. B. Case, D. W. Pierce, N. Hom-Booher, C. L. Hart and R. D. Vale, The directional preference of kinesin motors is specified by an element outside of the motor catalytic domain, Cell **90**, 959 (1997).

[13] U. Henningsen and M. Schliwa, Reversal in the direction of movement of a molecular motor, Nature **389**, 93-95 (1997).

[14] W. Hua et al., Coupling of kinesin steps to ATP hydrolysis, Nature **388**, 390-393 (1997).

[15] M. J. Schnitzer and S. M. Block, Kinesin hydrolyses one ATP per 8-nm step, Nature **388**, 386-390 (1997).

[16] H. Sosa et al., A model for the Microtubule-Ncd Motor Protein Complex Obtained by Cryo-Electron Microscopy and Image Analysis, Cell **90**, 217-224 (1997).

[17] B. Tipet, R. D. Vale and R. S. Hodges, Demonstration of coiled-coil interactions within the kinesin neck region using synthetic peptides, J. Biol. Chem. **272**, 8946-8956 (1997).

[18] C. Tucker and L. S. B. Goldstein, Probing the kinesin-microtubule interaction, J. Biol. Chem. **272**, 9481-9488 (1997).

[19] T. E. Dialynas, K. Lindenberg and G. P. Tsironis, Phys. Rev. E **56**, 3976-3985 (1997).

[20] G. P. Tsironis and Katja Lindenberg, Motor proteins have highly correlated Brownian engines, eds. S.K. Malhotra and J.A. Tuszynski. in *Advances in Structural Biology* vol. 5, Academic Press, (Stamford 1998).

[21] K. Lindenberg and B. J. West, *The Nonequilibrium Statistical Mechanics of Open and Closed Systems* (VCH Publishers, New York, 1990).

[22] G. Stratopoulos, T. Dialynas and G. P. Tsironis, Directional Newtonian Motion and Reversals of Molecular Motors, Phys. Lett. A **252**, 151 (1999).

[23] H. Frauenfelder, P. G. Wolynes and R. H. Austin, Biological Physics, Rev. Mod. Phys. 71, S419 (1999).

[24] A. Aubry and G. P. Tsironis, unpublished.

19
Nonlinear Dynamics of DNA

L.V. Yakushevich

ABSTRACT The main features of nonlinear dynamics of DNA as a new field of nonlinear science named nonlinear biomolecular dynamics are briefly described

1 Introduction

Nonlinear dynamics of DNA is one of the branches of a new and rapidly developing field of nonlinear science named nonlinear biomolecular dynamics. Only a few reviews on the theme can be pointed out, which were written by Scott [1], Zhou and Zhang [2], Yakushevich [3], and Gaeta and co-authors [4]. A collection of lectures made by participants of the International workshop in Les Houches (France, 1994) [5], selected paragraphs in the monograph of Davydov [6] and the monographs of Yakushevich [7],[8] should be also mentioned.

Nonlinear biomolecular dynamics can be determined as a part of the general dynamics that deals with internal mobility of biomolecules. Theoreticians define it also as the next (anharmonic or nonlinear) approximation after the first (harmonic or linear) one. In contrast to the linear approximation a nonlinear approximation is used when the amplitudes of internal motions in biomolecules are large. Conformational transitions, denaturation processes, formation of opening states in the processes of DNA-protein recognition are some of the better known examples of the large-amplitude motions.

Three events can be considered as having stimulated the appearance and rapid development of the nonlinear DNA dynamics. The first was the success of nonlinear mathematics and its applications to many physical phenomena [9]-[11]. Second were the new results in studies of the dynamics of biopolymers leading to understanding the important role of the dynamics in the biological functioning of biopolymers [12]-[14]. The third event was associated with the publishing of the series of works of Davydov, where for the first time the achievements of nonlinear mathematics were applied to biomolecules and the hypothesis of the occurrence of solitons in biopolymers (namely, in alpha-helical proteins) were suggested [15].

The nonlinear dynamics of DNA was started in 1980 when the article

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 373−391, 2000.

of five American authors: Englander, Kallenbach, Heeger, Krumhansl and Litwin was published [16]. The article was titled "Nature of the open state in long polynucleotide double helices: possibility of soliton excitations", and it was the first time that the nonlinear conformational excitations (or DNA solitons) imitating the local opening of base pairs has been introduced. In the article the first nonlinear Hamiltonian of DNA was presented, and this gave a powerful impulse for theoretical investigations. A large group of authors, including Yomosa [17]-[18], Takeno and Homma [19]- [20], Krumhansl and co-authors [21]-[22], Fedyanin and co-authors [23]-[25], Yakushevich [26]-[28], Zhang [29], Prohofsky [30], Muto and co-authors [31]-[33], van Zandt [34], Peyrard [35]-[36], Dauxois [37], Gaeta [38]-[39], Salerno [40], Bogolubskaya and Bogolubsky [41], Hai [42], Gonzalez and Martin-Landrove [43] made contributions to the development of this field by improving the model Hamiltonian and its dynamical parameters, by investigating corresponding nonlinear differential equations and their soliton-like solutions, by consideration of DNA solitons and calculation of corresponding correlation functions. The results obtained formed a theoretical basis of the nonlinear DNA dynamics.

An experimental basis of nonlinear DNA dynamics was formed by the results of experimental investigations on the DNA dynamics and interpretations some of them in the frameworks of the nonlinear concept. The most important results were obtained by Englander and co-authors [16] on hydrogen-tritium exchange in DNA, by Webb and Booth [44], Swicord and co-authors [45]-[47] on resonant microwave absorption (interpretations were made by Muto and co-authors [31] and by Zhang [48]), and by Baverstock and Cundall [49] on neutron scattering by DNA. All these results, however, admitted alternative interpretations (see the discussion in [3]), and only after publication of the work of Selvin and co-authors [50], where the torsional rigidity of positively and negatively supercoiled DNA was measured, reliable experimental basis for theoretical predictions was given.

In this article we describe briefly the main features of nonlinear DNA dynamics. We discuss the following questions:

- General description of DNA dynamics. Classification of the internal motions.
- Mathematical modeling of DNA dynamics. Hierarchy of the models.
- Nonlinear mathematical models. Solved and unsolved problems.
- Nonlinear DNA models and experiment. Resonant microwave absorption.
- Nonlinear conception and mechanisms of DNA functioning.

2 General description of DNA dynamics. Classification of the internal motions

An important feature of the DNA structure having the form of double helix [60]-[61] is that it is not rigid. On the contrary, the DNA molecule has a rather moveable structure. The DNA molecule is usually immersed in some thermal bath, and its structural elements such as individual atoms, groups of atoms (bases, sugar rings, phosphates), fragments of the double chain including several base pairs are in constant movement. The thermal bath is not the only source of the DNA internal mobility. Collisions with the molecules of the solution that surrounds DNA, local interactions with proteins, drugs or with some other ligands also lead to internal mobility. In this section we describe the main features of the DNA molecule as a dynamical system.

Because of the structure of a DNA molecule is rather complex, one can expect that the general picture of its internal mobility is also complex. In the first approximation, however, it can be described by a few simple characteristics: the time-scale, the amplitudes of the internal motions and the energies or frequencies associated with these motions. So, we can state that

i) the dynamical events in DNA occur on the time-scale ranging from femtoseconds to at least seconds;

ii) the amplitudes of the internal motions can be both small (for example, small vibrations of individual atoms or atomic groups near their equilibrium positions) and large (for example, large amplitude motions of the fragments of polynucleotide chains associated with the local unwinding or opening of the double helix);

iii) the frequencies associated with the internal motions are much lower than the frequencies of more internal vibrations in isolated small molecules.

To describe the picture of the internal DNA mobility in more detail, it is convenient to classify DNA internal motions according to their forms (types), energies, amplitudes and characteristic times. Some of the classifications have been proposed in the works of Fritzsche [89], Keepers and James [90], McClure [91], McCammon and Harvey [13], and Yakushevich [3], [53]. For example, classifications proposed in [53] are based on the time characteristics of internal DNA motions. The time-scale is divided there into several ranges, and for each range, the main types of internal motions, the main structural elements involved in these motions, the energies of activation and the amplitudes of the motions are described.

Investigators who are interested only in the part of the general picture that is assumed to be connected with the DNA functioning, usually restrict ourselves by consideration of the internal motions that belong to the

nanosecond range and its neighborhood. This range includes beside others the so-called solid-like motions of sugars, phosphates and bases, which are very important in many biophysical phenomena: conformational transitions, gene regulation, DNA-protein recognition, energy transmission, DNA denaturation, among others.

3 Mathematical modeling of DNA dynamics. Hierarchy of the models

3.1 Principles of modeling

Because of the complexity of the general picture of DNA internal dynamics, mathematical modeling of the dynamics is also rather complex. It requires detailed information about physical parameters such as coordinates, mass and moments of inertia of structural elements, and about interactions between the elements. The problem, however, can be simplified if we construct approximate models that imitate only the internal motions which make the main contribution to biological processes. This approach is widely used in studying DNA dynamics. Let us describe briefly the main principles of constructing the models in this way.

To construct an approximate model, first of all it is necessary to simplify the general picture of the DNA internal motion. This can be done by selecting of a limited set of internal motions that are dominant. This selection can be done in many ways, and this explains a large variety of the models proposed. Secondly, we need to describe these motions by mathematical equations. This can be done directly or through intermediate stage consisting in finding some physical (very often mechanical) analog with the same type of internal motions and interactions. Thirdly, we need to solve the equations and to interpret their solutions in terms of the parameters of the DNA internal dynamics.

If we do not plan to construct a new dynamical model and want only to choose an appropriate model among those proposed earlier by some other authors, it is convenient to use a special approach where each of the models is considered as an element of a hierarchy. This approach automatically gives us information about possible restrictions of the models used and about the relations between the models. Let us describe briefly the hierarchy of structural models of DNA, and discuss how it can be used to construct the hierarchy of the dynamical models.

3.2 Structural hierarchy

To construct a hierarchy of structural models, it is convenient to arrange all known structural models of DNA in order of increasing complexity. In

this case the models look like the elements of a hierarchy. Let us describe the main possible levels of the hierarchy.

The first level of the hierarchies formed by the simplest structural model of DNA, which is prompted by microphotos of the molecule where the DNA molecule looks like a thin elastic filament. So it can be suggested that a uniform elastic rod with a circular section can be considered as the simplest structural model of a fragment of DNA. The discrete analog of the rod-like model consists of a chain of coupled discs, each disc imitating a very small piece of the DNA molecule, which contains one base pair.

The second level of the hierarchy is formed by some more complex structural models of the DNA that take into account that the molecule consists of two polynucleotide chains interacting with one another by hydrogen bonds and being wound around each other to produce the double helix. In this approximation, the internal structure of the chains is neglected, and each of the chains is simulated by an elastic uniform rod (in the continuous case) or by a chain of coupled discs (in the discrete case). So, the complete model consists of two elastic rods (or two chains of coupled discs) weakly interacting with one another and being wound around each other. In the discrete case each of the discs imitates a very small peace of one of two polynucleotide chains, which contains only one base. To reduce calculations, a more simple version of the models described above is widely used, consisting of two straight uniform elastic rods weakly interacting with one another. And the discrete analog of the model has the form of two straight chains with discs connected with one another by longitudinal and transverse springs.

The third level of the hierarchy is formed by a group of structural models that take into account additional details of the internal DNA structure. Every polynucleotide chain is considered here as that consisting of mutually rigidly bound atomic subgroups: the bases, the sugar rings and the phosphate-carbon pieces, with relatively weak, flexible bonds connecting them with each other.

The fourth level of the hierarchy contains the so-called lattice models where a finite group of atoms (named nucleotide) forms a unit cell quasiperiodically repeating along the DNA molecule.

The fifth level of the hierarchy comprises the most accurate structural models taking into account the positions of every atoms of the molecule.

3.3 Dynamical hierarchy

The hierarchy of dynamical models can be easily constructed by the following way. Let us assume that the structural models described in the previous section are not static but dynamical. That is, all structural elements of the models are movable. Because the models have been arranged already in the order of increasing their complexity, we obtain automatically a hierarchy of the dynamical models that is briefly described.

The first level of the hierarchy is formed by the rod-like model (and by its discrete analog) having three types of internal motions: longitudinal and rotational motions of small elements of the rod bounded by adjacent cross sections, and bending motions.

The second level is formed by the model consisting of two elastic rods (or by their discrete analogs) weakly interacting with one another. The model has six types of internal motions: longitudinal, transverse and rotational motions in both rods.

The models of the third level take into account that each of the DNA strands consists of three types of atomic groups: sugars, phosphates and bases, and imitate their motions as solid like motions of the atomic groups weakly interacting with each other. To calculate how many types of internal motions are described in this model, we should take into account that in the extreme case when the connecting bonds between the groups are absent every group has six degrees of freedom, and that every chain has three types of groups. Thus, we obtain 36 degrees of freedom (instead of three for the first level models and six for the second level models). However, if we consider the connecting bonds, the number of degrees of freedom will be decreased.

The models of the fourth level describe internal motions in a lattice with a unit cell formed by a finite group of atoms (nucleotide) that periodically repeats along the DNA. In this approximation we consider all displacements of the nucleotide atoms but restrict ourselves by consideration of only homopolymer chains.

Finally, the fifth level is formed by the most accurate models of the DNA molecule where all motions of all atoms are taken into account.

In conclusion, let us consider one example illustrating how the hierarchy of the dynamical models can be applied. Assume, that we study the dynamical aspects of the process of local opening the double helix. Which of the model should be chosen? We could begin with the simplest models of the first level, but these models are not appropriate because they do not take into account the DNA internal structure at all. The second level models are more appropriate, and they can be used as the first approximation. The third level models are more accurate and their application permits one to describe the process in more detail. Their application can be consider as the second approximation, and so on.

4 Nonlinear mathematical models. Solved and unsolved problems

4.1 Ideal models

All internal motions mentioned above can be described by differential equations. If we do not restrict ourselves by consideration of only small amplitude motions, these equations are nonlinear. So, in the continuum approximation we can write three nonlinear coupled differential equations for the models of the first level, six equations for the models of the second level or more equations for the third or higher level models, the equations being arranged in the order of increasing complexity, thus forming a hierarchy.

As an example, let us discuss in more detail mathematical models of the second level. These models are very important because as was mentioned above the models of the second level can be appropriate for describing the formation and dynamics of open states in DNA. In thegeneral case the Hamiltonian of the model of second level consists of six terms. Three of them describe longitudinal, transverse and rotational subsystems and the other three describe interactions between them. In other words in the continuum approximation the model can be described by a system of six coupled nonlinear differential equations. No one has yet tried to investigate the system as a whole because of its complexity, but some particular cases were studied intensively.

One of the most interesting and promising particular case has been proposed by Peyrard and Bishop [35] . Their model takes into account only the transverse internal motions; thus it consists of two (instead of six) equations describing transverse displacements in two DNA strands. It was shown that the model equations has soliton-like solutions which have been successfully applied to explain the mechanism of DNA denaturation.

Another particular case has been considered in the works of Yomosa [17], [18] and other authors [16], [19]-[20], [23]-[24], [27], [29]. Their model takes into account rotational degrees of freedom. The model consists of two (instead of six) equations describing rotational displacements in DNA strands. It was shown that among solutions of the model equations there are soliton-like solutions, which successfully have applied to explain the mechanisms of long-range effects in DNA and to solve the problem of direction of the process of transcription.

There are also a few particular cases where investigators tried to take into account interactions between different types of motions. So, in the work of Muto et al. [33] transverse and longitudinal internal motions and their interactions were taken into account. In the work of Zhang [48] longitudinal and rotational motions and interactions between them were taken into account. We could add also to this list the works of Xiao et al. [51] and Volkov [52].

Concerning the whole problem consisting of six nonlinear differential

equations, it still remains unsolved. Moreover there are some difficulties even in finding analytical form for the terms describing the interactions. One of the approaches giving a possibility to construct these terms has been proposed recently in our work [53], and it is based on the representation of the model Hamiltonian in terms of vectors of displacements of the DNA structural units.

4.2 Nonideal models

Till now we discussed ideal DNA models where the effects of environment and inhomogeneity were not taken into account. These effects can be omitted in the first approximation of the theory, but they become important when we try to apply theoretical results to explain experimental data on the dynamics and functioning.

In the general case the modeling of DNA-environment interaction is a rather complex problem, but it can be reduced in the first approximation to two effects: the effect of dissipation and the effect of an external field. There are many approaches to modeling these two effects and some of them were discussed in [26]. In most of the approaches, investigators assume that the DNA-environment interaction leads to small perturbations of the solutions of the ideal model dynamical equations and use a linear perturbation technique to solve the equations.

Very similar approaches are used to solve the problem of nonlinear dynamics of inhomogeneous DNA. To receive inhomogeneous model equations investigators usually suggest that the coefficients of nonlinear differential equations are not constant but functions depending on a variable z, where the z axis is parallel to the DNA axis. So, there is no problem in obtaining inhomogeneous nonlinear equations, but there is a problem in solving them. If we suggest, however, that perturbations due to inhomogeneities are small, a linear perturbation technique to solve the equations can be used [26]. In the other cases, only computer simulations of the dynamics of the nonlinear inhomogeneous DNA system are possible [40].

4.3 Statistics of solitons in DNA

In the dynamical models discussed above it was assumed that only one nonlinear excitation (soliton) was present, so the possibility of exciting two or more nonlinear excitations, their collisions and interactions were not considered. But DNA is a rather long molecule, so one can expect that several nonlinear excitations can be excited simultaneously. In this case, we should consider an ensemble of solitons and discuss their statistics. The latter becomes very important when we try to interpret experimental data on scattering (neutrons or light) by DNA or the data on DNA denaturation.

Different approaches to consideration of statistics of solitons in DNA are possible. One of them was developed by Peyrard and co-authors [35], [54].

It is based on the method of transfer operators suggested by Krumhansl and Schrieffer [55]. This method gives a possibility to calculate classical partition function, free energy, specific heat and other characteristics of the nonlinear DNA system. The results of these calculations were successfully applied to the problem of DNA denaturation.

Another method was proposed by Fedyanin and Yakushevich [7], [23], which is based on the similarity between the dynamical properties of solitons and the dynamical properties of ordinary classical particles. This similarity has been demonstrated for many nonlinear systems, and in [7], [23] it was shown for DNA. According to the results obtained there, we can ascribe mass m, velocity v and energy E to DNA soliton, and consider instead of ensemble of solitons an ensemble of ordinary classical particle with the same dynamical characteristics. To simplify calculations it is often suggested that the number of solitons is not large and that ensemble of particles can be described as an ideal gas. With these assumptions, it is not difficult to calculate different macroscopic characteristics of the system such as large statistical sum, thermodynamical potential, correlation particles, density of the particles an others. This approach has been applied to calculations of the dynamical form-factor of slow neutron scattering by DNA solitons [23]. Predictions made in this application can be checked experimentally.

5 Nonlinear DNA models and experiment

Let us discuss now experimental methods of studying DNA dynamics and the data that are interpreted in terms of nonlinear theory. To be impartial, we present here both the arguments in favor of interpretation of experimental data in the frameworks of nonlinear conception and the arguments against. In addition, we describe new approaches in experimental studying nonlinear DNA properties, which might be able to solve the contradiction between these two positions.

5.1 Hydrogen-tritium (or hydrogen-deuterium) exchange

Method of hydrogen-tritium (or hydrogen-deuterium) exchange is widely used to study internal DNA dynamics [56]-[58]. The method is especially effective in studying the dynamics of open states.

Indeed, from the analysis of the data on hydrogen-tritium exchange, Englander and co-authors [16] came to the conclusion that open states with low energies and slow opening and closing rates, can be interpreted as structural deformations formed by several adjacent unpaired base pairs. They assumed a mobile character of the deformations, that is a capability to diffuse along the double helix. It was suggested also that the movement

of the deformations can be modeled mathematically as a propagation of solitary waves which are the kink- and antikink-like solutions.

These suggestions were widely discussed and even criticized by some theoreticians and experimenters. The criticism of theoreticians was directed to the oversimplicity of model. As a result of the discussion, many improvements of the initial model have been proposed [17]-[27].

The criticism of experimenters, and especially, the arguments of Frank-Kamenetskii [59]-[60], which were based on the statement that the value of probability of base-pair opening that was taken from the data of Mandal and co-authors [58] and used later by Englander and co-authors [16], are not correct. We should state that the question remains open, and some other additional experimental investigations are required to clarify it.

5.2 Resonant microwave absorption

One more useful tool in the research of the nonlinear properties of biomolecules is the studies of interaction of the molecule with microwaves. The most impressive example of application of the method is the history of investigations of resonant microwave absorption (in the range of several gigahertz) by aqueous solutions containing DNA.

Experimental evidence of resonant microwave absorption in DNA was received at first by Webb and Booth [44] and later by Swicord and co-authors [45]- [46]. Although their results are still controversial [61]-[64], they stimulated theoreticians to study the problem. As a result, many different approaches has been proposed. Some of them were based on the linear (harmonic) model [65]-[67]; others were based on the nonlinear conception [31], [48], [68].

The first nonlinear mathematical model of interaction of DNA with the external microwave field was proposed by Muto and co-authors [31], [68]. As the basis they used the ideal rod-like model of internal DNA dynamics, which describes longitudinal displacements in DNA. To imitate conditions of the microwave experiment, Muto et al. added two additional terms imitating the effects of dissipation and the effects of interaction with the external (microwave) field. The resulting equation had the form of the Ostrovskii-Sutin equation. Using a special numerical procedure Muto et al. [31] calculated the absorption spectrum and compared it with the spectrum calculated earlier by Van Zandt [66] in the framework of the linear approximation. The results obtained were in close agreement only for the fourth peak (the third overtone). The most marked differences were observed for the first (fundamental) peak, namely, in the nonlinear approximation, the resonance peak exhibited a multicomponent character (fine structure) and the linewidth of the peak was not simply related to the damping constant as in the case of the linear approximation. The difference between the spectra can be explained by the presence of the nonlinear term in the model equation.

The approach of Muto et al. was improved by Zhang [48]. He considered a rod-like model of DNA (as Muto et al. did), but took into account both longitudinal and torsional degrees of freedom. As a result, instead of one model equation he obtained two coupled nonlinear equations. To consider the microwave absorption by aqueous solutions containing DNA, Zhang added four terms (two damping terms and two driving terms) and solved the resulting system of two nonlinear coupled equations by the method of perturbation. As a result, he obtained that i) the resonant absorption of microwave energy is possible for both longitudinal and torsional modes, ii) the resonance frequencies are in the region of gigahertz and subterahertz, iii) for both modes the so-called subharmonic resonances are possible. However, these theoretical predictions have not been checked yet by experimenters.

5.3 Scattering of neutrons and light

There are a few attempts to explain the data on neutron scattering by DNA in terms of solitons. We describe here two of them. One was made in the works of Fedyanin and Yakushevich [3], [7], [23] and the other in the works of Baverstock and Cundall [49], [69]-[70].

In the works of Fedyanin and Yakushevich [3],[7], [23] some results on calculations of the dynamical form-factor of scattering thermal (slow) neutrons by DNA solitons were obtained. Calculations made by these authors on the basis of simple sine-Gordon model predict the existence of the central peak in the spectrum of scattering. The parameters of the peak (integral intensity and width) depend on the temperature and the wave vector which is a difference between the final and initial wave vectors of neutrons. The behavior of the parameters of the central peak predicted by theory could be checked experimentally.

We should note, however, that the results described above are changed if instead of a simple sine-Gordon model we take the helical version of the sine-Gordon model, proposed in [23]. The result of calculations of the dynamical form-factor can be interpreted as the splitting of the central peak into two components. This prediction could be checked experimentally.

Baverstock and Cundall [49], [69]-[70] used the soliton idea to interpret the experimental data on scattering of fast neutrons by DNA. We should note that interaction of fast (high energy) neutrons with DNA differs substantially from the interaction of the thermal (low energy and slow) neutrons described above. The main difference is that the fast neutron scattering is accompanied by the formations of radical anions. The studying of the yield of radicals is one of the powerful methods permitting one to obtain information about DNA internal dynamics. In 1986 Arroyo and coauthors [71] investigated the radical yield dependence on the direction of irradiation of oriented fibers of DNA. Their results showed that the radical anions of thymine were formed in roughly equal amounts to the guanine anions when the neutron flux was perpendicular to the axis of the DNA

molecules. When the flux was parallel to the DNA axes a protonated form of a thymine anions dominated, and overall radical yields were lower by a factor of three in the parallel sample. These results were interpreted by Miller et al. [72], who assumed a large asymmetry in the thermal conductivity of oriented films. With the help of a track structure model for protons, it was shown that irradiation by the protons parallel to the axes of the DNA molecule results both in the formation of the thymine anion and, from the same particle, further energy deposits of the vibrational excitations. Some of these may migrate to the sites of thymine anions, resulting in sufficient thermal stimulation of the thymine among donor interaction to promote protonation. Where the proton direction is perpendicular to the axes of the DNA molecules, such migration will be impeded by the low intra-molecular thermal conductivity. This idea was developed by Baverstock and Cundal [69]-[70]. Taking into account that the deposition of ionizing energy into condensed media is a highly nonlinear process, they suggested that it can give rise to soliton-like species that enables energy to be transferred without loss over long distances. Presently, there is not a possibility to check this interpretation experimentally.

5.4 Fluorescence depolarization

The method of fluorescence depolarization is widely used for measurements of the torsional constants of biopolymers. In 1992, Selvin et al. [50] used this method for measurements of the torsional rigidity of positively, relaxed and negatively supercoiled DNA. For the purpose they used Time Correlated Single Photon Counting (TCSPC) of intercalated Ethidium Bromide. The measurements were made in the wide range of superhelical density with a time-resolution of 75 picoseconds extending from 0.75 nanosecond, the range in which DNA twisting motions dominate the fluorescence depolarization signal.

The main result of the measurements was rather unusual: the torsional rigidity of the DNA molecule was not a constant, as it had been suggested before. Thus the DNA molecule cannot be considered as a linear system. Selvin et al. suggested that a more accurate mathematical model of the internal DNA dynamics should consist of coupled nonlinear torsional springs. According to their estimations, the anharmonic term in the model Hamiltonian should be approximately 15% for twist fluctuations at room temperature. These results provide rather reliable evidence of the nonlinear nature of the internal DNA dynamics.

6 Nonlinear conception and mechanisms of DNA functioning

An important and traditional problem of DNA science is to relate the DNA functional properties with its structural and dynamical properties. We shall describe here several examples where the mechanisms of conformational transitions, of long-range effects, of regulation of transcription process and of DNA denaturation are explained in terms of nonlinear dynamics.

6.1 Nonlinear mechanism of conformational transitions

An interesting application of the nonlinear theory is associated with the interpretation of the mechanism of transitions between different conformational forms of a DNA molecule. For the first time the relation between the phenomenon and the nonlinear theory was noticed and reported by Krumhansl in 1982 at the workshop in Gysinge [73], after which this approach was developed by many other authors [21]-[22], [29], [74]-[78].

As an example, let us consider the transition between A- and B-forms of the DNA molecule. The transition can occur due to the change of temperature, of pH, of hydration or of some other parameters. It is easily visualized by X-ray diffraction studies of DNA fibers: if the fibers, for example, are allowed to dry, they produce an A-type diffraction pattern, and if the fibers remain hydrated, the pattern is a B-type. So, the process of transition can be considered as a movement of the boundary between two (A and B) ranges. This movement is very similar to the movement of a boundary between two different phases in physical systems. It is well known, however, that in physics the transition processes are successfully described by kink and antikink solutions of corresponding nonlinear dynamical equations; thus one can expect that the movement of the boundary between two DNA regions are also described by soliton-like solutions of kink (or antikink) type. This suggestion was confirmed by theoretical results obtained in [76] where nonlinear equations describing A-B transitions were derived. One of the exact solutions of the equations had the form of kink, and this solution was interpreted as a moving boundary between two regions one of them having the A-form and the other having the B-form.

6.2 Nonlinear conformational waves and long-range effects

During the 70's and 80's a great deal of experimental work was done on long-range effects in DNA [79]-[83]. These effects can be described in the following way. Let us consider a system containing two protein molecules and one DNA molecule having two sites: 1 and 2. It is assumed that these proteins specifically interacts with the sites, namely, the first protein molecule can bind to site 1 and the other protein molecule can bind with the other

site: site 2. The effect is that the binding of the first protein with site 1 influences the binding of the second protein molecule with site 2 even when the distance between the sites reaches hundreds or thousands of base pairs.

Among different explanations of the effect, there is one which is of most interest. According to it, the effect of the binding of the first protein molecule to site 1 is accompanied by a local distortion of the DNA conformation, which can propagate along the double DNA chain. When reaching site 2 it changes the conformational structure of the site, which in turn changes the binding constants of the second protein with the site.

This mechanisms can be easily interpreted in terms of nonlinear dynamics. Indeed, formation of the local distortion can be interpreted as an excitation of a nonlinear wave (or soliton) and propagation of the distortion along the double helix as transmission of the nonlinear wave along the DNA. To model the long-range effect more correctly we should know many more details on the DNA-protein interaction and try to construct mathematical models that take them into account.

This approach was developed in [84]-[85] to give an interpretation of different stages of the transcription process. Moreover a more general assumption was made, namely, the nonlinear solitary waves moving along the DNA molecule could be a suitable model for the regulation and coordination of simultaneous transcription of many genes.

6.3 Direction of transcription process

From numerous biological experiments it is well known that the directions of transcription processes are different not only for DNA molecules of different living organisms but also for different promoter regions of the same DNA molecule. For some of the promoters the transcription process preferably develops in the up-stream direction, for some other promoters, in the down-stream direction. There are also promoters that do not have a preferable direction of transcription.

An explanation of these phenomena has been given by Salerno [40], [86]-[88], who used a discrete variant of the nonlinear model of the second level that took into account only rotational degrees of freedom. Corresponding model equations were studied numerically, and a kink-like solution of the continuum version of the equations was used as an initial condition. It was found that an initially static soliton can (i) stay without moving or (ii) oscillate or (iii) move along the DNA molecule in one of two possible directions: up-stream or down-stream. In the case (iii) when soliton moves the problem of the direction of the movement (or the problem of direction of transcription) arises. Salerno suggested that the direction depends on the sequence of bases near the starting point. To check the suggestion, he took as an example the real sequence of bases corresponding to T7A1 DNA promoter, and he found that in the case when initially soliton was placed outside the promoter the soliton remained static. When the soliton was

placed inside the promoter region, the solitary wave began to move to the left end of the chain. This result is in a good agreement with known data on the functional properties of T7A1 promoter. Moreover, this result shows the existence of dynamically active region inside the promoter region.

Thus this example shows that the approach of Salerno gives us an effective tool for indication of dynamically active regions in DNA. Moreover we could expect that these regions correspond to functionally active regions. If this suggestion is confirmed, scientists will have another method of analyzing and interpreting the DNA code.

7 References

[1] Scott, A.C., Solitons in biological molecules. Comments Mol. Cell. Biol. 3, 5-57, 1985.

[2] Zhou G.-F., Zhang Ch.-T., A short review on the nonlinear motion in DNA. Phys. Scripta 43, 347-352, 1991.

[3] Yakushevich L.V., Nonlinear dynamics of biopolymers: theoretical models, experimental data. Quart. Rev. Biophys. 26, 201-223, 1993.

[4] Gaeta G., Reiss C., Peyrard M., Dauxois T., Simple models of nonlinear DNA dynamics. Rev. Nuovo Cimento 17, 1-48, 1994.

[5] *Nonlinear Excitations in Biomolecules*. Ed. M. Peyrard. Springer, Berlin, 1995.

[6] Davydov A.S., *Solitons in Bioenergetics*. Naukova Dumka,

[7] Yakushevich L.V., *Methods of Theoretical Physics and Their Applications to Biopolymer Sciences*. Nova Science Publishers, New York, 1996.

[8] Yakushevich L.V., *Nonlinear Physics of DNA*, John Wiley and Sons, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto, 1998

[9] Scott A.C., Chu F.Y., McLaughlin D.W., The soliton: a new concept in applied science. Proc. IEEE 61, 1443-1483, 1973.

[10] *Solitons in Action*. Eds. K. Lonngren, A. Scott. Academic Press, New York, 1978.

[11] *Solitons*. Eds. R.K. Bullough, P.J. Caudrey. Springer-Verlag, Berlin, 1980.

[12] *Structure and Dynamics: Nucleic Acids and Proteins*. Eds. E. Clementi, R.H. Sarma. Adenine Press, New York, 1993.

[13] *Structure and Motion: Membranes, Nucleic Acids and Proteins*. Eds. E. Clementi, G. Corongiu, M.H. Sarma, R.H. Sarma. Adenine Press, New York, 1985.

[14] McCommon J.A., Harvey S.C., *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, 1987.

[15] Davydov A.S., Solitons in molecular systems. Physica Scripta 20, 387-394, 1979.

[16] Englander S.W., Kallenbach N.R., Heeger A.J., Krumhansl J.A., Litwin A., Nature of the open state in long polynucleotide double helices: possibility of soliton excitations. Proc. Natl. Acad. Sci. USA 77, 7222-7226, 1980.

[17] Yomosa S., Soliton excitations in deoxyribonucleic acid (DNA) double helices. Phys. Rev. A 27, 2120-2125, 1983.

[18] Yomosa S., Solitary excitations in deoxyribonucleic acid (DNA) double helices. Phys. Rev. A 30, 474-480, 1984.

[19] Takeno S., Homma S., Topological solitons and modulated structure of bases in DNA double helices. Prog. Theor. Phys. 70, 308-311, 1983.

[20] Homma S., Takeno S., A coupled base-rotator model for structure and dynamics of DNA. Prog. Theor. Phys. 72, 679-693, 1984.

[21] Krumhansl J.A., Alexander D.M., Nonlinear dynamics and conformational excitations in biomolecular materials. In: *Structure and Dynamics: Nucleic Acids and Proteins*. Eds. E. Clementi, R.H. Sarma. Adenine Press, New York, 1983, pp. 61-80.

[22] Krumhansl J.A., Wysin G.M., Alexander D.M., Garcia A., Lomdahl P.S., Layne S.P., Further theoretical studies of nonlinear conformational motions in double-helix DNA. In: *Structure and Motion: Membranes, Nucleic Acids and Proteins*. Eds. E. Clementi, G. Corongiu, M.H. Sarma, R.H. Sarma. Adenine Press, New York, 1985, pp. 407-415.

[23] Fedyanin V.K., Yakushevich L.V., Scattering of neutrons and light by DNA solitons. Stud. biophys. 103, 171-178, 1984.

[24] Fedyanin V.K., Gochev I., Lisy V., Nonlinear dynamics of bases in continual model of DNA double helices. Stud. biophys. 116, 59-64, 1986.

[25] Fedyanin V.K., Lisy V., Soliton conformational excitations in DNA. Stud. biophys. 116, 65-71, 1986.

[26] Yakushevich L.V., The effects of damping, external fields and inhomogeneity on the nonlinear dynamics of biopolymers. Stud. biophys. 121, 201-207, 1987.

[27] Yakushevich L.V., Nonlinear DNA dynamics: a new model. Phys. Lett. A 136, 413-417, 1989.

[28] Yakushevich L.V., Investigation of a system of nonlinear equations simulating DNA torsional dynamics. Stud. biophys. 140, 163-170, 1991.

[29] Zhang Ch.-T., Soliton excitations in deoxyribonucleic acid (DNA) double helices. Phys. Rev. A 35, 886-891, 1987.

[30] Prohofsky E.W., Solitons hiding in DNA and their possible significance in RNA transcription. Phys. Rev. A 38, 1538-1541, 1988.

[31] Muto V., Holding J., Christiansen P.L., Scott A.C., Solitons in DNA. J. Biomol. Struct. Dyn. 5, 873-894, 1988.

[32] Muto V., Scott A.S., Christiansen P.L., Thermally generated solitons in a Toda lattice model of DNA. Phys. Lett. A 136, 33-36, 1989.

[33] Muto V., Lomdahl P.S., Christiansen P.L., Two-dimensional discrete model for DNA dynamics: longitudinal wave propagation and denaturation. Phys. Rev. A 42, 7452- 7458, 1990.

[34] Van Zandt L.L., DNA soliton realistic parameters. Phys. Rev. A 40, 6134-6137, 1989.

[35] Peyrard M., Bishop A.R., Statistical mechanics of a nonlinear model for DNA denaturation. Phys. Rev. Lett. 62, 2755-2758, 1989.

[36] Dauxois T., Peyrard M., Willis C.R., Localized breather-like solutions in a discrete Klein- Gordon model and application to DNA. Phys. D 57, 267-282, 1992.

[37] Dauxois T., Dynamics of breathers modes in a nonlinear helicoidal model of DNA. Phys. Lett. A 159, 390-395, 1991.

[38] Gaeta G., On a model of DNA torsion dynamics. Phys. Lett. A 143, 227-232, 1990.

[39] Gaeta G., Solitons in planar and helicoidal Yakushevich model of DNA dynamics. Phys. Lett. A 168, 383-389, 1992.

[40] Salerno M., Discrete model for DNA-promotor dynamics. Phys. Rev. A 44, 5292-5297, 1991.

[41] Bogolubskaya A.A., Bogolubsky I.L., Two-component localized solutions in a nonlinear DNA model. Phys. Lett. A 192, 239-246, 1994.

[42] Hai W., Kink couples in deoxyribonucleic acid (DNA) double helices. Phys. Lett. A 186, 309-316, 1994.

[43] Gonzalez J.A., Martin-Landrove M., Solitons in a nonlinear DNA model. Phys. Lett. A 191, 409-415, 1994.

[44] Webb S.J., Booth A.D., Absorption of microwave by microorganisms. Nature 222, 1199-1200, 1969.

[45] Swicord M.L., Davis C.C., Microwave absorption of DNA between 8 and 12 GHz. Biopolymers 21, 2453-2460, 1982.

[46] Swicord M.L., Davis C.C., An optical method of investigating the microwave absorption characteristics of DNA and other biomolecules in solution. Bioelectromagnetics 4, 21-42, 1983.

[47] Edwards G.S., Davis C.C., Saffer J.D., Swicord M.L., Resonant absorption of selected DNA molecules. Phys. Rev. Lett. 53, 1284-1287, 1984.

[48] Zhang Ch.T., Harmonic and subharmonic resonances of microwave absorption in DNA. Phys. Rev. A 40, 2148-2153, 1989.

[49] Baverstock K.F., Cundal R.D., Are solitons responsible for energy transfer in oriented DNA?. Int. J. Radiat. Biol. 55, 152-153, 1989.

[50] Selvin P.R., Cook D.N., Pon N.G., Bauer W.R., Klein M.P., Hearst J.E., Torsional rigidity of positively and negatively supercoiled DNA. Science 255, 82-85, 1992.

[51] Xiao J.-X., Lin J.-T., Zhang G.-X., The influence of longitudinal vibration on soliton excitation in DNA double helices. J. Phys. A: Math. Gen. 20, 2425-2432, 1987.

[52] Volkov S.N., Conformational transition. Dynamics and mechanism of long-range effects in DNA. J. Theor. Biol. 143, 485-496, 1990.

[53] Yakushevich L.V., Nonlinear vector model of the internal DNA dynamics. In: *Mathematical Models of Non-linear Excitations, Transfer, Dynamics, and Control in Condensed Systems and Other Media*. Eds. Uvarova L.A., Arinshtein A.E., and Latyshev A.V. Plenum Publishing Corporation, New York, (to appear).

[54] Dauxois M., Peyrard M., Bishop A.R., Entropy-driven DNA denaturation. Phys. Rev. E 47, R44-R47, 1993.

[55] Krumhansl J.A., Schrieffer J.R., Dynamics and statistical mechanics of a one-dimensional model hamiltonian for structural phase transitions. Phys. Rev. B 11, 3535-3545, 1975.

[56] Nakanishi M., Tsuboi M., Saijo Y., Nagamure T., Stopped-flow ultraviolet spectroscopy for hydrogen-exchange studies of nucleic acids. FEBS Lett. 81, 61-64, 1977.

[57] Nakanishi M., Tsuboi M., Two channels of hydrogen exchange in a double-helical nucleic acid. J. Mol. Biol. 124, 61-77, 1978.

[58] Mandal C., Kallenbach N.R., Englander S.W. Base-pair opening and closing reactions in the double helix. J. Mol. Biol. 135, 391-411, 1979.

[59] Frank-Kamenetskii M.D., Fluctuational motility of DNA. Mol. Biol. 17, 639-652, 1983 (In Russian).

[60] Frank-Kamenetskii M.D., How the double helix breathes. Nature 328,17-18, 1987.

[61] Gabriel G., Grant E.H., Tata R., Brown P.R., Gestblom B., Noreland E., Microwave absorption in aqueous solutions of DNA. Nature 328, 145-146, 1987.

[62] Maleev V.Ya., Kashpur V.A., Glibitsky G.M., Krasnitskaya A.A., Veretelnik Ye.V., Absorption of DNA solutions in the 9-12 GHz frequency range. Biopolim. Kletka 2, 35-38, 1986.

[63] Foster K.R., Epstein B.R., Galt M.A., Resonances in the dielectric absorption of DNA ? Biophys. J. 52, 421-425, 1987.

[64] Garibov R.A., Ostrovskii A.V., Does the microwave radiation change the dynamical behavior of macromolecules? Uspekhi Sovrem. Biol. 110, 306-320, 1990 (In Russian).

[65] Davis M.E., Van Zandt L.L., Microwave response of DNA in solution. Theory. Phys. Rev. 37 A, 888-899, 1988.

[66] Van Zandt L.L., Resonant microwave absorption by dissolved DNA. Phys. Rev. Lett. 57, 2085-2087, 1986.

[67] Van Zandt L.L., Davis M.E., Theory of anomalous resonant absorption of DNA at microwave frequencies. J. Biomol. Struct. Dyn. 3, 1045-1053, 1986.

[68] Muto V., Scott A.C., Christiansen P.L., Microwave and thermal generation of solitons in DNA. J. de Phys. 50, C3, 217-222, 1989.

[69] Baverstock K.F., Cundall R.B., Solitons and energy transfer in DNA. Nature, 332, N3, 312-313, 1988.

[70] Baverstock K.F., Cundall R.B., Long-range energy transfer in DNA. Radiation Physics and Chemistry, 32, 553-556, 1988.

[71] Arroyo C.M., Carmichael A.J., Swenberg C.E., Myers L.S., Neutron induced free radicals in oriented DNA. Int. J. Radiat. Biol. 50, 789-793, 1986.

[72] Miller J.H., Wilson W.E., Swenberg C.E., Myers L.S., Charlton D.E., Stochastic model for free radical yields in oriented DNA exposed to density ionizing radiation at 77 K. Int. J. Radiat. Biol. 53, 901-907, 1988.

[73] Edholm O., Nilsson L., Berg O., Ehrenberg M., Claesens F., Grässlund A., Jönsson B., Taleman O., Biomolecular dynamics. A report from a workshop in Gysinge, Sweden, October 4-7, 1982. Quart. Rev. Biophys. 17, 125-151, 1984.

[74] Khan A., Bhaumic D., Dutta-Roy B., The possible role of solitonic process during A to B conformational changes in DNA. Bull. Math. Biol. 47, 783-789, 1985.

[75] Sobell H.M., Kink-antikink bound states in DNA structure. In: *Biological Macromolecules and Assemblies*. Eds. F.A. Jurnak, A. McPherson. John Wiley and Sons, New York, 1984, pp. 172-234.

[76] Volkov S.N., Conformational transition. Dynamics and mechanism of long-range effects in DNA. J. Theor. Biol. 143, 485-496, 1990.

[77] Jensen P., Jaric M.V., Bannemann K.H., Soliton-like processes during right-left transition in DNA. Phys. Lett. 95 A, 204-208, 1983.

[78] Volkov S.N., Nonlinear waves and conformational mobility of DNA. Preprint ITP-84- 52P, Inst. Theor. Phys., Kiev.

[79] Pohl F.M., Jovin T.M., Baehr W., Holbrook J.J., Ethidium bromide as a cooperative effector of a DNA structure. Proc. Natl. Acad. Sci. USA 69, 3805-3809, 1972.

[80] Kolata G.B., Bacterial genetics: action at a distance on DNA. Science 198, 41-42, 1977.

[81] Hogan M., Dattagupta N., Crothers D.M., Transmission of allosteric effects in DNA. Nature 278, 521-524, 1979.

[82] Crothers D.M., Fried M., Transmission of long-range effects in DNA. Cold Spring Harbor Symposia on Quantitative Biology 47, 263-269, 1983.

[83] Wells R.D., Goodman T.C., Hillen W., Horn G.T., Klein R.D., Larson J.E., Mller U.R., Nevendorf S.R., Panayatatos N., Stirdivant S.M., DNA structure and gene: *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, New York, Vol. 24, 1980, pp. 167-267.

[84] Yakushevich L.V., Non-linear DNA dynamics and problems of gene regulation. Nanobiology 1, 343-350, 1992.

[85] Polozov R.V., Yakushevich L.V., Nonlinear waves in DNA and regulation of transcription. J. Theor. Biol. 130, 423-430, 1988.

[86] Salerno M., Dynamical properties of DNA promoters. Phys. Lett. A 167, 49-53, 1992.

[87] Salerno M., Kivshar Yu. S. DNA promoters and nonlinear dynamics. Phys. Lett. A 193, 263-266, 1994.

[88] Salerno M., Nonlinear dynamics of plasmid pB R322 promoters. In: *Nonlinear Excitations in Biomolecules*. Ed. Peyrard M., Springer, 1995, pp. 147-153.

[89] Fritzsche H., New structural and dynamic aspects of DNA as revealed by nuclear magnetic resonance". Comm. Mol. Biophys. 1, 325-336, 1982.

[90] Keepers J. W., James Th. L., Models for DNA backbone motions: an interpretation of NMR relaxation experiments. J. Am. Chem. Soc. 104, 929-939, 1982.

[91] McClure W. R., Mechanism and control of transcription in procaryotes. Ann. Rev. Biochem. 54, 171-204, 1985. revealed by nuclear magnetic resonance". Comm. Mol. Biophys. 1, 325-336, 1982.

20

From the FPU Chain to Biomolecular Dynamics

A. V. Zolotaryuk
A. V. Savin
P. L. Christiansen

ABSTRACT The molecular dynamics simulations originally performed by Fermi, Pasta, and Ulam for an isolated one-dimensional chain with cubic anharmonicity had led afterwards to the discovery of stable coherent structures called "solitons". Any study of the stability of solitons on such a one-dimensional lattice with respect to transverse motions of chain atoms or molecules requires introduction of a secondary structure realized for biological macromolecules in the form of a helix. In the simplest case of intermolecular interactions with spherical symmetry, the straightforward generalization of the Fermi-Pasta-Ulam chain to higher dimensions gives rise to the helical structure: zigzag in two dimensions and α-helix in three dimensions. The planar zigzag structure is provided by the first- and second-neighbor intermolecular bonds, whereas the helical structure in three dimensions requires for its stabilization, at least, three types of interactions. The coupled nonlinear field equations that describe longitudinal and transverse displacements of molecules in the helix backbone are studied. In particular, stable non-topological two- and three-component soliton solutions in two and three dimensions, respectively, are shown to exist. These solutions describe supersonic pulses of longitudinal compression propagating together with localized transverse thickening ("bulging") and torsional stretching (twisting). Other, more specific, types of solitons are investigated in two dimensions for the zigzag backbone.

1 Introduction

One of the most widespread models used for studies of nonlinear dynamics is a one-dimensional (1D) anharmonic lattice termed the Fermi-Pasta-Ulam (FPU) chain [10], the original investigation of which gave rise to discovery by Zabusky and Kruskal of a *soliton* [19]. Any intermolecular potential in such a 1D lattice has a hard ("positive") anharmonicity. This is a conventional type of anharmonicity in nonlinear lattices and its physical meaning is as follows. When nearest-neighbor atoms or molecules of the chain are

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 393–407, 2000.

displaced from their equilibrium positions, a repulsion force between them becomes stronger than the harmonic approximation of this interaction. In other words, a hard anharmonic force contributes to this repulsion with the positive sign. As a result of the presence of such a positive anharmonicity, dynamically stable solitary waves can propagate along the chain with supersonic velocities [19, 17]. These nonlinear collective excitations are referred to as the lattice (or acoustic) solitons.

In applications to real biological macromolecules [2, 18], the standard 1D FPU model should be generalized in order to include transverse motions of chain molecules. However, as shown by numerical simulations [14, 13], soliton excitations in anharmonic chains are extremely sensitive to their transverse perturbations and therefore the problem of soliton propagation in such 3D entities is far from fully understood. Consequently, the question on the existence and stability of moving solitary waves along realistic biomolecules, considered in three dimensions, is of great interest.

Here we deal with looking for "numerically exact" solitary wave solutions for a helix backbone, the molecules of which are allowed to move in two or three dimensions. All intermolecular interactions are assumed to be of the point-point type with spherical symmetry. The backbone is considered as an *isolated* object which is not subjected to any substrate potential. Since only point-point intermolecular interactions are involved, the 3D helix backbone will be stabilized only if, besides the nearest-neighbor interactions, at least, two other types of intermolecular forces are involved. Particularly, for the α-helix macromolecule, these interactions are assumed to take place between the first, second, and third neighbors.

Finally, it should be emphasized that even if the molecules are assumed to be coupled by harmonic forces, an effective anharmonicity appears due to geometry of the backbone. However, this "geometric" anharmonicity appears to have the opposite ("negative") sign compared with intrinsic hard anharmonicity of chemical bonds. For other purposes, namely breather-like solutions, the effects of such a geometric nonlinearity have previously been investigated by Cadet [1].

2 Helices in two and three dimensions

Let us consider molecules or groups of atoms (e.g., amino acids) in three dimensions which are linked together in a periodic linear (valence-bonded) sequence of a helical form, as shown in Fig. 1, by the first-, second-, and νth-neighbor (when counted within this sequence) forces of spherical symmetry. The integer $\nu > 2$ (in the case of three dimensions) is required to satisfy the condition that a distance d_ν between the νth neighbors is the shortest length among other intermolecular distances, except for d_1 and d_2, distances between the first and second neighbors. Then, ν deter-

mines the number of "spines" within the helix backbone (e.g., $\nu = 3$ for the α-helix), periodic subsequences with the shortest lattice spacing d_ν. Since all intermolecular forces in the chain are assumed to be of spherical symmetry and the molecules are allowed to move in all the three dimensions, the forces between the nearest neighbors (valence bonds) together with the forces between the νth-neighbor molecules (e.g., hydrogen bonds in the α-helix) only *partially* stabilize the helix backbone. At least, some third type of intermolecular bonds must be involved for stabilizing the helix structure. For instance, it can be the three-particle interaction fixing a certain valence angle ψ, the angle between the nearest-neighbor (valence) bonds. The simplest and most straightforward way of fully stabilizing the 3D helical structure is to add a second-neighbor coupling. In this respect, such a helix will be the most simple generalization of the 1D FPU chain to three dimensions. Similarly, for stabilizing a planar zigzag backbone, at least, two types of intermolecular interactions (e.g., between the first and second neighbors) must be present in the system.

The 3D geometry of a regular (undistorted) helix backbone, when its molecules are found in equilibrium positions, can be uniquely given by a set of three parameters. Let us locate the Cartesian system of coordinates $(X,\ Y,\ Z)$ with respect to the helix backbone as shown in Fig. 1. Then, one of these sets can be $\{R_0,\ \phi,\ h\}$ where R_0 is the radius of the cylinder which spans the helix backbone, ϕ is the angle in the XY plane formed by each three successive chain molecules (this angle is the projection of the angle ψ between the nearest valence bonds onto the XY plane), and $h = Z_0/R_0$ with Z_0 being the Z projection of the distance between the nearest-neighbor molecules of the linear sequence. In what follows, we term ϕ and ψ the torsional and valence angles, respectively. In terms of the parameters described above, the radius-vector of each molecule of the helix backbone can be represented as $\mathbf{R}_n = R_0\,(\cos(n\phi),\ \sin(n\phi),\ nh)$ with $n = 0,\ \pm 1,\ \ldots$. When $\phi > 2\pi/\nu$ or $\phi < 2\pi/\nu$, we refer to such a chain as a right- or left-handed helix, respectively. Correspondingly, the screw-sense of each spine is clockwise or counter-clockwise. As regards for the α-helix, each amino acid is related to the next one by a translation of $Z_0 = 1.5\ \mathring{A}$ along the helix axis and a clockwise rotation of $\phi = 100^\circ$ that gives 3.6 amino acids per turn of the helix.

Alternatively, the helix can be described by other three parameters, namely by the equilibrium distances between the first-, second-, and νth-neighbor molecules, $\{d_1,\ d_2,\ d_\nu\}$. The length of the vector

$$\begin{aligned} \mathbf{a}_{jn} &= (\mathbf{R}_{n+j} - \mathbf{R}_n)\,/R_0 \\ &= (\cos[(n+j)\phi] - \cos(n\phi),\ \sin[(n+j)\phi] - \sin(n\phi),\ jh) \end{aligned} \tag{20.1}$$

that connects the nth and $(n+j)$th vertices of the regular helix backbone, does not depend on the number of the chain site:

$$|\mathbf{a}_{jn}| = \sqrt{2[1-\cos(j\phi)] + j^2h^2} \equiv a_j = d_j/R_0\ ,\quad j = 1,\ 2,\ \nu. \tag{20.2}$$

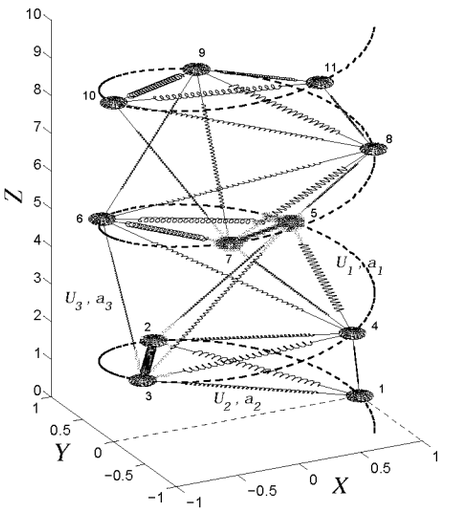

FIGURE 1. A fragment of the helix backbone consisting of 11 molecules. The geometry of the backbone is in accordance with the structure of α-helix, i.e., $\nu = 3$ and $\phi = 100^\circ$. The intermolecular interactions U_j, $j = 1, 2, 3$, are schematically shown by springs of different thickness. The thicker springs correspond to stronger stiffness of the bonds.

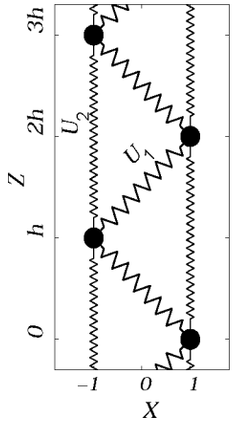

FIGURE 2. The planar zigzag backbone as a particular case of the helical structure ("2D helix").

Next, the third and fourth sets of parameters which also may be used for description of the helical geometry are $\{d_1,\ \psi,\ d_\nu\}$ and $\{R_0,\ \phi,\ \psi\}$. Using Eqs. (20.2) as well as the relations for the angles ϕ and ψ: $\cos\phi = 1 - \sqrt{a_1^2 - a_2^2/4}$ and $\cos\psi = 1 - d_2^2/2d_1^2$, one finds immediately a set of equations that establish one-to-one mappings between the four above-described sets of geometric parameters: $\{R_0,\ \phi,\ h\} \Longleftrightarrow \{d_1,\ d_2,\ d_\nu\} \Longleftrightarrow \{d_1,\ \psi,\ d_\nu\} \Longleftrightarrow \{R_0,\ \phi,\ \psi\}$. Moreover, one can find the relation [4]

$$\frac{a_\nu^2}{2} = 1 - \cos(\nu\phi) - \nu^2 \frac{(\cos\phi + \cos\psi)(1 - \cos\phi)}{1 + \cos\psi} \tag{20.3}$$

which gives the dependence of the distance a_ν on the integer ν for given angles ϕ and ψ. The minimization of this distance with respect to all integers $\nu \geq 3$ yields the required integer ν. Thus, for the α-helix macromolecule $\phi = 100^\circ$ and $\psi = 110^\circ$, and as a result, the minimal distance a_ν occurs at $\nu = 3$. Note that the inequality $\cos\phi + \cos\psi < 0$ is always valid and therefore $2a_1 > a_2$.

Consider now the "helical" structure in two dimensions. In this particular case $\phi = \pi$ and $\nu = 2$, and the helix is simply reduced to the *zigzag* backbone shown in Fig. 2 (two spines). The only two types of bonds, namely between the first- and second-neighbor molecules are sufficient to stabilize secondary structure [20]. Note that in the other particular case $\phi = 2\pi$ and $\nu = 1$, the helical structure is reduced to the standard 1D FPU chain (one spine) with the nearest-neighbor interactions.

3 Equations of motion for a helix backbone

Let M be the mass of chain molecules, K the characteristic stiffness of intermolecular forces, x_n, y_n, and z_n the local coordinates of an instant displacement of the nth molecule from its equilibrium position given by the radius-vector $\mathbf{R}_n$. Then the total Hamiltonian of the helix backbone with three types of interactions as described above, i.e., between the first-, second-, and νth-neighbor molecules, can be written in the following dimensionless form (scaled by the factor by KR_0^2):

$$\mathcal{H} = \sum_n \left[\frac{1}{2} \left(\frac{d\mathbf{q}_n}{d\tau} \right)^2 + \sum_j U_j \left(|\mathbf{q}_{n+j} - \mathbf{q}_n| \right) \right] \tag{20.4}$$

where $\tau = \omega_0 t$ with $\omega_0 = \sqrt{K/M}$. Each of the three potential functions $U_j(r_{jn})$'s describes a bond connecting the nearest-neighbor ($j = 1$), second-neighbor ($j = 2$), or νth-neighbor ($j = \nu$) molecules. These interactions are supposed to have spherical symmetry, i.e., to depend only on the intermolecular distances $r_{jn} = |\mathbf{q}_{n+j} - \mathbf{q}_n|$ between the nth and $(n + j)$th

molecules, where

$$\mathbf{q}_n = (q_{1n},\ q_{2n},\ q_{3n}) = (\cos(n\phi) + x_n/R_0,\ \sin(n\phi) + y_n/R_0,\ nh + z_n/R_0)\,. \tag{20.5}$$

The functions $U_j(r)$, $j = 1, 2, \nu$, are assumed to be of the standard form (like the Lennard-Jones or Morse potential) with $U_j(a_j) = 0$ and $U_j'(a_j) = 0$; $\kappa_j = U_j''(a_j)$ is the dimensionless stiffness constant of the bond connecting the nth and $(n + j)$th molecules.

The equations of motion that correspond to the Hamiltonian (20.4) are

$$\frac{d^2\mathbf{q}_n}{d\tau^2} = \sum_j \left[W_j(r_{jn})\left(\mathbf{q}_{n+j} - \mathbf{q}_n\right) - W_j(r_{j,n-j})\left(\mathbf{q}_n - \mathbf{q}_{n-j}\right)\right] \tag{20.6}$$

with $W_j(r_{jn}) = U_j'(r_{jn})/r_{jn}$. The summation over $j = 1,\ 2,\ \nu$ also includes the particular cases of the 1D chain and the planar zigzag: (one spine, $\nu = 1$) $j = 1$ and (two spines, $\nu = 2$) $j = 1, 2$.

4 Small-amplitude limit

The linearized equations of motion (20.6) admit the plane wave solution of the type $\exp[i(kn - \Omega\tau)]$ with the dispersion law

$$\begin{vmatrix} \Omega^2 - c_{11} & -ic_{12} & -ic_{13} \\ ic_{12} & \Omega^2 - c_{22} & -c_{23} \\ ic_{13} & -c_{23} & \Omega^2 - c_{33} \end{vmatrix} = 0 \tag{20.7}$$

where the coefficients c_{ij}'s are given by

$$\begin{aligned} c_{11} &= 16\sum_j \frac{\kappa_j}{a_j^2}\sin^4\frac{j\phi}{2}\cos^2\frac{jk}{2}\ , \quad c_{22} = 4\sum_j \frac{\kappa_j}{a_j^2}\sin^2(j\phi)\sin^2\frac{jk}{2}, \\ c_{12} &= 4\sum_j \frac{\kappa_j}{a_j^2}\sin^2\frac{j\phi}{2}\sin(j\phi)\sin(jk),\ c_{33} = 4h^2\sum_j \frac{j^2\kappa_j}{a_j^2}\sin^2\frac{jk}{2}, \\ c_{13} &= 4h\sum_j \frac{j\kappa_j}{a_j^2}\sin^2\frac{j\phi}{2}\sin(jk),\ c_{23} = 4h\sum_j \frac{j\kappa_j}{a_j^2}\sin(j\phi)\sin^2\frac{jk}{2}. \end{aligned} \tag{20.8}$$

In the 1D case ($\phi = 2\pi$) with only nearest-neighbor interactions ($j = 1$ and $a_1 = h$), we obtain immediately the well-known dispersion law for the harmonic 1D chain: $\Omega = \sqrt{c_{33}} = 2\sqrt{\kappa_1}\sin(k/2)$ that describes one acoustic branch. However, in the 2D case for the planar zigzag ($j = 1, 2$), Eq. (20.8) is reduced to $(\Omega^2 - c_{11})(\Omega^2 - c_{33}) - c_{13}^2 = 0$ that admits a two-branch solution: one branch is of acoustic type and the other one of optical type.

In three dimensions, using the explicit form of Eqs. (20.7) and (20.8), one can be convinced of the existence of two acoustic branches. These two

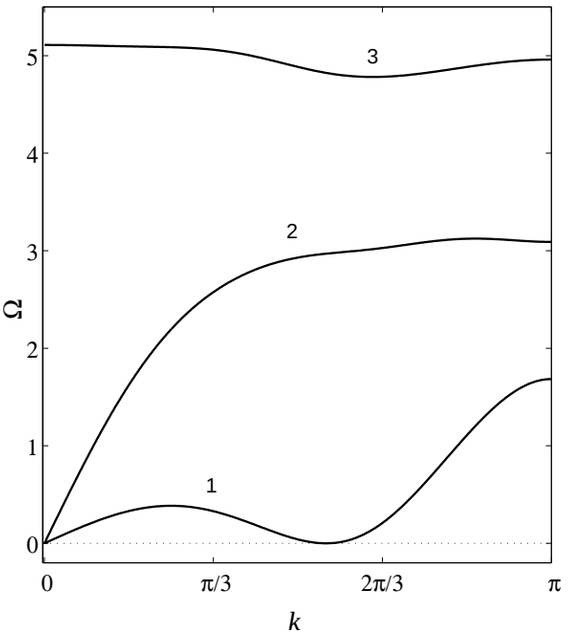

FIGURE 3. Frequencies Ω_t (curve 1), Ω_l (curve 2), and Ω_{op} (curve 3) against wave number k, $0 \le k \le \pi$, for the α-helix.

$\Omega_l(k)$ and $\Omega_t(k)$ (see Fig. 3) correspond to the longitudinal and torsional oscillations of the chain molecules, respectively. The third root of Eq. (20.7) gives an optical branch $\Omega_{op}(k)$, corresponding to the transverse oscillations of molecules in the radial direction of the helix backbone. Explicitly, at $k = 0$ one finds $\Omega_{op}^2(0) = c_{11} + c_{22} + c_{33} = 16 \sum_j (\kappa_j / a_j^2) \sin^2(j\phi/2)$.

The presence of two acoustic branches results in the existence of two speeds of sound: longitudinal (v_l) and torsional (v_t). In dimensionless form, they can be defined as the limits $s_{l,t} = v_{l,t}/v_0 = h \lim_{k\to 0} \Omega_{l,t}/k$ where $v_0 = (K/M)^{1/2} R_0$ is the characteristic velocity of small-amplitude waves in the helix backbone.

For numerical computations we used the following values of the system parameters:

$$\nu = 3, \quad \phi = 100^\circ, \quad h = 1, \quad \kappa_1 = 10, \quad \kappa_2 = 5, \quad \kappa_3 = 1 \ . \tag{20.9}$$

In the parameter set (20.9), the values for the number ν and the angle ϕ correspond to the α-helix protein molecule. The stiffness constants κ_1, κ_2, and κ_3 are related to each other approximately as elasticities of the valence bond, valence angle, and hydrogen bond. The form of all the three

dispersion curves for the parameter set (20.9) is shown in Fig. 3. At $k = 0$, the frequencies are $\Omega_l = \Omega_t = 0$ and $\Omega_{op} = 5.11$. It follows from the explicit representation of the coefficients (20.8) that at a certain value of the wave number $k = k_0$, the free term in Eq. (20.7) becomes zero. This means that at this value, softening torsional oscillations happens $[\Omega_t(k_0) = 0]$. For the parameter set (20.9) this value is $k_0 = 1.748$. As follows from Fig. 3, the frequency spectrum of torsional oscillations lies inside the frequency spectrum of longitudinal oscillations. Moreover, the velocity of the longitudinal sound $s_l = 3.39$ significantly exceeds the speed of the torsional sound $s_t = 0.75$.

5 Three-component soliton solutions

In this section we describe a numerical scheme for seeking solitary wave solutions of stationary profile for the basic equations of motion (20.6). This scheme can be applied if the profile of solutions appears to be sufficiently smooth while varied from site to site along the chain [3, 4]. In order to study discreteness effects, we should use more complicated numerical techniques such as the pseudo-spectral method suggested by Eilbeck and Flesch [9] and afterwards developed by Duncan *et al* [8]. When the soliton solutions to the equations of motion have been found, they can be chosen as initial conditions for numerical simulations of these equations. The final profile of the vector lattice field $\mathbf{q}_n(\tau)$ obtained under the simulations at sufficiently large times τ allows us to conclude whether or not the initial soliton profile is a stable solution of Eqs. (20.6). The main point in such a numerical approach is an appropriate choice of a discrete functional for minimization.

To accomplish the soliton analysis of the equations of motion (20.6), we rewrite them, introducing the new variables $q_{1n} = (1 + \eta_n)\cos(n\phi + \theta_n)$, $q_{2n} = (1+\eta_n)\sin(n\phi+\theta_n)$, and $q_{3n} = nh+\beta_n$ where η_n describes the radial displacement of the nth molecule from the cylinder surface that spans the helix backbone when its molecules are situated at the equilibrium positions. It is positive if the displacement is outside and a bulging of the helix occurs in this region. If the displacement is directed inside the helix, then η_n is negative. The second generalized coordinate θ_n describes the azimuthal deviation of the nth molecule from its equilibrium position, and β_n is the Z coordinate of the displacement. The Lagrangian of the helix backbone written in terms of these new variables takes the form

$$\mathcal{L} = \sum_n \left\{ \frac{1}{2} \left[\left(\frac{d\eta_n}{d\tau} \right)^2 + (1+\eta_n)^2 \left(\frac{d\theta_n}{d\tau} \right)^2 + \left(\frac{d\beta_n}{d\tau} \right)^2 \right] - \sum_j U_j(r_{jn}) \right\} \tag{20.10}$$

where the distance r_{jn} is given by

$$r_{jn}^2 = (1+\eta_n)^2 + (1+\eta_{n+j})^2 - 2(1+\eta_n)(1+\eta_{n+j})$$

$$\times \cos(j\phi + \theta_{n+j} - \theta_n) + (jh + \beta_{n+j} - \beta_n)^2 \ . \qquad (20.11)$$

The corresponding equations of motion are

$$\begin{aligned}
\frac{d^2\eta_n}{d\tau^2} &= (1+\eta_n)\left(\frac{d\theta_n}{d\tau}\right)^2 \\
&- \sum_j W_j(r_{j,n-j})\left[1+\eta_n-(1+\eta_{n-j})\cos(j\phi+\theta_n-\theta_{n-j})\right] \\
&+ \sum_j W_j(r_{jn})\left[1+\eta_n-(1+\eta_{n+j})\cos(j\phi+\theta_{n+j}-\theta_n)\right],
\end{aligned} \qquad (20.12)$$

$$\begin{aligned}
\frac{d^2\theta_n}{d\tau^2} &= \frac{1}{1+\eta_n}\left[-2\frac{d\eta_n}{d\tau}\frac{d\theta_n}{d\tau}\right. \\
&+ \sum_j W_j(r_{jn})(1+\eta_{n+j})\sin(j\phi+\theta_{n+j}-\theta_n) \\
&- \left.\sum_j W_j(r_{j,n-j})(1+\eta_{n-j})\sin(j\phi+\theta_n-\theta_{n-j})\right],
\end{aligned} \qquad (20.13)$$

$$\frac{d^2\beta_n}{d\tau^2} = \sum_j \left[W_j(r_{jn})(jh+\beta_{n+j}-\beta_n) - W_j(r_{j,n-j})(jh+\beta_n-\beta_{n-j})\right]. \qquad (20.14)$$

In the simplest case, $\nu = 1$ and $j = 1$, Eq. (20.14) describes the 1D FPU model with the nearest-neighbor interactions, while for $j > 1$ one obtains the 1D generalization of this model including long-range intermolecular interactions. For the particular case of the first- and second-neighbor interactions, $\nu = 2$ ($j = 1, 2$), this model has extensively been studied in Refs. [12, 11].

We assume that the soliton solutions for the 3D backbone have moving permanent profile, so that one can write $\eta_n = \eta(nh - s\tau)$, $\theta_n = \theta(nh - s\tau)$, and $\beta_n = \beta(nh - s\tau)$ where $s = v/v_0$ with v_0 being the characteristic sound velocity defined in the previous section. There is no need to take into account the dispersion of the optical mode [3], and therefore we can approximate the first and second time derivatives of η_n by the simplest spatial difference derivatives as follows

$$d\eta_n/d\tau = -s\eta'(nh - s\tau) \simeq -s(\eta_{n+1} - \eta_{n-1})/2h, \qquad (20.15)$$

$$d^2\eta_n/d\tau^2 = s^2\eta''(nh - s\tau) \simeq s^2(\eta_{n+1} - 2\eta_n + \eta_{n-1})/h^2. \qquad (20.16)$$

However, for the longitudinal and torsional displacements we need to take into account the dispersion which arises from the discreteness of the chain backbone. To this end, we represent the time derivatives of θ_n and β_n by

differences which additionally contain higher-order spatial difference derivatives chosen in an appropriate way, namely:

$$d\theta_n/d\tau \simeq s(\varphi_{n+1} - 5\varphi_n - 2\varphi_{n-1})/6h, \quad (20.17)$$

$$d^2\theta_n/d\tau^2 \simeq -s^2(\varphi_{n+1} - 15\varphi_n + 15\varphi_{n-1} - \varphi_{n-2})/12h^2, \quad (20.18)$$

$$d^2\beta_n/d\tau^2 \simeq -s^2(\rho_{n+1} - 15\rho_n + 15\rho_{n-1} + \rho_{n-2})/12h^2, \quad (20.19)$$

where $\varphi_n = \theta_{n+1} - \theta_n$ and $\rho_n = \beta_{n+1} - \beta_n$ are the relative torsional and longitudinal displacements, respectively. In the continuum limit, when expanded up to the third [in the right-hand side of Eq. (20.17)] or fourth [in the right-hand sides of Eqs. (20.18) and (20.19)] order, Eqs. (20.17)–(20.19) become $\theta_\tau \simeq -s\theta_x$, $\theta_{\tau\tau} \simeq s^2\theta_{xx}$, and $\beta_{\tau\tau} \simeq s^2\beta_{xx}$ with $x = nh$. Replacing then the time derivatives in Eqs. (20.12)–(20.14) by the corresponding discretized versions (20.15)–(20.19), we obtain *discrete* equations for the displacements η_n, φ_n, and ρ_n. Having solved them, for instance, by minimization techniques [4] for a given value of velocity s, one finds a three-component soliton profile.

5.1 3D case: solitons of longitudinal compression

For the particular case of the α-helix macromolecule, the $\{\eta_n, \varphi_n, \rho_n\}$ soliton profile found by the discretization technique as described above is shown in Fig. 4 by rhombus marks. Each of these components has zero asymptotics at the chain ends. The anharmonicity parameters which were chosen for the studies of existence and stability of such three-component soliton solutions in the α-helix backbone are given according to Eqs. (20.9). With this choice, note that the nonlinear behavior of the chain molecules in the first turn is caused by the anharmonicity of soft hydrogen bonds. Therefore we took into account only the anharmonicity of the third-neighbor ($\nu = 3$) interactions, more precisely, we put $\gamma_1 = \gamma_2 = 0$ and $\gamma_3 > 0$. Here the dimensionless anharmonicity parameters γ_j, $j = 1, 2, 3$, are defined as coefficients at the cubic term expansions of the Morse potentials $U_j(r)$'s. The numerical solution of the corresponding discrete equations has shown that at weak anharmonicity (e.g., $\gamma_3 = 0.1$), there are no soliton solutions at all, while for the value $\gamma_3 = 1$, the soliton solutions exist only with velocities in the segment $1 < s/s_l < 1.065$. In this interval of velocities, the solutions have bell-shaped soliton profiles, smoothly depending on the chain site n. As illustrated by Fig. 4, in the region of localization of the soliton solution, a compression ($\rho_n < 0$) of the backbone along the helix axis occurs accompanied by a localized transverse "bulging" ($\eta_n > 0$) and a torsional stretching (twisting, $\varphi_n > 0$) of the helix. Thus, the soliton solutions were shown to exist only in a narrow interval of supersonic velocities. When $s \to s_l$, the soliton energy and amplitudes monotonically tend to zero while the width increases to infinity. With the growth of the velocity, the energy and absolute values of the amplitudes monotonically

increase whereas the soliton width monotonically decreases. It is important to emphasize that the soliton solutions exist only if the anharmonicity of intermolecular interactions is sufficiently strong, enough to suppress the "negative" anharmonicity caused by the geometry of helical structure (so-called geometric anharmonicity [1]).

We have studied the soliton dynamics in the chain consisting of $N = 100$

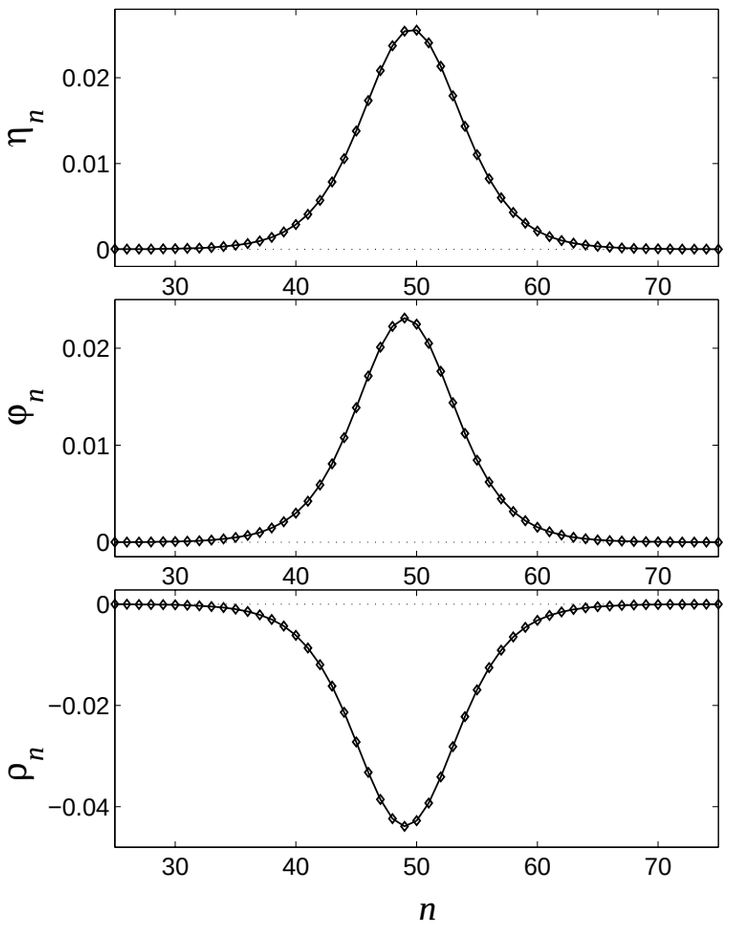

FIGURE 4. Three-component profile of the soliton in the α-helix with anharmonicities $\gamma_1 = \gamma_2 = 0$ and $\gamma_3 = 1$ found numerically by solving the discrete equations (rhombus marks) and at the final instant $\tau = 28062.45$ when the soliton has passed 100000 chain sites (solid lines). The initial velocity of the soliton was $s = 1.05 s_l$.

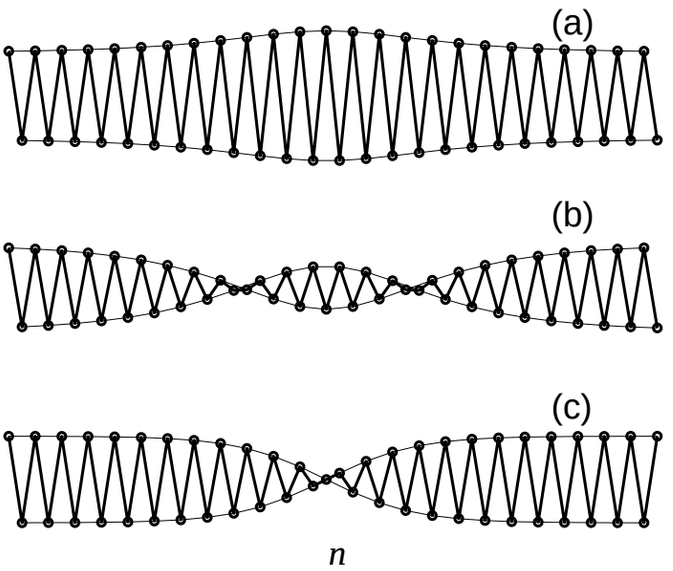

FIGURE 5. Three types of localized deformation of the planar zigzag ($\nu = 2$, $\gamma_1 = 0$ and $\gamma_2 = 1$): (a) solitons of longitudinal compression and transverse "bulging", solitons of longitudinal stretching and transverse slendering with (b) two nodes (non-topological soliton) and (c) one node ("metastable topological soliton").

molecules with the parameters (20.9) when $\gamma_1 = \gamma_2 = 0$ and $\gamma_3 = 1$. When the initial speed was $s = 1.05 s_l = 3.5645$, then the soliton has passed 100000 chain sites during the time $\tau = 28062.45$. This propagation corresponds to the velocity $s = 100000/\tau = 3.5635 = 1.0497 s_l$ and the final three-component soliton profile is shown in Fig. 4 by solid lines. It is remarkable that the soliton shape at the final instant of time perfectly coincides with the initial profile. The simulations of the equations of motion have demonstrated that the above-described procedure of looking for soliton solutions gives "numerically exact" soliton profiles and the solitons themselves are dynamically stable.

5.2 2D case: other types of solutions

In the 2D case for the planar zigzag backbone, the solitons with longitudinal compression and transverse bulging also exist [20]. These solutions may be considered as a particular case of the three-component solitons described above for the 3D helical structure if the φ_n-component is identically zero. The two-component profile of the η_n and ρ_n displacements is shown

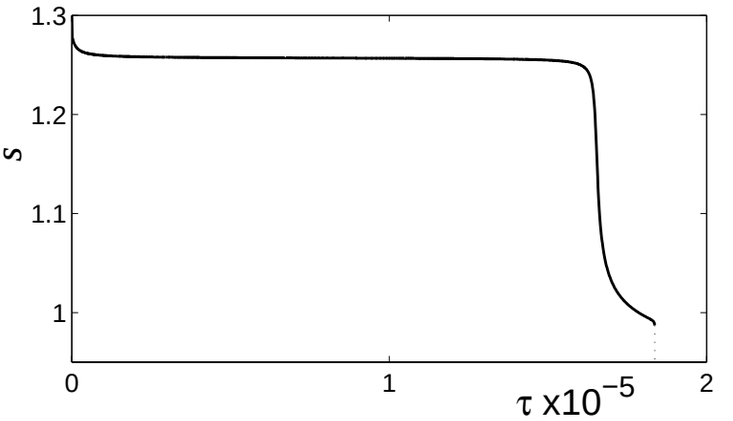

FIGURE 6. Velocity of the metastable topological soliton against time τ, propagating in the planar zigzag ($\nu = 2$) with $\gamma_1 = 0$ and $\gamma_2 = 1$.

schematically in Fig 5(a). The shape of these components is the same as that depicted in Fig. 4.

The geometric anharmonicity prevents the formation of solitons with longitudinal compression, but in turn, it can provide the existence of solitons with longitudinal stretching, but transverse thickening. One of these solitons has been shown, both analytically and numerically [20], to exist with the two-component profile shown schematically in Fig. 5(b). For this soliton, the zigzag transversely "overslenders" at the soliton center, so much that two nodes for the profile of transverse displacements appear as illustrated by Fig. 5(b). The second type of soliton with longitudinal stretching and transverse bulging has only one node in the transverse displacement field as illustrated by Fig. 5(c), but in this case the boundary conditions are kink-like, namely: $\eta_n \to 0$ if $n \to -\infty$ and $\eta_n \to -1$ if $n \to +\infty$, whereas $\rho_n \to 0$ if $n \to \pm\infty$. The time dependence of the velocity of such a "metastable topological soliton" is given by Fig. 6. At $\gamma = 1$, the minimal emission is at $s = 1.257 s_l$. The further decrease of the velocity results in increasing emission of small-amplitude waves.

6 Conclusions

It is an attractive point of view to study the transport of vibrational energy in biopolymers such as protein on the basis of 1D nonlinear lattice models. The first attempt in this direction was formulated by Davydov and Kislukha [7] (see the book [5] and review [16], and references therein) who

suggested that the intramolecular vibrational Amide-I mode could be self-trapped through its interaction with deformation in the protein structure. Further, Davydov *et al* [6] and Scott [15] generalized this theory to the α-helical structure, taking into account the dipole-dipole coupling between the three spines. While Davydov and Scott considered the intramolecular mode Amide-I, Yomosa [18] modeled the energy transfer in protein, essentially applying the 1D FPU model. Here we also focus only on the large-amplitude supersonic lattice solitons and all intermolecular (or interatomic) interactions are represented by pair central forces. We show that the minimal number of intermolecular pair interactions (of spherical symmetry) in the linear sequence of molecules, which are necessary for stabilizing a helix backbone, is two in the 2D case (zigzag) and three in the 3D case (helix). In other cases, when the intermolecular interactions are non-spherically symmetric, a helix Hamiltonian becomes much more complicated. Therefore the Hamiltonian (20.4) may be considered as the most simple generalization of the standard 1D FPU chain to three dimensions when each chain molecule has three degrees of freedom. The resulting geometric structure is a helix backbone with $\nu \geq 3$ spines. Three-component non-topological solitons have been shown to exist and propagate with supersonic velocities in the helix backbone with any number of spines. Similarly to the 1D case, for the existence of the soliton solutions, the presence of anharmonicity, at least, in the (hydrogen) bonds connecting each nth and $(n+\nu)$th molecules, is necessary. However, compared to the 1D FPU model, the existence of stable soliton solutions in the 3D helix backbone appears to be more limited. First, they exist only if the anharmonicity of the ν-neighbor forces is sufficiently strong. Second, the segment of admissible (supersonic) velocities of solitons is always finite. It is quite narrow and for all velocities from this segment, the soliton propagation is uniform with retaining velocity and profile. However, when the solitons collide, their behavior after collision depends on velocity. At the lower edge of the segment, the soliton collision is elastic whereas for higher velocities, at the upper edge, the solitons collide inelastically, with radiation of small-amplitude waves. Summarizing, we conclude that in a 3D isolated helix backbone, three-component solitons describing supersonic pulses of longitudinal backbone compression, still exist as dynamically stable objects. However, the range of their existence is more limited if compared to the 1D FPU lattice.

Acknowledgments: This work was partially carried out with the financial support from the European Economic Community (EEC) under the INTAS Grant No. 96-0158. Both of us (A.V.S. and A.V.Z.) would also like to express his gratitude to The MIDIT Center and The Institute of Mathematical Modelling of The Technical University of Denmark for hospitality. Stimulating and useful discussions with A.C. Scott as well as his suggestions in the process of this investigation are gratefully acknowledged.

7 References

[1] S. Cadet. *Phys. Lett.*, 121:77, 1987.

[2] M.A. Collins. *Chem. Phys. Lett.*, 77:342, 1981.

[3] P.L. Christiansen, A.V. Savin, and A.V. Zolotaryuk. *J. Comput. Phys.*, 134:108, 1997.

[4] P.L. Christiansen, A.V. Zolotaryuk, and A.V. Savin. *Phys. Rev. E*, 56:877, 1997.

[5] A.S. Davydov. *Solitons in Molecular Systems*. Reidel, Dordrecht, 1985.

[6] A.S. Davydov, A.A. Eremko, and A.I. Sergienko. *Ukr. Fiz. Zh. (Ukrainian Phys. J.)*, 23:983, 1978.

[7] A.S. Davydov and N.I. Kislukha. *Phys. Stat. Sol. B*, 59:465, 1973.

[8] D.B. Duncan, J.C. Eilbeck, H. Feddersen, and J.A.D. Wattis. *Physica D*, 68:1, 1993.

[9] J.C. Eilbeck and R. Flesch. *Phys. Lett. A*, 149:200, 1990.

[10] E. Fermi, J. Pasta, and S. Ulam. *Collected Works of Enrico Fermi*, volume II, page 978. Chicago, 1965.

[11] N. Flytzanis, St. Pnevmatikos, and M. Peyrard. *J. Phys. A: Math. Gen.*, 22:783, 1989.

[12] N. Flytzanis, St. Pnevmatikos, and M. Remoissenet. *J. Phys. C: Solid State Phys.*, 18:4603, 1985.

[13] P.S. Lomdahl, O.H. Olsen, and M.R. Samuelsen. *Phys. Lett. A*, 152:343, 1991.

[14] O.H. Olsen, P.S. Lomdahl, and W.C. Kerr. *Phys. Lett. A*, 136:402, 1989.

[15] A.C. Scott. *Phys. Rev. A*, 26:578, 1982.

[16] A.C. Scott. *Phys. Rep.*, 217:1, 1992.

[17] M. Toda. *Phys. Rep.*, 18:1, 1975.

[18] S. Yomosa. *Phys. Rev. A*, 32:1752, 1985.

[19] N.J. Zabusky and M.D. Kruskal. *Phys. Rev. Lett.*, 15:241, 1965.

[20] A.V. Zolotaryuk, P.L. Christiansen, and A.V. Savin. *Phys. Rev. E*, 54:3881, 1996.

21

Mutual Dynamics of Swimming Microorganisms and Their Fluid Habitat

John O. Kessler
G. David Burnett
Katherine E. Remick

ABSTRACT "Organisms alter their material environment, and their environment constrains and naturally selects organisms." Lenton's [17] statement applies especially well to populations of swimming micro-organisms. The mutual dynamic of themselves and their fluid habitat orders and constrains them, generates concentration-convection patterns [12], [15], enhances transport of metabolites and, at all scales, guides many of their interactions. Our objective is to describe mathematical models sufficient for reaching insights that can further guide theory and experiment. These models necessarily include nonlinear and stochastic features. To illustrate self-organization and the type of experimental statistical inputs available, we present some rather astonishing data concerning the motile bacteria *Bacillus subtilis* and hydrodynamics associated with their activity. The inescapable interdependence of physics and biology emerges from the analysis.

1 Introduction

Swimming unicellular micro-organisms are important and ubiquitous. They include varieties of protozoa, algae, and bacteria. They propel themselves by waving, flexing, or rotating various organelles. Although our discussion is organized around phenomena involving aerobic bacteria, much of it is applicable to other organisms as well. A summary pertaining to algae at low volume fractions and in "large" containers can be found in [22].

Swimming bacteria convert dissolved chemical potential energy into life processes: growth, storage, metabolism, and the power to swim. Sensing spatially and temporally varying significant conditions, and responding by directional swimming, are biological processes. They are not entirely deterministic. Even for the most uniform populations, swimming speed and swimming direction can only be specified through probability den-

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 409−426, 2000.

sity functions $P(\vec{V};\{q\},t)$, where $\{q\}$ represents a manifold of histories, including chemical concentrations at the present time t, and encountered earlier along specific trajectories. Not only the swimming vector, but other active processes, such as sensing, are likely subject to probabilistic rules [18]. Stochastic influences include intracellular thermal fluctuations, rotational Brownian motion, time-delayed responses to particular trajectories through chemical concentration space, and systematic (but unknowable) variabilities such as position of an individual in the cell division cycle. Capriciousness, or "free will", may also contribute.

The subjectively judged geometry of bioconvection patterns is robust, for particular cell varieties, conditions, culture age, fluid depths, initial conditions, etc. Nevertheless, no two patterns are ever exactly alike nor is their time evolution. The stochastic modulation of behavior of the living component of these micro-bio-fluid dynamic systems complicates and enriches the actuality, and the models that attempt to describe it [10], [3], [20].

This paper is organized around the idea that there are two regimes of phenomena that require rather different approaches, both for mathematical modelling and for experiments. The "Continuum" Regime (I) concerns bioconvection. It is characterized by large concentrations n of organisms occupying small volume fractions $\phi \equiv nv$, where v is the volume per organism. For bacteria, $n = 10^8 - 10^9 cm^{-3}, v = 10^{-11} - 10^{-12} cm^3$. The dynamics of this regime take place in "large" volumes of water, e.g. for $\geq 10^{-1} cm$ depth and $\geq 1 cm^3$ volume. The organisms interact indirectly only, via (1) convective motions due to buoyancy effects that result from directional swimming, and (2) variations in concentration of significant molecules, such as oxygen, caused by consumption and directional supply. The governing nonlinear coupled partial differential equations describe continua of molecular and cell concentrations $c(\vec{x},t)$ and $n(\vec{x},t)$, and the fluid velocity field $\vec{u}(\vec{x},t)$.

The other regime (II) considers bacteria constrained in fluid layers of depth $h \geq d$, where d is the diameter of one cell body, $\simeq 1\mu m$. The cells may be isolated, or included in populations that are almost close-packed. The organisms interact directly through the fluid velocity fields that they generate, by collisions, and perhaps via chemical signals. They interact not only with each other but with the surfaces, free or rigid, that bound the fluid in which they find themselves [23], [24]. This regime, the subject of current and planned experiments, is not as well understood as (I). It is especially exciting, since mean behavior appears to be affected by constraint: In very shallow layers of fluid, cells swim quickly, and the probability distributions alter. These constrained situations have practical significance, as in the generation of biofilms.

2 Bioconvection (I)

2.1 Observations

The geometry and time dependence of bioconvection patterns depend on initial and boundary conditions, concentration, growth/death rate, geometry of the containers, depth of the suspension of organisms, and swimming style. The figures shown here are intended to motivate modelling. They illustrate generic patterns obtained with suspensions of *Bacillus subtillis*, wild type 1085 (G. Ordal) in TB growth medium, $n \geq 10^8 cm^{-3}$, contained in plastic Petri dishes, $5.3cm$ internal diameter. Depths ranged from $< 1mm$ to $\simeq 5mm$; the surface of the fluid rises at the periphery because of capillarity. The photographic method was "dark field". Thus, regions containing many cells are bright, and conversely.

In Fig. 1, a shallow suspension ($\leq 2mm$ at the center) exhibits a heavily dislocated hexagonal pattern of plumes. At the edges, a pattern typical for a greater depth emerges. The high bacterial concentration surrounding the floating air bubbles are due to oxygen taxis: swimming up the O_2 concentration gradient that diffuses from the bubbles. In Fig. 2, the depth of the suspension is greater than in Fig. 1, but less than Fig. 3. In spite of appearance, the images are not three dimensional. Viewed in "real time", the images appear steady. Using time lapse photography, one observes strong movements and continual reorganization, especially of the transverse narrow dark bands that, as far as we know, appear only in aerobic bacterial bioconvection. In Fig. 4, a very shallow layer of fluid exhibits capillary rise on both sides of a solid, smooth stick, resulting in a linear pattern of definite wavelength. The shallow, flat regions of fluid cannot support a pattern. Fig. 4 was obtained during ongoing investigations of boundary conditions.

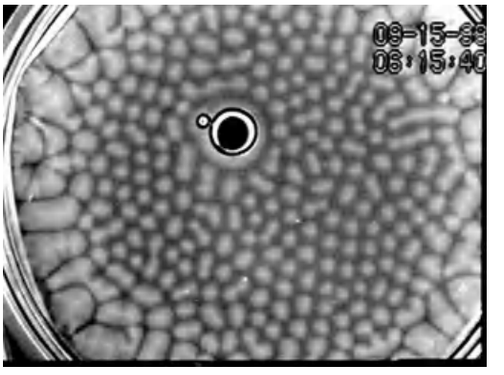

FIGURE 1. Plan view of shallow bacterial bioconvection, with floating bubbles. See text.

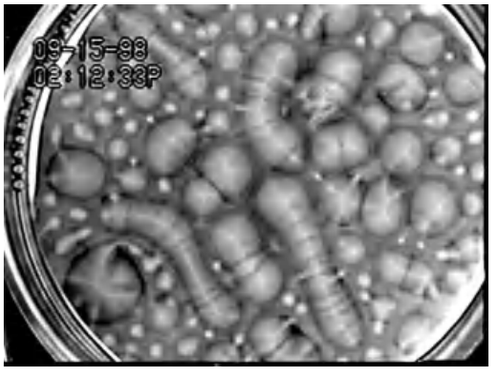

FIGURE 2. Plan view of bacterial bioconvection. The layer of fluid is deeper than in Fig. 1. See text.

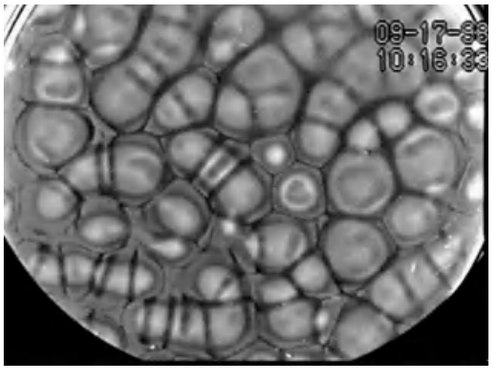

FIGURE 3. Plan view of bacterial bioconvection. The layer of fluid is deeper than in Fig. 2, approximately 5 mm. See text.

2.2 Continuum theory

The Navier-Stokes equations govern the fluid velocity field $\vec{u}$. They are driven by a gravitational body force that results from nonuniform concentrations of n, caused by directional swimming. The mean density increase, relative to the water in which the organisms swim and due to variations in n, is

$$\Delta\rho = nv[\rho(organism) - \rho(water)]. \tag{21.1}$$

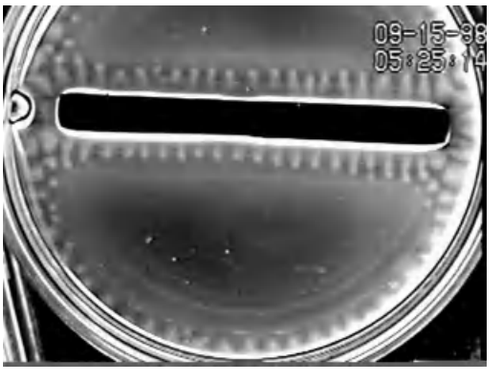

FIGURE 4. Plan view of shallow bacteria bioconvection. See text. Very shallow layers of fluid cannot support a pattern. The capillary ascent of fluid up the side of the flat black bounding surface and inside vertical surfaces of the Petri dish provides the necessary depth.

Bacterial densities can be estimated at 10% greater than that of water; algal densities are $\sim 2\%$ greater. The total mean density of the fluid is

$$\overline{\rho} = \rho(water) + \Delta\rho \approx \rho(water). \tag{21.2}$$

Then

$$\overline{\rho}\frac{D\vec{u}}{Dt} - \mu\nabla^2\vec{u} + \nabla p = \vec{g}\Delta\rho \tag{21.3}$$

and

$$\nabla \cdot \vec{u} \approx 0, \tag{21.4}$$

where μ is the viscosity of water, p is pressure, and $\vec{g}$ is the acceleration of gravity. Even though organisms can be thought of as incompressible, volume is not precisely conserved in the usual manner because some of it *swims* relative to $\vec{u}$. For small volume fractions and $V/u < 1$, where V is the swimming speed, Eq. (21.4) is nearly exact.

The fluid dynamics is driven by n, the cell concentration, which is subject to conservation, including directional swimming, birth and death. These factors depend on chemical concentration(s) c, local and time delayed. Conservation of c is given by

$$\frac{\partial c}{\partial t} - D\nabla^2 c + \nabla \cdot (\vec{u}c) = -\gamma(c)n \tag{21.5}$$

where D is the diffusivity of c and $\gamma(c)$ is the rate at which organisms consume (or, in some cases, emit) c. Gradients of c that skew the swimming velocity probability density function (SVPDF) $P(\vec{V}\ldots)$ arise, for example,

when c diffuses in from a bounding surface, as a result of consumptive depletion in the bulk fluid. It is likely that

$$\gamma(c) \sim \gamma_0 \frac{c}{(c_1 + c)} \tag{21.6}$$

so that at high c, $\gamma \approx \gamma_0$. When several molecules are important or limiting, Eq. (21.5) must be replaced by a set of equations, one for each c_i, and with $\gamma_i = \gamma_i(c_j \ldots)$. For example, there will not be much respiration after the food runs out. Conversely, assimilation of substrate ceases when oxygen levels are too low.

The nonlinear coupling term $\nabla \cdot (\vec{u}c)$ in Eq. (21.5) is due to advection, i.e. the transport of c by the fluid velocity field $\vec{u}$, caused through Eq (21.3) by directional accumulation of n in response to ∇c.

The case for photosynthetic organisms, such as motile algal cells, requires equations for the n-dependent attenuation of illumination and for illumination direction-oriented swimming of the cells [13].

The conservation of organisms is modelled by

$$\frac{\partial n}{\partial t} + \nabla \cdot (\vec{u}n) = -\nabla \cdot \vec{j} + Bn \tag{21.7}$$

where, as before, $\nabla \cdot (\vec{u}n)$ represents the nonlinear advective coupling. The birth/death rate is B, for organisms that multiply asexually. For "short" observation times B can be ignored. Otherwise, B can be a complicated function of various c_i, including exudates emitted by the cell population. Loss of mobility, when the cells are "like dead", e.g. due to temporary lack of oxygen, could also be modelled by a (negative) term B_n. When organisms interconvert, e.g. normal cells to/from swarmers, several conservation equations are required.

The most interesting and mathematically complex factor is the cell flux

$$\vec{j} = \text{“}n\vec{V}\text{”} \tag{21.8}$$

where $\vec{V}$ is the swimming velocity vector and the quotes indicate complexity. In the Keller-Segel model, "$n\vec{V}$" was taken as a deterministic term $n\langle\vec{V}\rangle$, plus diffusion $-D_c\nabla n$, where D_c is an effective diffusivity of cells [21]. The distributions of swimming speeds and directions of a large population of bacteria ought to be described by SVPDF's that depend on many factors, such as gradients and magnitudes of the oxygen concentration. We have recently shown that the usually encountered distribution of bacterial swimming speeds [14], [16] is made up of cohorts with various mean speeds $\langle|V|\rangle$, unlike the well known gas dynamic case where each component particle sweeps out the entire Maxwellian. Furthermore, each cell exhibits wide fluctuations in speed (Figs. 9a, 9b) during a typical observation period of ~ 5 seconds.

The description of cell flux can be given by

$$\vec{j}(\vec{x},t) = n(\vec{x},t)\int_{\vec{v}} P(\vec{V},\{q\},t)\vec{V}d\vec{V}, \tag{21.9}$$

where q includes $\vec{u}(\vec{x},t)$ via gyrotactic orientation [12], $\nabla f[c]$, and $f'[c]$. The function $f[c]$ is appropriate for chemo- or oxygen taxis and $f'[c]$ comprises the local concentrations of c and previous ones along the bundle of trajectories that pass through $(\vec{x},t)$. Furthermore, $\{q\}$ may include age dependence of the population, concentration, and factors that model sensory efficiency.

Since $\vec{j}$ determines n, and n is coupled into the equations for $\vec{u}$, c, and hence $P(\vec{V},\ldots)$, it is evident that we have a set of nonlinear, stochastic, coupled integro-partial differential equations. Simplification, sorting out the most crucial features, and performing the measurements that these processes suggest will take time - but then, the 21st Century is one hundred years long!

3 Bacteria in constraining environments (II)

During the last two decades, experiments have generated much insight into the locomotory apparatus of single bacterial cells. Relatively less attention has been given to problems and phenomena of group dynamics, hydrodynamics, and the statistical variability of individual cell behaviors. We present experimental results obtained with geometrically constrained populations of *Bacillus subtillis*. Such will be required for inclusion into new models of the dynamics of micro-organisms and their environment.

Bacteria swimming near a rigid plane, such as the surface of a microscope slide or a plastic Petri dish, tend not to reenter the main volume of fluid. Their trajectories remain approximately parallel to the plane. This effect may be due to combinations of Van der Waals and electrostatic forces, attraction due to inward volume conserving flows that surround a swimming cell, and a motion-generated torque that orients cells propelled from the rear toward nearby surfaces. This torque arises because, at the distal side, the drag on the moving body is less than at the side close to the rigid surface, where the fluid's velocity must vanish. It is convenient that one need not account statistically for projective foreshortening of velocity vectors when trajectories parallel bounding surfaces perpendicular to the axis of observation.

In a very shallow layer of water, ranging from zero, at the air/water/glass contact line, to a depth between air and glass of a few cell diameters, the organisms' trajectories are triply constrained. They interact with each bounding surface and with each other. They may swim singly, but often form platoon-like groups. In the shallowest regions, they often swim parallel to the contact line. Neither the hydrodynamics nor the microbiological

aspects of these phenomena have been modelled. We have measured relevant distributions of speed and swimming direction (see Figs. 4-9). Such SVPDF's must be included in collective models of individual behavior and hydrodynamics.

The fluid mechanics of micro-swimmers is characterized by Reynolds numbers $Re = Va/\mu << 1$ (see Appendix II for definitions and magnitudes). Nevertheless, linearity and reversibility need not be valid. When swimming changes boundary conditions, and when flow, and/or proximity of boundaries, affect swimming rate, style, and direction, neither linearity or reversibility of the equations remain. Some observations recently made in our laboratory [16], and others shown below, illustrate these remarkable phenomena.

The mean speed of the bacteria *Bacillus subtilis*, that swim in water whose depth h ranges from one to two bacterial body diameters d, is often more than double the speed of the bacteria in "deep" water. This effect may be explained by Ramia's [23] calculations that indicate a 10% greater propulsive efficiency per rotating flagellum, in the vicinity of a rigid surface. *B. subtilis* cells contain many flagella (e.g. 5-10).

L	Far	$36 \pm 10\mu m$(Fig. 4)	$16 \pm 10\mu m$
Speed	27.4 ± 0.3	35.8 ± 0.9	48.7 ± 1.1
Fastest Speed Segment $\mu m s^{-1}$	71.9	139	189

TABLE 21.1. Speeds of cells swimming near a glass surface as a function of distance L from the contact line. The depth of the water layer in which the cells swim increases with L. Data of Figs. 3-5, Method (b), see Appendix I.

Table 1 shows the average speeds and the fastest speed segment in a thin, wedge-shaped film of water. The film increases in thickness from "zero" at the air/glass/water contact line. These data are a summary of the speed averages in Figs. 5, 6, and 7, which show histograms of the speed of cells obtained using statistical data analyses a, b, and c (see Appendix I). It is noteworthy that mean speed and top speed both increase with decreasing thickness of the layer of water.

In a shallow wedge, where $h(x)$ decreases to zero and $\vec{x}$ is in the plane of one surface and perpendicular to the contact line, bacteria tend to swim perpendicular to dh/dx, often in nearly close packed groups. This effect implies that the directional uncertainty of $P(\vec{V}\ldots)$ can be quashed by interactions. This remarkable self-organization of swimming directions is illustrated in Fig. 8. Directional groups coalesce and disperse at various time intervals, as also observed by [19], and much as do pedestrians in corridors, as cited by Helbing [6].

The speed with which cells swim approximately parallel to the contact

Swimming Direction		Mean Speed	[max,min]	No. of Tracks
Parallel to	$60° - 120°$	51.2 ± 2.6	[151,15]	63
contact line	$240° - 300°$	52.5 ± 2.0	[115,21]	101
Swimming Away	$120° - 240°$	43.9 ± 2.0	[86,15]	51
Swimming Toward	$300° - 60°$	44.0 ± 1.9	[97,18]	61

TABLE 21.2. Anisotropy of bacterial swimming speed in a shallow wedge shaped layer of water bounded on one side by the air/water/glass contact line. Speeds are in $\mu m s^{-1} \pm S.E.M.$. The frame-by-frame per track method (b) is used for speed and angle measurements. The maximum and minimum speed encountered somewhere along some track in an entire sample is given by [,].

line is greater than the speed of swimmers in the perpendicular direction. This surprising phenomenon, summarized in Table 2, is probably connected with the statistical preference for swimming in the parallel direction, shown in Fig. 8.

For modelling purposes, one requires some knowledge of the primitive behaviors that are summarized by the histograms. If, for example, the cell population consists of cohorts, each with a definite, stable speed, the biological and computational situation is quite different from the "Maxwell-Boltzmann" case, where, in due course, each swimmer acquires all possible speeds. Figs. 9a and 9b show that the swimming speed of two randomly chosen bacteria fluctuates within wide limits, and the average speeds are quite different also. In the bioconvection situation, averaging over the SVPDF's, after the equations are solved for a particular $\vec{V}$, is inadmissible because of the nonlinearity of the equations. Fast cohorts will swim ahead, changing the environment for the slower cells. Solving the equations simultaneously for many $\vec{V}'s$ determined by the distribution functions is possibly a starting approach.

4 Possibilities for computer simulation

Digital techniques have been used to analyze various aspects of the experiments, e.g. the histograms of cell swimming. Partial solutions of the continuum pde's, [22], but without inclusion of the experimentally determined SVPDF's, have been computed. Results of computations for various approximations are summarized in [22], [8], and [4]. Simulations have begun that track individual model automata "swimming" up concentration gradients that they cause. Calculations without fluid dynamical effects [10], and calculational experiments that include fluid dynamics, but for 2 dimensional organisms [7], are a useful beginning. The latter work produces self-organized plumes rather similar to those observed in experiments. The continuation of this approach, using experimentally determined SVPDF's has much promise.

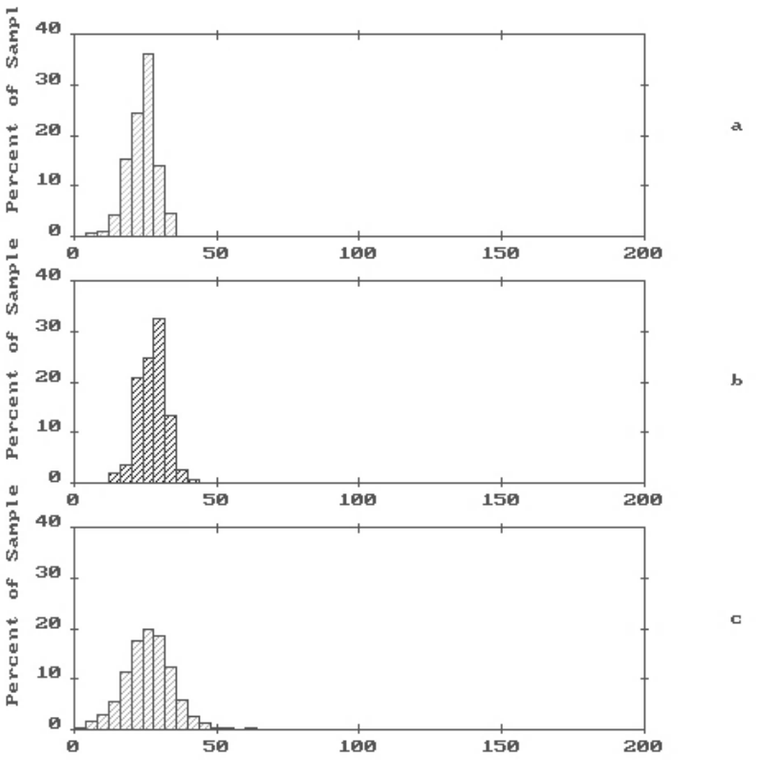

FIGURE 5. Swimming speeds of bacteria near a glass surface, in deep water far from the air/water/glass contact line. The three histograms pertain to one sample. They refer to the three methods described in the appendix. Units for mean speeds $\langle V(i)\rangle \pm$ standard error of the mean (S.E.M.) are in $\mu m s^{-1}$. (a) Head to tail, $\langle V(a)\rangle = 23.9 \pm 0.3$. (b) Frame-by-frame per track, $\langle V(b)\rangle = 27.4 \pm 0.3$. (c) Collective frame-by-frame, $\langle V(c)\rangle = 26.1 \pm 0.1$. This sample contained 217 tracks comprised of 5299 frames. The fastest track segment in (c) was $71.9 \mu m s^{-1}$. The distributions of swimming directions was approximately uniform.

The extraction of useful phenomenological data from bioconvection patterns has proven to be a significant challenge. However, as digital capability increases, we believe that better acquisition of statistical data from the patterns (e.g. mean populated area per unit time) will enrich the understanding of the global system through interactive computational correlation of observation, control, and analysis.

One issue that must be dealt with in any computational model is that, though there is similarity in the development of every pattern, the details of the experimentally observed evolution are different every time. This

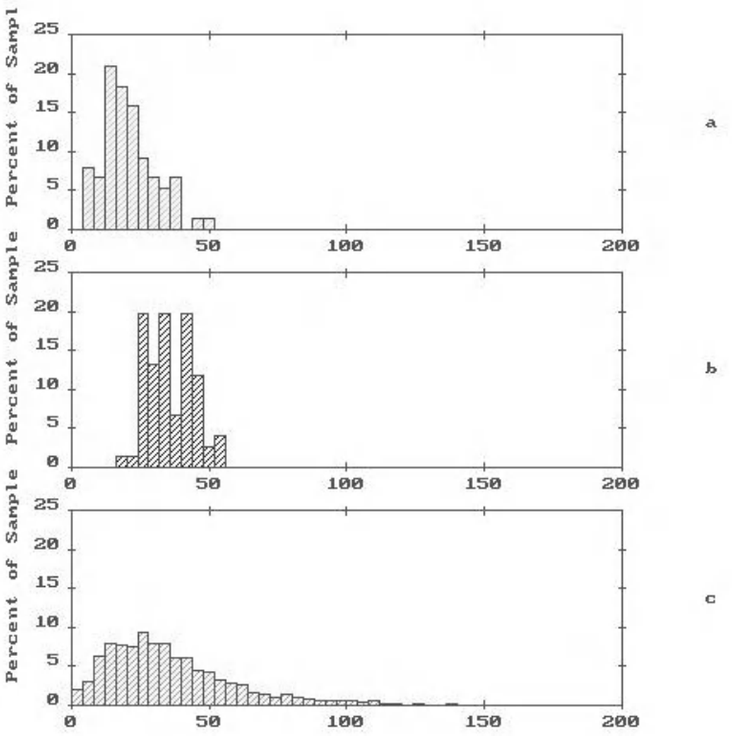

FIGURE 6. Swimming speeds of bacteria near a glass surface. The depth of water is $\geq 10\mu m$. The location where this sample was measured was between 26 and 46 μm from the air/water/glass contact line. Units and statistical methods as in Fig. 5. (a) Head to tail, $\langle V(a)\rangle = 20.9 \pm 1.1$. (b) Frame-by-frame per track, $\langle V(b)\rangle = 35.8 \pm 0.9$. (c) Collective frame-by-frame, $\langle V(c)\rangle = 35.4 \pm 0.6$. The sample contained 76 tracks comprised of 1239 frames. The fastest track segment was $139\mu m s^{-1}$. The angular distribution was approximately uniform.

stochasticity, and its dependence on the experimental initial conditions, represents both the physical and biological characteristics of the global system. Random events, such as a fast swimmer suddenly veering off in a new direction (the heretofore mentioned "free will"), drastically increase the nonlinearity of the global system. They result in a change of details, but not general morphology of the subsequent pattern. Inclusion of such events into 3-D digital simulations will be a major step in reaching insights by "recreating" the naturally occurring events with the computer.

An analysis of a population of the swimmers parallels analysis of the evolution of 2-D foams which "break" (i.e. individual cell walls rupture) when gently heated [1]. In that case, it was important to understand both

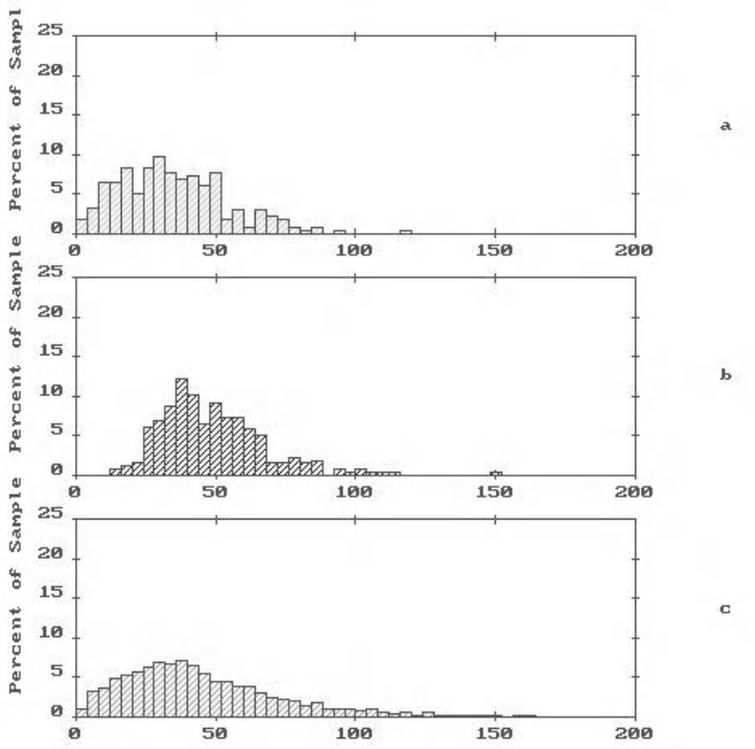

FIGURE 7. Swimming speeds of bacteria near a glass surface. The depth of water for this sample, located between 6 and 26 μm from the contact line, ranged from $\sim 1\mu m$ to $10\mu m$. Units and statistical methods as in Fig. 5. (a) Head to tail, $\langle V(a)\rangle = 34.7 \pm 1.1$. (b) Frame-by-frame per track, $\langle V(b)\rangle = 48.7 \pm 1.1$. (c) Collective frame-by-frame, $\langle V(c)\rangle = 44.8 \pm 0.5$. This sample contained 276 tracks comprised of 3143 frames. The fastest track segment was $189\mu m s^{-1}$.

the physico-chemical and random, but constrained, geometric properties of the foams at discrete times during the evolutionary process. By doing so, it became possible to generate a digital model of both the global and local behavior of the foam [5]. Once verified as being both physically and phenomenologically correct, this computer model became an invaluable tool with which to analyze the patterns created during foam evolution. The foams are analogous to the swimmers in that the patterns generated in each system appear quite similar, though their details differ for each dynamical evolution. It is important to note that a) the physics within the foam network does not change during the evolutionary process, and b) the foam has no "live" component which will exhibit "free will", as the bacterial

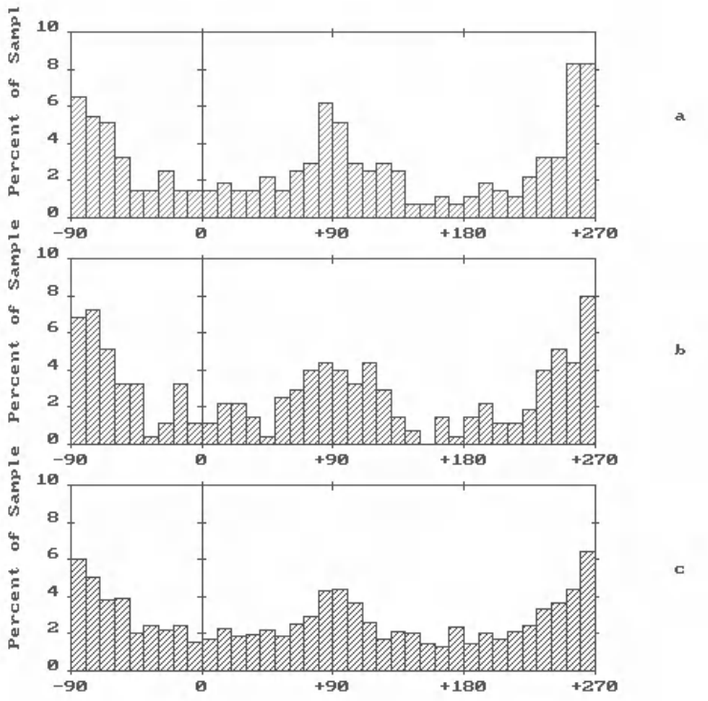

FIGURE 8. Distribution of directions for the sample of swimming trajectories whose speed analysis is shown in Fig. 6. Methods (a), (b), and (c) as before. The angular position of the contact line is $\pm 90^\circ$. The cells tend to swim parallel to it, in either direction.

groups do, but it is subject to external an internal random events.

5 Conclusion

Conventional microbiological research is concerned with the biochemistry, subcellular dynamics, and behavior of individual organisms. The amazingly complete insights that have been attained do not generally extend to fluid dynamical and other interactions among swimming cells that live within either convective or strongly constraining geometries. Yet, large interacting populations are the norm in the natural environment. The behavior of individuals determines, and is determined by, biological imperatives and associated behaviors of neighbors, by chance events, or — in appropriate settings — by vast segments of populations and their inadvertently created

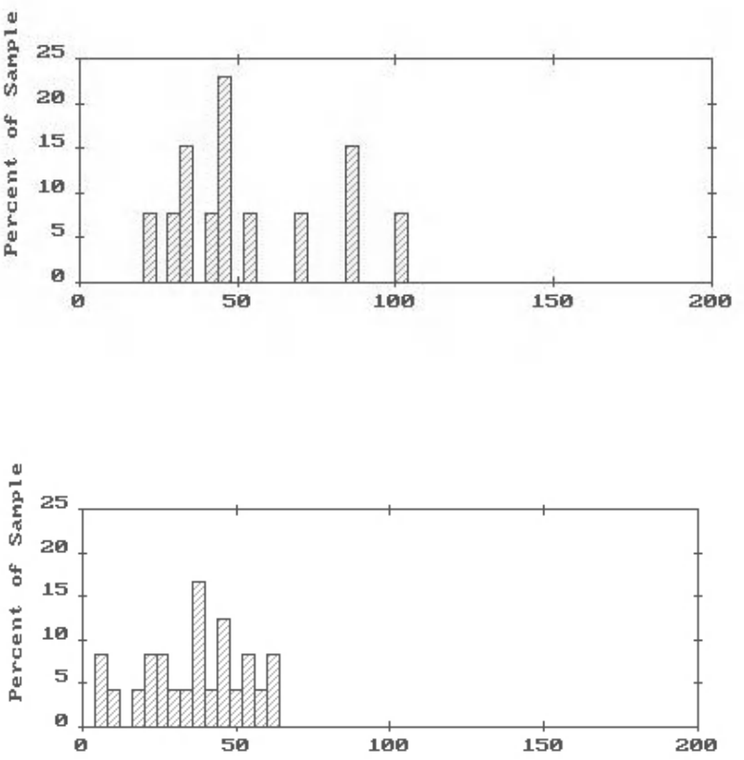

FIGURE 9. Swimming speeds of individual bacterial cells near a glass surface, and near the contact line. These histograms are for two tracks from the sample of Figs. 7 and 8, randomly chosen: (Top) from the group of tracks oriented within $\pm 30^\circ$ of the contact line, and (Bottom) within $\pm 60^\circ$ of the perpendicular to the contact line. These track histograms illustrate the typical variability of speeds along any given trajectory.

fields of fluid motion.

This paper brings into focus some of the physical, biological, and mathematical ideas needed for modelling these systems. The experimental data presented are intended to demonstrate that some of the required underlying information, at all scales, is novel, astonishing, and within reach. These observations relate not only to biological dynamics and its mysteries, but to fundamental issues in the physics of dynamical systems.

How ought one interpret and model convection that exhibits macroscopic

chaos and reproducible categories of patterns, while the underlying driving mechanism, the directional swimming of cells, must be described by probability density functions? At the scale of bacterial cells, interactions among individuals, also between individuals and bounding surfaces, circumvent the usual assumptions of linearity and reversibility governing viscous flow phenomena. New discoveries concerning flow at very small scales [9] may relate. Can bacteria be used as micro-probes of physical phenomena? How ought such effects be incorporated in the modelling equations? The appearance of evanescent groups of fast swimming, directionally coherent cells in shallow wedge-shaped layers of fluid is reminiscent of the "cybotactic swarms" featured in the ancient literature concerning liquid crystals. Is there a self-generated mean director field? Are van der Waals or immersion forces [11] responsible for weak aligned bound states? The statistical data suggest that neither the central limit theorem, nor ergodicity, apply. What are the broader implications, for biology, and for local as well as convective transport-related nonlinearities?

In the context of this volume, we conclude that the subject is exciting, applicable, nonlinear, and sufficiently nontrivial to extend well into the next millennium.

Acknowledgments: This research was supported by grant DMR-9812526 from the National Science Foundation. Generous support was provided by Alice and Ralph Sheets, through a gift to the University of Arizona Foundation.

We also wish to thank D. Wiseley for his assistance with the gathering and analysis of data, and N.H. Mendelson for many illuminating discussions.

6 Appendix I: Statistical methods

Video tapes of cells swimming in shallow layers of water were analyzed, using a computer system supplied by the Motion Analysis Corporation (Santa Rosa, CA). Speed and direction of a given trajectory, and statistics of many trajectories, were analyzed by three methods. When a given trajectory begins at $(\vec{x}_1, t_1)$ and its recording is terminated at $(\vec{x}_2, t_2)$, one may take

$$\frac{|\vec{x}_2 - \vec{x}_1|}{(t_2 - t_1)} \tag{21.10}$$

as an estimate of the speed V, and the direction of $\vec{x}_2 - \vec{x}_1$ as a direction of travel. This head-to-tail method "a" is performed on all tracks in the sample.

The speed and direction increments measured frame-by-frame along a trajectory using

$$\frac{|\vec{x}_{i+1} - \vec{x}_i|}{(t_{i+1} - t_i)} \tag{21.11}$$

estimates the actual speed and direction in the time interval of one frame, e.g. 1/30 *sec.* for standard (NTSC) video. The average of these measurements $\langle V_i \rangle$ and $\langle \theta_i \rangle$ is calculated for a particular track. When these data are collected for all tracks and plotted as histograms, they constitute method "b", "frame-by-frame per track". Method "c" generates histograms for all frame-by-frame elements, for all tracks, without intermediate averaging. This "collective frame-by-frame" method shows the overall distribution and range of local velocities.

Methods (a) and (b) relate to statistics for individual tracks. The ratio of means,

$$S \equiv \frac{\langle V(a) \rangle}{\langle V(b) \rangle} \leq 1 \tag{21.12}$$

is a measure or tortuosity, where $S = 1$ implies straight trajectories; Brownian motion would produce $S \simeq 0$.

7 Appendix II

Typical dimensions of *Bacillus subtilis* or *Escherichia coli*:

Bacterial body: Length 1-5 μm, diameter $d \approx 0.7 \mu m$.
Helical flagella: Length 1-10 μm, diameter $\sim 15 nm$. Helix pitch $2 - 3 \mu m$, helix diameter $\sim 0.5 \mu m$.

Swimming speeds:

Average $V \leq 20 \mu m s^{-1}$ in deep layers, far from surfaces
$V \leq 50 \mu m s^{-1}$ in shallow layers of water.
Spurts $\geq 100 \mu m s^{-1}$

Reynolds number $= \frac{aV}{\nu} = \frac{(dimension) \times (speed)}{kinematic\ viscosity}$

For bacteria:
Normal, average case: $\frac{(3 \times 10^{-4}) \times (3 \times 10^{-3})}{10^{-2}} = 9 \times 10^{-5} << 1$
Choice of biggest parameters: $\frac{(10 \times 10^{-4}) \times (10 \times 10^{-3})}{10^{-2}} = 10^{-3} << 1$
(Units: cm, sec)

Rate of oxygen consumption $\approx 10^6$ molecules per cell per second [2].

Diffusivity of oxygen $= 2 \times 10^{-5} cm^2 s^{-1}$

Kinematic viscosity of water $\nu \equiv \frac{\mu}{\rho} \approx 10^{-2} cm^2 s^{-1}$, with $\rho \approx 1 gm cm^{-3}$.

Kinetic theory based estimate for diffusivity of swimming cells:

For $10 \mu m$ free path, speed of $30 \mu m s^{-1}$

Diffusivity of swimming cells:

$D(cell) \approx \frac{(10^{-3}) \times 3 \times (10^{-3})}{3} = 10^{-6} cm^2 s^{-1}$.

8 References

[1] G.D. Burnett, J.J. Chae, W.Y. Tam, R.M.C. de Almeida, and M. Tabor. Structure and dynamics of breaking foams. *Phys. Rev. E*, 51(6):5788–5796, 1995.

[2] H.C. Berg. *Random Walks in Biology.* Cambridge University Press, Cambridge, 1983.

[3] M.A. Bees and N.A. Hill. Wavelengths of bioconvection patterns. *J. Exp. Biol.*, 200:1515–1526, 1997.

[4] M.A. Bees and N.A. Hill. Linear bioconvection in a suspension of randomly swimming, gyrotactic micro-organisms. *Phys. Fluids*, 10:1864–1881, 1998.

[5] J.J. Chae and M. Tabor. Dynamics of foams with and without wall rupture. *Phys. Rev. E*, 55(1):598–610, 1997.

[6] D. Helbing. Traffic modelling by means of physical concepts. In D.E. Wolf et al., editors, *Traffic and Granular Flow*, pages 87–117. World Scientific, Singapore, 1996.

[7] M. Hopkins. *Computational Modeling of the Collective Fluid Dynamics of Motile Microorganisms.* PhD thesis, Department of Mathematics, Tulane University, 1997.

[8] A.J. Hillesdon, T.J. Pedley, and J.O. Kessler. The development of concentration gradients in a suspension of chemostatic bacteria. *Bull. Math. Biol.*, 57:299–344, 1995.

[9] C.M. Ho and Y.C. Tai. Micro-electro-mechanical systems (MEMS) and fluid flows. *Annu. Rev. Fluid Mech.*, 30:579–612, 1998.

[10] I.M. Jánosi, J.O. Kessler, and V.K. Horváth. The onset of bioconvection in suspensions of *Bacillus subtilis. Phys. Rev. E*, 58:4793-4800, 1998.

[11] P.A. Kralchevsky, K.D. Danov, and N.D. Denkov. Chemical physics of colloid systems and interfaces. In K.S. Birdi, editor, *Handbook of Surface and Colloid Chemistry*, pages 333–494. QRC Press, Boca Raton, 1997.

[12] J.O. Kessler. Cooperative and concentrative phenomena of swimming microorganisms. *Contemporary Physics*, 26:147–166, 1985.

[13] J.O. Kessler and N.A. Hill. Complementarity of physics, biology, and geometry in the dynamics of swimming micro-organisms. In H. Flyvbjerg et al., editor, *Physics of Biological Systems: From Molecules to Species*, pages 325–340. Springer, Heidelberg, 1997.

[14] J.O. Kessler, R.P. Strittmatter, D.L. Swartz, D.A. Wiseley, and M.F. Wojciechowski. Paths and patterns: The biology and physics of swimming bacterial populations. In C.P. Ellington and T.J. Pedley, editors, *Biological Fluid Dynamics*, number XLIX, pages 503–526. Symposia of the Society for Experimental Biology, 1995.

[15] J.O. Kessler and M.F. Wojciechowski. *Bacteria as Multicellular Organisms*, chapter Collective Behavior and Dynamics of Swimming Bacteria, pages 415–417. Oxford University Press, New York, 1997.

[16] J.O. Kessler, D.A. Wiseley, K.E. Remick, and D.E. Marthaler. Individual and collective dynamics of swimming bacteria. In *Proceedings: Traffic and Granular Flow*, M. Schreckenberg and D.E. Wolf, Eds, pages 37 - 50. Springer, Singapore, 1998.

[17] T.M. Lenton. Gaia and natural selection. *Nature*, 394:439–447, 1998.

[18] M.D. Levin, C.J. Morton-Firth, W.N. Abouhamad, R.B. Bourret, and D. Bray. Origins of individual swimming behavior in bacteria. *Biophys. J.*, 74:175–181, 1998.

[19] N.H. Mendelson, 1997, Private communication. Also, N. H. Mendelson, A. Bourque, K. Wilkening, K. R. Anderson, and J. C. Watkins. Organized cell swimming motions in *Bacillus subtilis* colonies: Patterns of short-lived whirls and jets. J. Bact,181:600-609, 1999.

[20] N.H. Mendelson and J. Lega. A complex pattern of travelling stripes is produced by swimming cells of *Bacillus subtilis*. *J. Bact.*, 180:3285–3294, 1998.

[21] T.J. Pedley and J.O. Kessler. A new continuum model for suspensions of gyrotactic microorganisms. *J. Fluid Mech.*, 212:595–613, 1990.

[22] T.J. Pedley and J.O. Kessler. Hydrodynamic phenomena in suspensions of swimming microorganisms. *Annu. Rev. Fluid Mech.*, 24:313–358, 1992.

[23] M. Ramia, D.L. Tullock, and N. Phan-Thien. The role of hydrodynamic interaction in the locomotion of microorganisms. *Biophys. J.*, 65:755–778, 1993.

[24] M.A.S. Vigeant and R.M. Ford. Interactions between motile *E. coli* and glass. *Appl. Environ. Microbiol.*, 63:3474–3479, 1997.

22

Nonlinearities in Biology: The Brain as an Example

Hermann Haken

1 Introduction

The human brain is probably the most complex system we know. The number of its individual parts, the neurons, is estimated to be one hundred billions. A neuron can be connected with up to ten thousand other neurons. Other characteristic features of neurons have long been studied, and this field is still subject of intense investigations both experimental and theoretical [1], [3], [11], [30]. At any rate it is well-established that neurons are highly nonlinear elements. At the same time it is evident from our daily experience that these neurons must cooperate in a well coordinated fashion, be it the steering of movements or processes of pattern recognition, speech production, and so on. This high coordination becomes also macroscopically visible through MEG and EEG measurements of brain activity under different circumstances [19].

In my article I wish to address two questions that seem to be quite fundamental, namely how can we model the behavior of the individual elements and their interaction among each other, and how is it that in spite of the extremely complicated interactions, at a macroscopic level some kind of coherent structures, be it in the form of the electric and magnetic fields of the brain, be it in the form of perceived patterns and so on, emerge.

2 Some salient features of neurons

A neuron can be conceived as an element that receives inputs from other neurons, processes them and then sends out outputs to other neurons. The scheme is shown in Fig. 1.

The soma of a neuron receives its inputs through dendrites in the form of electric currents. It then sums up over these inputs, which might be excitatory or inhibitory, and once a certain threshold is reached it fires. This means, it sends out short pulses along its axon, which then splits up and connects with the dendrites of other neurons via synapses. Once the electric signal carried by the pulse reaches a synapse, it causes vesicles to open and to release neurotransmitters that diffuse across the synaptic cleft

P.L. Christiansen, M.P. Sørensen, and A.C. Scott (Eds.): LNP 542, pp. 427−445, 2000.

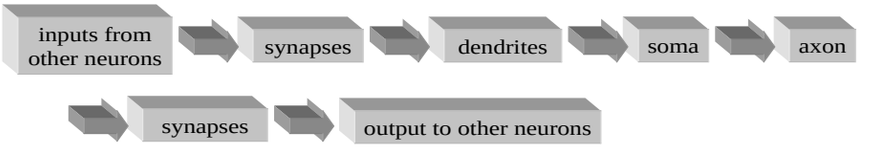

FIGURE 1. Block diagram of neuronal connections and activities

and eventually cause currents in the dendrites of other neurons. Of this conventional wisdom, two features are to be noted:

1. The messages sent out from a neuron via the axon are pulse-coded, i.e. the signal is encoded by means of time intervals between the individual pulses.

2. These time intervals may be rather noisy.

While this pulse code seems to be universal for all neurons, there may be still individual differences, for instance in how the incoming signals are transformed into the outgoing ones. Here we focus our attention on one kind of neurons that are well studied experimentally with respect to the visual cortex. These results are based on experiments with anesthetized cats to which a bar with a specific orientation and direction of movement is shown. Depending on these features, the neurons fire more or less rapidly. Following related results by W. Freeman [10] on the olfactory bulb, an important feature was discovered more recently [8], [13]. When two bars move across the visual fields, two corresponding neurons (or ensembles of them) fire in phase, i.e. phase-locking between nerve pulses occurs. Since the locking occurs only for a certain time interval, the results are described as spindles.

In the first part of this article I shall describe a model that I recently developed and that takes into account these experimental facts. It thus adopts a middle position between two well-known extreme models. The one widely known model is that of McCulloch and Pitts [27], which assumes that the neurons have only two states, one resting state and one firing state. The firing state is reached when the sum of the inputs from other neurons exceeds a certain level. Then the neuron fires, but in this model only once. The other case is represented by modelling neurons by means of the Hodgkin-Huxley model. This model was originally devised to understand the properties of axonal pulses, but is now applied to the generation of pulse trains by neurons [4]. For further references the reader is referred to section 8.

3 The noisy lighthouse model of a neural network

We first consider the generation of dendritic currents by means of axonal pulses via the synapses. We formulate the corresponding equation for the dendritic current ψ as follows:

$$\dot{\psi}(t) = aP(t-\tau) - \gamma\psi(t) + F_\psi(t) \tag{22.1}$$

Here and in the following a dot means the time-derivative, i.e. $\dot{\psi} = d\psi/dt$. In a first approximation we assume that a is a constant. P is the axonal pulse, τ a time delay, γ is a decay constant and F_ψ is a fluctuating force. As is known from statistical physics, whenever there occurs damping, fluctuating forces are present. As usual we shall assume that the fluctuating forces are δ-correlated in time. A more general relationship between pulses and the generation of dendritic currents is given by the equation

$$\dot{\psi}(t) = \int_{-\infty}^{t} G(t,\sigma)P(\sigma)d\sigma - \gamma\psi(t) + F_\psi(t). \tag{22.2}$$

As is known, vesicles that release neurotransmitters and thus eventually give rise to the dendritic current can spontaneously open. This will be the main reason for the fluctuating force F_ψ, but also other noise sources may be considered here. When a pulse comes in, the opening of a vesicle occurs with only some probability. Thus we have to admit that in a more appropriate description a is a randomly fluctuating quantity. Similarly the kernel G in (22.2) may contain such fluctuations. While F_ψ in (22.1) and (22.2) represent additive noise, a and G represent multiplicative noise. In order to describe the pulses properly, we introduce a phase angle ϕ and connect P with ϕ through a function f

$$P(t) = f(\phi(t)). \tag{22.3}$$

We require the following properties of f

$$a) \quad f(0) = 0, \tag{22.4}$$

$$b) \quad f(\phi + \phi_0) = f(\phi), \quad \text{periodic}, \tag{22.5}$$

$$c) \quad \text{sharply peaked}. \tag{22.6}$$

Finally we have to establish a relationship between the phase angle ϕ of the pulse P produced by the neuron under consideration and the dendritic currents. To this end, we write

$$\dot{\phi}(t) = S(X) + F_\phi(t), \tag{22.7}$$

where the function $S(X)$ has the following properties: S is equal to zero for X smaller than a threshold Θ, then it increases in a quasi linear fashion until it saturates. Using the dendritic currents of other neurons ψ_m, we write S in the form

$$S(X) = S\left(\sum_m c_m \psi_m(t-\tau') + p_{\text{ext}}(t-\tau'') - \Theta\right). \tag{22.8}$$

Here, τ' and τ'' are delay times, and p_{ext} is an external signal that is transferred to the neuron under consideration from sensory neurons. A simple explicit representation of (22.8) obeying the properties just required for S is given by

$$\left.\begin{aligned} \dot{\phi} &= \sum_m c_m \psi_m(t-\tau') + p_{\text{ext}}(t-\tau'') - \Theta \quad \text{for} \quad S > 0 \\ \dot{\phi} &= 0 \\ \phi &= 0 \end{aligned}\right\} \quad \text{otherwise.} \tag{22.9}$$

The interpretation of equation (22.7) is based on the functioning of a lighthouse in which a light beam rotates. The rotation speed $\dot{\phi}$ depends on S according to (22.7). The fluctuating forces F_ϕ lead to a shift of the phase at random instances. The relationships (22.1) or (22.2), (22.3) and (22.8) can be easily generalized to the equations of a whole network. The index m or k refers to the location and to the property "excitatory" or "inhibitory". The generalizations are straight forward and read

$$\dot{\psi}_m(t) = \sum_k a_{mk} P_k(t-\tau) - \gamma \psi_m(t) + F_{\psi,m}(t), \tag{22.10}$$

$$P_k(t) = f(\phi_k(t)), \tag{22.11}$$

$$\dot{\phi}_k(t) = S\left(\sum_m c_{km} \psi_m(t-\tau') + p_{\text{ext},k}(t-\tau'') - \Theta_k\right) + F_{\phi,k}(t). \tag{22.12}$$

4 The special case of two neurons

In order to elucidate the meaning of these equations, we consider the special case of two neurons. In this case, equations (22.10) and (22.12) acquire the form

$$\dot{\psi}_1 = a f\left(\phi_2(t-\tau)\right) - \gamma \psi_1, \tag{22.13}$$

$$\dot{\phi}_2 = S\left(\psi_2(t-\tau') + p_{\text{ext},2} - \Theta\right), \tag{22.14}$$

$$\dot{\psi}_2 = af\left(\phi_1(t-\tau)\right) - \gamma\psi_2, \tag{22.15}$$

$$\dot{\phi}_1 = S\left(\psi_1(t-\tau') + p_{\text{ext},1} - \Theta\right). \tag{22.16}$$

In these equations, we neglect fluctuations. If the relation

$$p_{\text{ext},1} = p_{\text{ext},2} \tag{22.17}$$

holds, for symmetry reasons among the possible solutions, there is the following:

$$\psi_2 = \psi_1, \qquad \phi_2 = \phi_1. \tag{22.18}$$

(22.18) implies that the phase locked solution is a possible one. Using (22.18), equations (22.13) - (22.16) simplify to

$$\dot{\psi}_1 = af\left(\phi_1(t-\tau)\right) - \gamma\psi_1 \tag{22.19}$$

and

$$\dot{\phi}_1 = S\left(\psi_1(t-\tau') + p_{\text{ext}} - \Theta\right). \tag{22.20}$$

Now let us assume that both neurons operate in the linear regime according to (22.9). Equation (22.20) then acquires the form

$$\dot{\phi}_1 = b\left(\psi_1(t-\tau') + p_{\text{ext}} - \Theta\right), \tag{22.21}$$

where b is a constant. Note that the equations (22.19) and (22.21) are still highly nonlinear because of the function f. In order to make our reasoning as transparent as possible, we put

$$a = b = 1. \tag{22.22}$$

It is a simple matter to eliminate ψ from the equations (22.19) and (22.21), which leads to

$$\ddot{\phi}_1 = f\left(\phi_1\right) - \gamma\dot{\phi}_1 + \gamma p_{\text{ext}} - \gamma\Theta \equiv -\frac{\partial V(\phi_1)}{\partial\phi_1} - \gamma\dot{\phi}_1. \tag{22.23}$$

As is indicated by the last part of this equations, the first and the last two terms in the middle part can be expressed as the derivative of a potential function that consists of two parts. As a simple analysis reveals, the first part f gives rise to a potential function that consists of decreasing steps when considered as a function of ϕ, whereas the last two terms give rise to a linearly decreasing function. The motion of the phase ϕ_1 can thus be

interpreted as the motion of a particle in this potential field, i.e. the particle is sliding down at a constant speed on the average on which, however, jumps are superimposed. The effect of noise can easily be considered, too, namely it gives rise to a diffusion of the phase ϕ_1.

Let us now consider the stability of phase locking. To this end we again consider the linear regime of S in which case the equations (22.13) - (22.16) can be simplified to

$$\dot{\psi}_2 = f(\phi_1) - \gamma\psi_2, \tag{22.24}$$

$$\dot{\phi}_1 = \psi_1 + p_{\text{ext}} - \Theta, \tag{22.25}$$

$$\dot{\psi}_1 = f(\phi_2) - \gamma\psi_1, \tag{22.26}$$

$$\dot{\phi}_2 = \psi_2 + p_{\text{ext}} - \Theta. \tag{22.27}$$

Again we can eliminate ψ_1 and ψ_2, which leads to

$$\ddot{\phi}_1 = f(\phi_2) - \gamma\dot{\phi}_1 + \gamma p_{\text{ext}} - \gamma\Theta \tag{22.28}$$

and

$$\ddot{\phi}_2 = f(\phi_1) - \gamma\dot{\phi}_2 + \gamma p_{\text{ext}} - \gamma\Theta. \tag{22.29}$$

In order to make our approach as simple as possible, we use the following explicit form of f

$$f(\phi) = a\dot{\phi}\sum_n \delta(\phi - 2\pi n - \epsilon), \quad \epsilon \quad \text{small}. \tag{22.30}$$

We further introduce the equation for the phase-locked state ϕ

$$\ddot{\phi} = f(\phi) - \gamma\dot{\phi} + \gamma p_{\text{ext}} - \gamma\Theta. \tag{22.31}$$

Making the hypothesis

$$\phi_j = \phi + \xi_j; \quad j = 1, 2, \tag{22.32}$$

subtracting (22.31) from (22.29) and integrating over time yields

$$\dot{\xi}_2 + \gamma\xi_2 = -a\left[\phi - \phi \bmod 2\pi - ((\phi + \xi_1) - (\phi + \xi_1) \bmod 2\pi)\right]. \tag{22.33}$$

A somewhat lengthy but rather straightforward analysis reveals that for small ξ_j's (22.33) can be transformed into

$$\dot{\xi}_2 + \gamma\xi_2 = 2\pi a\dot{\phi}^{-1}\sum_\ell \delta\left(t - t_\ell^+\right)\xi_1, \tag{22.34}$$

where $\phi(t_\ell^+) = 2\pi\ell$. Similarly we obtain

$$\dot{\xi}_1 + \gamma\xi_1 = 2\pi a\dot{\phi}^{-1} \sum_\ell \delta\left(t - t_\ell^+\right) \xi_2. \tag{22.35}$$

Addition or subtraction of (22.34) and (22.35) leads to two uncoupled equations for $\xi = \xi_2 - \xi_1$ and $\sum = \xi_1 + \xi_2$, e.g.

$$\dot{\xi}(t) + \gamma\xi(t) = -2\pi a\dot{\phi}^{-1} \sum_\ell \delta\left(t - t_\ell^+\right) \xi(t). \tag{22.36}$$

In between the peaks of the δ-functions ξ shows an exponential decay that is superimposed by jumps of ξ at the δ-peaks. Provided γ is large enough, ξ (and $\sum$) decay to zero in spite of the jumps so that in this case the phase-locked state is stable.

5 Time-averages

When we write $P(t)$ in the form

$$P(t) = p_0 \sum_j \delta\left(t - t_j\right), \tag{22.37}$$

where t_j are the times where the pulses occur, and p_0 is the pulse strength, then

$$\frac{1}{Tp_0} \int_{t-T/2}^{t+T/2} P(\sigma)d\sigma \equiv A(t) \tag{22.38}$$

can be interpreted as the number of pulses of height p_0 during the time T. Such averages can be introduced in the equation for the dendritic currents (22.1). The solution of that equation reads

$$\psi(t) = a \int_{-\infty}^{t} e^{-\gamma(t-\sigma)} P(\sigma - \tau)d\sigma + \hat{F}_\psi(t), \tag{22.39}$$

where we may consider

$$1/\gamma = T \tag{22.40}$$

as a time-window. By means of the definition (22.38) we may cast (22.39) into the form

$$\psi(t) = ap_0\gamma T \cdot A(t - \tau) + \hat{F}_\psi(t). \tag{22.41}$$

In order to introduce averages into equation (22.7), we consider a quasi-steady state, so that

$$t_j = jt_0, \quad j = 0, \pm 1, ... \tag{22.42}$$

and we conclude from

$$\dot{\phi} = C \tag{22.43}$$

that

$$\phi = Ct_0 \mathrel{\hat{=}} 1 \quad \text{pulse.} \tag{22.44}$$

The notation "$\hat{=}$" means "corresponding to". More generally we can cast equation (22.7) with (22.8) into

$$\frac{1}{T}\int_{t-T/2}^{t+T/2} \dot{\phi} d\sigma = \frac{1}{T}\int_{t-T/2}^{t+T/2} S\left(\sum_m c_m \psi_m(\sigma - \tau') + p_{\text{ext}} - \Theta\right) d\sigma \tag{22.45}$$

and using (22.38) further into

$$A(t) \approx S\left(\sum_m c_m \frac{1}{T}\int_{t-T/2}^{t+T/2} (\psi_m(\sigma - \tau') + p_{\text{ext}}) d\sigma - \Theta\right). \tag{22.46}$$

When we add the corresponding indices, referring to the individual neurons, we finally obtain well-known equations that we shall discuss in the next section.

6 The averaged neural equations

In this section we resume the general approach of section 3 that resulted in the equations (22.10) - (22.12). We now perform the time-average over these equations according to section 5. In order to allow a description that covers a number of special cases, we shall adopt a slightly different notation. We denote the axonal pulse rates at time t by

$$A_{jk}(t), \tag{22.47}$$

where the first index j refers to excitatory or inhibitory, and the second index enumerates the neurons. Later we will adopt a continuous space variable, and we shall then make the identifications

$$\begin{aligned} j &= e \quad \text{(excitatory)}, \quad k = x \\ &= i \quad \text{(inhibitory)}, \quad k = x. \end{aligned} \tag{22.48}$$

The dendritic wave amplitudes will be denoted by

$$\psi_{\ell,m}(t), \tag{22.49}$$

where in analogy to (22.48) we can specify the indices ℓ, m according to

$$\begin{aligned} \ell &= e \quad \text{(excitatory)}, \quad m = x \quad : \quad \psi_e(x,t) \\ \ell &= i \quad \text{(inhibitory)}, \quad m = x \quad : \quad \psi_i(x,t). \end{aligned} \tag{22.50}$$

The conversion of the axonal pulse rates into dendritic current amplitudes is described by

$$\psi_{\ell,m}(t) = \sum_k f_{\ell;m,k} a_\ell A_{\ell,k}\left(t - t_{\ell,mk}\right) + F_{\ell,m}(t). \tag{22.51}$$

$f_{\ell;m,k}a_\ell$ describes the synaptic strengths between neurons m and k, where we assume that inhibitory (excitatory) neurons have only inhibitory (excitatory) connections. $t_{\ell,mk}$ describes the time delay according to finite propagation velocities v_e, v_i. The conversion from dendritic amplitudes to axonal firing rates is described by the equations

$$\begin{aligned} A_{jk}(t) = \hat{S}_j\Big(\sum_m \hat{f}_{j;k,m} \left(\psi_{e,m}\left(t - t_{e,km}\right) - \psi_{i,m}\left(t - t_{i,km}\right)\right) \\ + p_{j,m}\left(t - t_{km}\right)\Big) + \hat{F}_{\ell,m}(t), \end{aligned} \tag{22.52}$$

where the notation is similar to that of equation (22.51). $p_{j,m}$ is an external signal of kind j at neuron m and $\hat{F}$ are again fluctuating forces. We assume that the conversion from dendritic currents to axonal pulse rates is local so that with *locally averaged* ψs

$$A_{jk}(t) = \hat{S}_j\left(\psi_{e,k}(t) - \psi_{i,k}(t) + p_{j,k}(t)\right) + \hat{F}_{j,m}(t). \tag{22.53}$$

Using the approximation (22.53), we can simply eliminate either A from the equations (22.51) and (22.53) or the ψs. When we eliminate the ψs, we obtain

$$\begin{aligned} A_{jk}(t) &= \hat{S}_j\Big(\sum_n f_{e;k,n} a_e A_{e,n}\left(t - t_{e,nk}\right) - \sum_n f_{i;k,n} a_i A_{i,n}\left(t - t_{i,nk}\right) \\ &\quad + p_{j,k}(t) + F_{e,k}(t) - F_{i,k}(t)\Big) + \hat{F}_{j,k}(t). \end{aligned} \tag{22.54}$$

When we ignore the fluctuating forces and use an ensemble average that allows us to introduce connectivities that are smooth functions of the space variable, equations (22.54) are essentially the Wilson-Cowan [36], [37] equations provided the membrane decay time is short compared to the propagation times. We note that the Wilson-Cowan model has found applications to the modelling of hallucinations as was done by Cowan and Ermentrout [7], [9] and more recently in a very extensive study by Tass [33], [35].

The formulation (22.54) allows us immediately to make contact with approaches in neural nets. We consider only one kind of neurons, namely excitatory, and make the replacement

$$A_{ek}(t) \to q_k(t), \tag{22.55}$$

$$A_{ik} \to 0, \tag{22.56}$$

and ignore the fluctuating forces, i.e. we put

$$F = \hat{F} = 0. \tag{22.57}$$

Furthermore we put the delay times equal to a single time according to

$$t_{e,nk} = \tau. \tag{22.58}$$

Without loss of generality, we may put

$$a_e = 1. \tag{22.59}$$

With these assumptions, (22.54) is converted into

$$q_k(t+\tau) = \hat{S}\left(\sum_n f_{kn} q_n(t) + p_k(t)\right). \tag{22.60}$$

Equation (22.60) is the conventional network model in the sense of McCulloch and Pitts [27]. (For a physicist the spin glass analogy by Hopfield [20] is most appealing, but we shall not dwell on it here.) f_{kn} serve as synaptic strengths. When the sigmoid function is a step function, having the values 0 or 1, with a threshold h, (22.60) describes a discrete time process during which the variables q_k that may be interpreted as neuronal activity, adopt the values 0 or 1. One may easily supplement equations (22.60) by a damping term

$$-\alpha q(t) \tag{22.61}$$

on the right-hand side. The reader hardly need be reminded of the contents of equation (22.60). The network consists of neurons that can occupy the states

$$q_k = 0, 1. \tag{22.62}$$

The inputs into neuron k coming from neurons n are summed up by means of synaptic weights f_{kn}. Depending on the size of the total sum, the neuron with index k may change its state or remain in the former one. It is important to note that in this model the geometrical form of the network is entirely unimportant. What counts is the connections. Kohonen's idea of feature maps [25] brings equations (22.60) closer to specific spatial arrangements. However, there is no link with the spatio-temporal patterns found in EEG or MEG measurements.

To elucidate this case, we again start from equations (22.51) – (22.53) and eliminate the axonal pulse rates so that we obtain equations for the dendritic wave amplitudes. A simple calculation shows that these equations have the form

$$\begin{aligned}\psi_{\ell,m}(t) = \sum_k a_\ell f_{\ell;m,k} \hat{S}_\ell(\psi_{e,k}\,(t - t_{\ell,mk}) \\ -\psi_{i,k}\,(t - t_{\ell,mk}) + p_{\ell,k}\,(t - t_{m,k})),\end{aligned} \tag{22.63}$$

where we omit the fluctuating forces. As we shall see immediately, equations (22.63) are a good starting point for deriving field equations of brain activity. In order to do so, we identify the indices m or k with the space point x, which may be one-, two-, or three-dimensional. It may well be that the synaptic strengths $f_{\ell;m,k}$ are extremely complicated functions of the space coordinates m, k. In order to proceed further, we shall assume that we average over ensembles of neurons so that the functions f become smoothed.

In the following then we shall make simple assumptions on the form of these smooth functions. We shall assume that the neuron at position x receives inputs from dendrites X more strongly, if the distance is small, or, in other words, we shall assume that f decreases with increasing distance. In order to make the following approach as transparent as possible, we shall assume that we may put $\psi_i = 0$, which actually is a good approximation and its inclusion would lead only to some kind of changes of numerical factors [23], [24]. Then equations (22.63) can be specified as equation

$$\begin{aligned}\psi_e(x,t) = a_e \int dX f_e(x,X) \hat{S}_e \Big[\psi_e \left(X, t - \frac{|\, x - X\,|}{v_e}\right) \\ + p_e \left(X, t - \frac{|\, x - X\,|}{v_e}\right)\Big],\end{aligned} \tag{22.64}$$

where we replaced the sum over k by an integral over X, and we made the replacement

$$f_{e,m,k} \to f_e(x, X) = \beta(x - X). \tag{22.65}$$

Equation (22.64) had been derived previously [23], [24] in a way where continuous space variables were used from the very beginning and the explicit form (22.31) of β was used. In its linearized form, this integral equation corresponds to the model equation by Nunez [28] and is discussed in [30]. (Nunez included the equation for ψ_i!) Equation (22.64) is the starting point for our further considerations (see also [23], [24]). It will be our goal to convert the integral equation (22.64) into a local field equation. By use of the δ-function

$$\delta\left(t - T - \mid x - X \mid / v_e\right), \tag{22.66}$$

we can write equation (22.64) in the form of equation

$$\psi_e(x, t) = \int dX \int_{-\infty}^{+\infty} G(x - X, t - T)\rho(X, T)dT, \tag{22.67}$$

where

$$G(x - X, t - T) = \delta\left(t - T - \frac{\mid xX \mid}{v_e}\right)\beta(x - X) \tag{22.68}$$

can be interpreted as a Green function, and

$$\rho(X, T) = a_e \hat{S}\left[(\psi_e(X, T) + p(X, T))\right] \tag{22.69}$$

as a density that occurs as an inhomogeneity in a wave or field equation. In order to derive such an equation, we proceed in a conventional manner by making Fourier transformations of the following form

$$\psi_e(x, t) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{ikx - i\omega t}\psi_e(k, \omega)dkd\omega, \tag{22.70}$$

$$\rho(x, t) \leftrightarrow \rho(k, \omega), \tag{22.71}$$

$$G(\xi, t_0) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{ik\xi - i\omega t_0} g(k, \omega)dkd\omega \tag{22.72}$$

with

$$\xi = x - X, \quad t_0 = t - T. \tag{22.73}$$

Then we may write the Fourier transform of ψ_e as

$$\psi_e(k,\omega) = g(k,\omega)\rho(k,\omega). \tag{22.74}$$

The explicit form of $g(k,\omega)$ depends on both the dimension and the specific form of β in (22.68). To elucidate the general procedure, we proceed with a simple example in which case we assume β in the form

$$f_e(x,X) = \beta(x-X) = \frac{1}{2\sigma_e} e^{|x-X_e|}. \tag{22.75}$$

The Fourier transform g can be obtained by means of

$$g(k,\omega) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-ikx+i\omega t} G(x,t)dxdt. \tag{22.76}$$

Using (22.68) and (22.75), we can easily evaluate the integral (22.76), which leads to

$$g(k,\omega) = \left(\omega_0^2 - i\omega_0\omega\right)\left(v_e^2k^2 + (\omega_0 - i\omega)^2\right)^{-1}, \tag{22.77}$$

where we use the abbreviation

$$\omega_0 = \frac{v_e}{\sigma_e}. \tag{22.78}$$

The form of (22.77) suggests to multiply both sides of the Fourier transform of the equation (22.67) by

$$\left(v_e^2k^2 + (\omega_0 - i\omega)^2\right) \tag{22.79}$$

and to back transform. This leads immediately to the equation

$$\ddot{\psi}_e + \left(\omega_0^2 - v_e^2\frac{d^2}{dx^2}\right)\psi_e + 2\omega_0\dot{\psi}_e = \left(\omega_0^2 + \omega_0\frac{\partial}{\partial t}\right)\rho(x,t), \tag{22.80}$$

where we used the abbreviation (22.78). Equation (22.80) is of the form of a damped wave equation with an inhomogeneity that is defined by (22.69) and was previously derived in [23], [24]. It must be stressed that these evaluations were done in one dimension. As a more detailed analysis [17] shows, the general form of (22.80) is rather insensitive to the special form of (22.65) provided we are dealing with slow, long-wavelength solutions. The whole procedure applies also to the two- and three-dimensional case, though in the two-dimensional case the form of the l.h.s. of (22.80) is more complicated.

7 How to make contact with experimental data? Synergetics as a guide

In this section, I wish to briefly elucidate the way how the microscopic field equations, that were derived in the previous sections, can be used to make contact with experimental data, where we take as an example MEG measurements by Kelso and his group [26]. Here a person was subjected to a periodic acoustic signal and had to respond by the tapping of his or her finger in between the acoustic signals. When the frequency of the acoustic signals was increased beyond a critical value, the subject could no more follow the task, but moved his or her finger in phase with the acoustic signal. A nonequilibrium phase transition had occurred. During these tasks, the MEG was measured by a squid array that contained about three dozens of sensors of the magnetic field produced by the brain. From these data the spatio-temporal pattern of the magnetic field across an area that covers the sensory and motor cortex is reconstructed. To bridge the gap between these macroscopic patterns and the microscopic approach, we invoke basic results from synergetics [16], [18]. This interdisciplinary field studies the behavior of complex systems, in particular close to their instability points. According to the general results of that field, close to instability points, the dynamic behavior even of complex systems is – in many cases – governed by few order parameters that occur as the amplitudes of a corresponding number of spatial modes (plus some small additional terms). Since experimentally a nonequilibrium phase transition is observed, according to an earlier suggestion [15], we expected for that MEG experiment only few order parameters. As our data analysis showed [12], the brain area studied indeed behaves as if it is (essentially) governed by two order parameters. Furthermore it is possible to establish order parameter equations that are in good agreement with experimentally observed time series [21], [22].

Now the gap has to be bridged between the microscopic field equations and the phenomenologically derived order parameter equations. In the context of the present article I wish to present only the first few steps, namely how to incorporate the external signals $p(x,t)$ in the field equations [23], [24]. To this end, we decompose the external signals into a sum over individual signals p_j and use

$$\rho(x,t) = a_e S_e \left[\psi_e(x,t) + p_a(x,t) + p_{ms}(x,t)\right]. \tag{22.81}$$

We shall write the input of the acoustic signal in the form

$$p_a(x,t) = \beta_a(x) \sin \Omega t, \tag{22.82}$$

where β_a describes its spatial distribution in the auditory cortex and the sine-function contains the periodic input of frequency Ω. Furthermore, the efferent and afferent nerve fibres generate a feedback loop, which involves

the finger movement. To this end, we assume that the motor cortex generates a motor signal described by

$$\overline{\psi}_m(t) = \int dx \beta_m(x) \psi_e(x,t), \tag{22.83}$$

where $\beta_m(x)$ describes the spatial distribution of the origin of the efferent nerve fibres. It is assumed that the nerve signal (22.83) acts on the finger as if it is a driving force at a delay time t_f, because of the propagation of the signal in the nerve fiber. We model the finger movement as a simple harmonic oscillator and thus write

$$\ddot{z} + \gamma \dot{z} + \omega_f^2 z = c\overline{\psi}_m(t - t_f). \tag{22.84}$$

Anticipating a result that can be derived only later, we write (22.83) in the form

$$\overline{\psi}_m(t) \approx \overline{\psi}_0 \cos \Omega t \tag{22.85}$$

that allows us to solve (22.84) immediately. Under conditions where a phase shift is negligible, we can write the solution $z(t)$ in the form

$$\delta \approx 0 \to z(t) = c_0 \overline{\psi}_m(t), \quad c_0 = \frac{c}{\sqrt{(\omega_f^2 - \Omega^2)^2 + (\gamma\Omega)^2}}. \tag{22.86}$$

This allows us to write the sensory input that stems from the motor activity in the form

$$p_{sm}(x,t) = \beta_{sm}(x) s_{sm}(t), \tag{22.87}$$

where we put

$$s_{sm}(t) = z(t - t_f). \tag{22.88}$$

The basic field equation now reads

$$\ddot{\psi}_e + \left(\omega_0^2 - v^2\Delta\right)\psi_e + 2\omega_0\dot{\psi}_e = a_e\left(\omega_0^2 + \omega_0\frac{\partial}{\partial t}\right) \cdot S_e\left[\psi_e(x,t) + p_a(x,t) + p_{sm}(x,t)\right]. \tag{22.89}$$

When we insert (22.87) with (22.88) and (22.86) into (22.89), we obtain a field equation in which all variables are expressed by ψ_e except for p_a, which is an external input signal. For the further analysis, the sigmoid function S is approximated by a linear term in p_{sm} and a linear and cubic term in $\psi_e + p_a$, i.e.

$$S_e(Q) \approx aQ - bQ^3. \tag{22.90}$$

Taking all these steps together, we, eventually, arrive at the following equation

$$\ddot{\psi}_e + \left(\omega_0^2 - v^2\Delta\right)\psi_e + \gamma_0\dot{\psi}_e + A_e^3 + B\psi_e^2\dot{\psi}_e + \sum_{j=1}^{4} K_j = 0. \tag{22.91}$$

It is a nonlinear damped wave equation, in which the terms K_j are defined as follows: The most important term is

$$K_1 = \epsilon\left(2\Omega\sin 2\Omega t\psi_e - \cos 2\Omega t\left(\dot{\psi}_e + \omega_0\psi_e\right)\right), \tag{22.92}$$

which represents a parametric driving at double the frequency of the incoming signal. The term

$$K_2 \propto \psi_e^2, \psi_e\dot{\psi}_e \tag{22.93}$$

contains quadratic expressions in ψ_e and their corresponding coefficients contain the linear signal. K_3 contains terms linear and cubic in the signal, but no functions ψ_e or $\dot{\psi}_e$, i.e.

$$K_3 \propto \cos(\Omega t + \phi), \quad \cos(3\Omega t + \phi'). \tag{22.94}$$

Finally, K_4 contains the feedback loop of the motor signal as it was explained above

$$K_4 = \gamma_1\dot{\overline{\psi}}_m. \tag{22.95}$$

The solutions of (22.91) are very sensitive to the choice of parameters and, therefore, appropriate solutions of this equation can be found only in connection with a detailed analysis of experimental results. As had been shown elsewhere, an analysis [21], [22] of the Kelso experiments requires that the field equations are parametrically driven by the signal, otherwise the transition from a syncopating to a synchronized motion would not be possible. This implies that the term (22.92) is the most important one. The terms (22.93) may still be there, but play a minor role, while the term (22.94) must be very small or must vanish altogether. One may speculate that this is brought about when an average over an ensemble of neurons is taken. The term (22.95) must, of course, be included. In order to treat (22.91) further [23], [24], a mode expansion

$$\psi_e(x,t) = \sum_{n=-1}^{1} \xi_n(t)\exp(inkx) \tag{22.96}$$

is performed, where

$$\xi_n(t) = \xi_{-n}^*(t). \tag{22.97}$$

We assume periodic boundary conditions of the neural sheet and take only standing waves into account. The hypothesis (22.90) gives rise to equations for the mode amplitudes ξ_n. For these equations contact can be made with previously phenomenologically derived equations [21], [22] so that the gap between the microscopic and macroscopic treatments could be closed. It should be stressed that the theoretical reconstruction of the experimental data depends very sensitively on the proper choice of the constants in the equations. This means, on the one hand, that the brain has learned to manipulate such operations in a highly delicate fashion. It further means that it will be hardly possible to treat brain functioning in all details from first principles, because many of the required constants are not known precisely enough at the microscopic level. On the other hand, it is satisfying to see that by proper choices of the constants the experimental data can, indeed, be recovered. Quite evidently, there is still a lot of work to be done, because experimentally also higher order modes, though of small amplitude, have been observed and it might be interesting to penetrate more deeply into the fine structure of the MEG patterns. At any rate we may state that at least globally the dynamics of the brain states is governed by few order parameters only. Of course, these results do not imply that the whole brain activity is governed by few order parameters only. Rather when specific tasks are to be fulfilled or during episodes, the brain is governed by few order parameters, but these order parameters, in general, change all the time so that in reality there is a huge set of order parameters available for the brain. The above mentioned experiments show clearly that there are critical fluctuations in the transition region. Such fluctuations are well-known from transitions in other synergetic systems [16], [18].

8 Concluding remarks and outlook

In my article I described the "noisy lighthouse model", showing that it can represent phase locking effects between neurons and leads, after the performance of time-averages, to equations of a type that were formerly introduced to model neural nets (see e.g. [2], [14]). Because of the inherent nonlinearities, the solutions can show a rich repertoire of spatio-temporal patterns. I took the analysis of MEG-patterns according to basic concepts of synergetics as an example. Clearly, because of lack of space in my article, only a small cross section of the highly modern and exciting endeavor of modelling brain functions could be presented. Such research is going on at all levels. For instance, at the level of neurons, the integrate- and fire model deserves attention (for recent developments cf. [5], [6]) and its relationship to other models, including the lighthouse model, must be studied. Axonal pulse propagation can be studied in terms of solitary waves and a wide field is opening here [31], [32]. At the macro-level refined methods of EEG and

MEG analysis are being developed [34]. All in all it can be said that brain research is an exciting field in which still many surprises are waiting for us.

Acknowledgments: I thank Peter Christiansen and Alwyn Scott for stimulating discussion.

9 References

[1] Abeles, M.: Corticonics, Cambridge University Press (1991).

[2] Beurle, R.L.: Properties of a mass of cells capable of regenerating pulses, Philos. Trans. Soc. London, Ser. A 240, 55-94 (1956).

[3] Braitenberg, V., Schüz, A.: Anatomy of the cortex. Statistics and geometry, Springer, Berlin (1991).

[4] Braun, H.: Stochastic encoding with intrinsically oscillating neurons, talk given at Sigtuna workshop (1998).

[5] Bressloff, P.C., Coombes, S.: Phys. Rev. Lett. 81, 2168 (1998).

[6] Bressloff, P.C., Coombes, S.: Phys. Rev. Lett. 81, 2384 (1998).

[7] Cowan, J.D.: Brain mechanisms underlying visual hallucinations, in: Paines, D. (ed.), Emerging Syntheses in Science, Addison Wesley, 123-131 (1987).

[8] Eckhorn, R., Reitböck, H.J.: In Synergetics of Cognition (eds. Haken, H., Stadler, M.), Springer, Berlin (1986).

[9] Ermentrout, G.B., Cowan, J.D.: A mathematical theory of visual hallucination patterns, Biol. Cybern. 34, 137-150 (1979).

[10] Freeman, W.J.: Biol. Cybern. 56, 139 (1987).

[11] Freeman, W.J.: Tutorial on neurobiology: From single neurons to brain chaos, International Journal of Bifurcation and Chaos, Vol. 2, No. 3, 451-482 (1992).

[12] Fuchs, A., Kelso, J.A.S., Haken, H.: Phase transitions in the human brain: Spatial mode dynamics, International Journal of Bifurcation and Chaos, 2, 917-939 (1992).

[13] Gray, C.M., König, P., Engel, A.K., Singer, W.: In Synergetics of Cognition (eds. Haken, H., Stadler, M.), Springer, Berlin (1986).

[14] Griffith, J.S.: A field theory of neural nets: I: Derivation of field equations, Bull. Math. Biophys., Vol. 25, 111-120 (1963), and II: Properties of field equations, Bull. Math. Biophys., Vol. 27, 187-195 (1965).

[15] Haken, H.: Synopsis and introduction, in: Synergetics of the brain, (eds.) Başar, E., Flohr, H., Haken, H., Mandell, A.J., Springer, Berlin (1983).

[16] Haken, H.: Synergetics. An Introduction, 3rd, ed., Springer, Berlin (1983).

[17] Haken, H.: A physicist's view on brain functioning: Noise, coherence, chaos, to be published in special issue by World Scientific.

[18] Haken, H.: Advanced Synergetics, 2nd ed., Springer, Berlin (1987).

[19] Haken, H.: Principles of brain functioning: A synergetic approach to brain activity, behavior and cognition, Springer, Berlin (1996).

[20] Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. 79, 2554 (1982).

[21] Jirsa, V.K., Friedrich, R., Haken, H., Kelso, J.A.S.: A theoretical model of phase transitions in the human brain, Biol. Cybern. 71, 27-35 (1994).

[22] Jirsa, V.K., Friedrich, R., Haken, H.: Reconstruction of the spatio-temporal dynamics of a human magnetoencephalogram, Physica D 89, 100-122 (1995).

[23] Jirsa, V.K., Haken, H.: A field theory of electromagnetic brain activity, Phys. Rev. Lett., 22 July (1996).

[24] Jirsa, V.K., Haken, H.: A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics, Physica D, in press.

[25] Kohonen, T.: Associative memory and self-organization, 2nd ed., Springer, Berlin (1987).

[26] Kelso, J.A.S., Bressler, S.L., Buchanan, S., DeGuzman, G.C., Ding, M., Fuchs, A., Holroyd, T.: A phase transition in human brain and behavior, Physics Letters A 169, 134-144 (1992).

[27] McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity, Bulletin of Math. Biophysics 5, 115-133 (1943).

[28] Nunez, P.L.: The brain wave equation: A model for the EEG, Mathematical Biosciences 21, 279-297 (1974).

[29] Nunez, P.L.: Electric fields of the brain, Oxford University Press (1981).

[30] Nunez, P.L.: Neocortical dynamics and human EEG rhythms, Oxford University Press (1995).

[31] Scott, A: Nonlinear Science: Emergence and Dynamics of Coherent Structures, Oxford University Press (1999).

[32] Scott, A.: private communication (1998).

[33] Tass, P.: Cortical pattern formation during visual hallucinations, J. Biol. Phys. 21, 177-210 (1995).

[34] Tass, P.: Phys. Rev. Lett., in press (Oct. 12, 1998).

[35] Tass, P.: Oscillatory cortical activity during visual hallucination, J. Biol. Phys., in press.

[36] Wilson, H.R., Cowan, J.D.: Excitatory and inhibitory interactions in localized populations of model neurons, Biophysical Journal, Vol. 12, 1-24 (1972).

[37] Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, Kybernetik 13, 55-80 (1973).

Index

Ablowitz, M.J. 57
Abraham, F.F. 276
Acetanilide, crystalline, see ACN
ACN soliton
  absorption intensity 353–354
  binding energy 350, 352
  effective mass 352
  local modes in 348–354
  overtone series 352–353
  peak in 349, 350
Adenosine triphosphate, see ATP
Algal cells, motile 414
Amide-I (peptide) modes 348, 349, 406
Amplitude, complex 26, 340
Anharmonicity, geometric 403
Anisotropy, spatio-temporal 252
Anticompacton 330
Anti-ferromagnetism 138
Arroyo, C.M. 383
ATP (adenosine triphosphate) 358, 366, 368, 369
Aubry, S. 369
Averaged variational principle 49–50, 52
Axon 427

Bäcklund transform, for SG equation 267
Bacteria
  aerobic 409
  Bacillus subtilis 409, 411, 415, 416, 424
  Escherichia coli 424
  free will of? 410, 419, 420
  locomotion 415
  swimming 409
    speeds 416–417, 418, 419, 420, 421
Barashenkov, I.V. 180
Barthes, M. 354
Baverstock, K.F. 374, 383, 384
Beam, optical 173
  blow-up in 225
  multi-splitting of 223, 224
  trapping 182
Belova, T.I. 281
Benard convection 4
Benjamin, T.B. 50
Bernstein, L. 343, 354
Bigio, I. 354
Bioconvenction 411–412, 413, 418
Biofilms 410
Biomagnetism, 74
Bishop, A.R. 307, 308, 379
Bjerrum defects 274
Black hole(s) 4, 20, 21–26, 30, 38
Bloembergen, N. 50
Blow-up in finite time 231
Boesch, R. 155
Bogdan, M.M. 180
Bogolubov, N.N. 48
Bogolubskaya, A.A. 374
Bogolubsky, I.L. 374
Booth, A.D. 374, 382
Born–Oppenheimer approximation 349
Boson
  gas condensate 180
  lowering and raising operators 341
  number levels 341
Boudon, V. 335
Bound states, fluxon-fluxon and fluxon-antifluxon 104, 127–128
Boundary layers, hydrodynamic 53
Bragg scattering 251
Brain, human 427
  MEG and EEG measurements 427, 437, 440, 443–444
Braun, O.M. 276
Breakdown, optical 235
  avalanche 230

facet damage 230
Breathers
biomolecular 369
collisions of 302–303
discrete 293
moving, longitudinal 298, 299, 300, 301
stationary 299
Brillouin scattering 195
Brownian motion 358, 359, 362, 363, 410

Campillo, A.J. 182
Caputo, P. 283
Carbon-hydrogen, see CH
Careri, G. 348, 349, 355
Cascading 247
Cauchy-Schwarz inequality 215
CH (carbon-hydrogen) stretching oscillations 339, 342
Chains, hydrogen bonded 273–275
Channeling, optical, in air 233–239
Chaos, three wave 8
Characteristics, method of 33
Charge, topological 140
Chaux, R. 335
Christiansen, P.L. 191, 444
Christodoulides, D.N. 187, 189
Coherence, induced 185
Coherent phenomena 3–41
Collapse (in) 4, 214
arrest of 29, 235
contraction law 220
critical 220, 230, 231
dimension for 219
in 2D 223
power for 235
dimension, supercritical 220
domain, localizing vs. nonlocalizing 223
generalized KP 38
point of 19, 20
recurring 237, 238
scenario 17
self-similar 28, 41, 216–220, 237
singularity 230, 231, 232
strong 19, 24, 32
superstrong 19–20, 23
threshold for 17
time 34, 218
transverse 221, 222
viscous Benjamin–Ono equation 39
weak 18–19, 24, 25, 29–30, 32, 38
Collective variable 142, 146
Compacton(s) 281, 324–325, 329, 330, 331, 332
Compatibility conditions 48
Complex Swift–Hohenberg, see CSH
Conductors, superionic 274
Connections, neuronal 428
Conservation of organisms 414
Cooper pairs 77, 128
Corkum, P.B. 183
Correlation function 7
Cosmic rays 293
Cowan, J.D. 436
Critical angle of gravity waves 32
Crowdions 270, 274
Cryotron 75
Crystal growth 275
CSH (complex Swift-Hohenberg) model
extended 233
for semiconductor laser 241
Cundal, R.D. 374, 383, 384
Current amplitudes, dendritic 435
Currents, dendritic 429, 435, 437–438
Cytoskeleton 357

Damping
Gilbert 157, 158
Landau-Lifshitz 157
Dauxois, T. 374
Davydov, A.S. 348, 355, 373, 405, 406
Davydov's product approximation 349, 350, 352
Debye radius 21
Defects in
crystals 265, 266, 269–270
decorated mica 295
Dehlinger, U. 266
Dentrite(s) 427
Deoxyribonucleic acid, see DNA
Diddams, S.A. 183, 224

Diffusion constant, effective, for vortices 155–156, 161
Dihalomethanes 339, 344
  classical model 340
  experimental data 344
  local mode 340–341
  symmetric vs. anti symmetric mode 340
Dimitrov, D. 167
Directional couplers, nonlinear 208
Dispersion
  anomalous 249
  anomalous vs. normal 213
  weak 26–32
DNA (Deoxyribonucleic acid)
  anharmonicity in 384
  atomic model 377, 378
  conformational transitions 373, 385
  coupled discs model 377
  denaturation 373
  dynamics
    experimental observations of 381–384
    FK models for 276–277
    modeling principles 376–378
    proton exchange 381–381
  elastic rod model 377, 378
  fast neutron scattering 383
  fluorescence depolarization 382
  helical structure of 277
  internal dynamics of 375–376
  lattice models 377, 378
  long range effects 385–386
  mathematical models 379–381
  nonideal models 380
  nonlinear dynamics of 373–387
    mechanisms 385–387
  open state dynamics 379
  relation to FK 266
  resonant microwave absorption 382–383
  SG model for 383
  structural model 377, 378
  transcription 386–387
  two rod model 377, 378
Domínique-Adame, F. 167
Domain walls, ferromagnetic and ferro electric 270–271, 324, 325–330
Drummond, P. 261
Duncan, D.B. 400
Dusuel, S. 281, 335
Duwel, A.E. 283
Dynamics, of magnetic vortices
  above KT transition 163–166
  Langevin 141, 162, 155–166
  Newtonian 140, 146, 147, 155
  non-Newtonian 141
Dynes, R.C. 281

EEG (Electroencephalogram)
Eilbeck, J.C. 333, 346, 355, 400
Einstein oscillator 348, 349, 350
Eisenberg, H.S. 189, 190
Electron-positron showers 295
Energy eigenvalues, for dihalomethanes 342–343
Energy transfer, targeted 369
Englander, S.W. 374, 381, 382
Enneper, A. 267
Enol'ski, V. 355
Ensembles of neurons 428, 442
Equation(s)
  Benjamin–Ono 28, 30, 39, 309
  Bloch 239
  Burgers 29, 30
  CSH 232, 239, 240
  double SG 278
  Euler–Lagrange 13
  Fokker–Planck 162
  four-wave 252
  generalized Klein–Gordon 324, 325
  generalized KP 37
  Ginzburg–Landau 4
  Hamilton–Jacobi 24
  Hilbert–NLS 309
  Hill's 433
  Hopf 28, 33
  KdV 50, 52, 309
    generalized 30, 32
  Kadomtsev–Petviashvilli (KP) 27–28, 29, 35, 36–37, 61
  Khokhlov–Zabolotskaya–Kuzhetsov 61
  Landau–Lifshitz 140, 156–157
    stochastic 157, 160–161, 162

Langevin 365
Laplace 33
Laplacian growth 34
Manakov 187
Maxwell's 49, 176, 224, 231, 249
Navier–Stokes 412
nonlinear Klein–Gordon, double well potential 327–328
SG (sine–Gordon) 70
elastically coupled 110
perturbed 82, 88
Thiele 140, 141–143, 144, 145, 151, 162, 163
three-wave 250, 257, 258
Zakharov 22
Equivalence principle, in perturbation theory 57
Ermentrout, G.B. 436
Escherichia coli 424
Euler strut 329, 332
Excitations, gyrotropic 146, 151
Experiments, neural 439–443

Fabry–Perot etalon 240, 241
Fedyanin, V.K. 374, 381, 383
Feir, J.F. 50
Femptosecond pulse 183, 235
Feng, Q.Y. 244
Fermi–Pasta–Ulam, see FPU
Ferro- and antiferromagnetism, relation to FK 266
Ferromagnetism 37, 138, 266
Fiber optics, nonlinear 195–209
Fibers, optical, dispersion flattened 198
Fibers, optical, single mode 196
Field equations, neural 437, 440
Filaments, collapsing 243
Fiske steps 94, 108, 112, 113
satellites 87, 94–98
FK (Frenkel–Kontorova) model (for) 266–269, 324
discreteness effects in 279
generalized 277
magnetic systems 283–285
properties of kinks 277–281
Flach, S. 286
Flagella 415
Flesch, R. 400
Fluctuation-dissipation theorem 155
Fluctuations
neural 429
velocity 39
Fluids
classical 139
ideal 9, 33
Flux quantization 69
Fluxons 272
bound state energy of 104
fractional 115
oscillating 282
Foams, in 2D 419 -420
Forces, stochastic 1158–159
FPU chain 393, 406
Fraiman, G.M. 220
Frank–Kamenetskii, M.D. 382
Frauenfelder, H. 369
Fraunhofer pattern, in JJs 72
Freeman, W.J. 428
Frenkel, Ya.I. 265, 266
Frenkel–Kontorova, see FK
Fritsche, H. 375
Fulton, T.A. 281
Functionals, positive definite 31
Funnel, anisotropic 24
Fusion plasma heating 4

Göhmann, F. 167
Gaeta, G. 373, 374
Gaididei, Y. 167
Galin, L.A. 34
Garmire, E. 182
Gaussian noise 157
Gaussons 178
Gilhøj, H. 355
Ginibre, J. 226
Ginzburg–Landau theory 151
Goldobin, E. 282
Goldstone mode 278
Gonzales, J.A. 374
Gouvêa, E. 166
Grønbech-Jensen, N. 167, 273
Grantham, J.W. 183
Gratton, E. 355
Gravity wave detector 75
Green function for
neurons 438
vortex trajectories 159

Grimshaw, R. 52
Group velocity dispersion (GVD) 197–198, 213
  anomalous 214, 220
  normal 232
  weak 231
Gyrovector, density of 145

Hai, W. 374
Hall effect 145
Hamiltonian (for)
  classical for dihalomethanes 340
  coupled JJ system 104
  DNA 374
  focusing NLS 11, 14, 15
  generalized KP 37
  Heisenberg spin lattice 138
  helix backbone 397
  lower-hybrid waves 25
  magnetic chain 271
  negative 16, 31, 40
  NLS 214, 217
  nonlinear Klein–Gordon equation 324
  nonlocal NLS 311
  plasma 22
  QDNLS 345
  quantum for dihalomethanes 341
  self-focusing 39
  2D lattice NLS 316
  unbounded 16
  viscous Benjamin–Ono 40
    plasma 35
Harmonic conversion 249
  oscillator, displaced 351
Hartree approximation 347, 349
Harvey, S.C. 375
Hechtfischer, G. 282
Helbing, D. 416
Helices in 2D and 3D 394–397
  dynamics of 397–398
  small amplitude limit 398–400
Hexagons 4
Hierarchy of DNA models 376–378
Hilbert transform 309
Hochheiser, D. 245
Hodgkin Huxley nerve model 428
Holyst, J.A. 284
Homma, S. 374

Hopf bifurcation 233
Hopfield, J.J. 436
Hydrodynamics, free surface 33, 34
  ideal 26–27
Hydrolysis of ATP 358, 366, 368, 369
Hyman, J.M. 324

Index of refraction, intensity dependent 175
Instability
  decay 6
  drift 180
  Kelvin–Helmholtz 32
  KP 39
  modulational 198–199, 253
  monochromatic wave 8
  oscillatory 181
  short wave 226
  transverse modulational 243
  wave pairs 8
Instantons 162
Interactions, long range 401
Interferometer, plasma wave 25–26
Inverse scattering transform method, see ISTM
Ionization, multiphoton 230
Ionosphere, auroral 25
ISTM 47, 248
  for optical solitons 200
  perturbation theory for 55–57
Ivanov, B. 167

JJ (Josephson junction) devices
  ac voltage standards 80
  analog vs. digital 69
  arrays 81
    phase locking of 81–82
  cavity modes 108
  cooling problems 83
  detectors 77–78
  digital 75–77
  flux flow (in)
    microwave fields 91–94
    oscillators 82, 87–100
    resonances 96
    uniform magnetic field 88–91
  general 69–83
  I–V characteristic 71, 88, 93, 94, 111, 113, 115

long 81, 82, 94, 95, 96, 281–283
coupled 103–118
fluxon 122
overlap vs. annular geometry 122
power balance 10, 108
radiation from 107
triangular cells 113–118
zero field step 105, 106, 107, 116
magnetic field sensors 69
microwave oscillators 80–82
oscillators
phase locking 91
radiation from 112–113
parallel arrays 109–113
receiver, sub-mm wave 82
SQUID systems 70
stacked 123–135
annular 282
plasma oscillations 128
two-fluxon mode 125–126
two resonance modes 134
voltage standards 78–80
JJ (Josephson junction) phenomena
annular geometry 282
basic theory 70–72
chaos in 79
Cherenkov rediation from 282, 283
coupled 273
experimental studies of 281–283
large, two-dimensional 109
penetration length 88, 104
power balance in 272
relation to FK 266
resonant flux-flow steps 117–118
SG model for 272–273
Shapiro steps 79
small vs. large 71–72
zero field steps 111, 112
Joseph, R.I. 189
Josephson junction, see JJ
Jutzi, W. 75

Kallenbach, N.R. 374
Kamppeter, T. 167
Kaplan, A.E. 178
Karpman, V.I. 55
KdV (Korteweg de Vries) equation 50, 52, 309
generalized 281
solitonic properties of 268
with nonlinear dispersion 325
Keener, J.P. 54
Keller–Segal model 414
Kelso, J.A.S. 440, 442
Kerr effect 17, 49, 174, 175, 182, 189, 231, 247, 248, 307
Ket products 341, 342
Khokhlov, R.V. 50, 60
Kholmurodov, Kh.T. 180
Kinesin 361
Kinks and breathers, in FK model 267, 268
Kislukha, N.I. 405
Kivshar, Yu.S. 180, 184, 261, 273, 276, 280, 282
Koch, S. 230
Kodama, Y. 57
Kohonen, T. 437
Kontorova, T. 265, 266
Korteweg-de Vries, see KdV
Kosterlitz, J.M. 163
Kosterlitz-Thouless (KT)
transition 166
temperature 138
Kovalev, A.S. 167, 270
KP (Kadomtsev–Petviashvilli)
Krolikowski, W. 180
Kronig–Penny model 308
Krumhansl, J.A. 355, 374, 381, 385
Kruskal, M.D. 393
KT (Kosterlitz–Thouless)
Kudryavtsev, A.E. 281
Kuznetsov, E.A. 214, 226

Landau, A.I. 270
Langmuir-Blodgett films 138
Laser, Raman 232
Laser, semiconductor 232, 233, 244
Lattice(s)
discrete modes in 286
NLS equation 344–347, 350
solitons (in)
ACN 348–354
dihalomethane 341
mica 293–299

quantum 339–355
Laval, G. 226
Lederer, F. 261
Lega, J. 245
Lennard–Jones potential 299, 301
Lenton, T.M. 409
LIDAR, remote sensing by 239
Light bullets 260
Lighthill, M.J. 50
Lighthill's criterion 50
Likharev, K.K. 75, 76
Lindenberg, K. 369
Linewidth enhancement 239
Litwin, A. 374
Local modes in dihalomethanes 340–341
Local modes, in ACN 348–354
Lomdahl, P.S. 276
London penetration depth 70, 103
Lorentz
contraction, in SG 268, 269
force 140
invariance, in FK model 278
Lukens, J.E. 81
Luther, G. 244
Luther-Davies, B. 184
Lyapunov function 180
Lythe, G. 167

Mach–Zehnder interferometer 202
Magnetic chains 270, 271
Magnetite 295
Magnetocardiography 74
Magnetoencephalogram, see MEG
Magnetometer, see SQUID
Magnets, two-dimensional 137–166
Magnon modes 152
Magnus force 140
Mak, W. 261
Malkin, V.N. 220
Malomed, B.A. 273, 280, 282
Mamaev, A.V. 184
Manley–Rowe invariant 253
Marburger, J.H. 229
Marin, J.L. 333
Martin-Landrove, M. 374
Maslov, E.M. 55
Masses, electron and ion 21
Matisoo, J. 75
Matthieussent, G. 226
Maxwell–Boltzmann distribution 417
McCammon, J.A. 375
McClure, W.R. 375
McCulloch, W. 428, 436
McLaughlin, D.W. 54, 272
Mechanical models
experimental 331–333
lattices 333–335
MEG (Magnetoencephalogram) measurements 440, 443
Mendelson, N.H. 423
Mermin–Wagner theorem 138
Mica 293, 294, 296, 301
structure of 297
Michaux, P. 335
Microorganisms 409–425
Microscope, scanning tunnel 275
Microscopy, magnetic 75
Miller, J.H. 384
Mineev, M.B. 273
Mitchell, M. 188
Mitropolsky, Y.A. 48
Mixing
four-wave 181
resonant 6, 8
Mlejnek, M. 245
Model, Heisenberg 139, 141, 146
Modes, guided 173
Modulation, optical, cross-phase vs. self-phase 197
Molecular nonlinearity, intrinsic vs. extrinsic 339
Momentum, three wave 7
Monte Carlo simulations 138, 141, 367
Moro, E. 167
Morse potential 301, 402
Motion, synchronized 442
Multi-photon absorption 235
Muons 295
Muscovite mica, see Mica
Muto, V. 374, 379, 382, 383

Neural equations, averaged 434–439
Neuron(s) 427
excitatory 436
interacting 430–433
Neurotransmitters 429

Newell, A.C. 244, 245
Newtonian dynamics 140
of SG kink 280
Niemeyer, J. 83
NLS (nonlinear Schrödinger) equation 4, 9–11, 50, 213, 253, 307, 309
collapse 15–18
compact focusing 10, 11–15
coupled 187, 209
cubic 177
discrete 189–190, 300
generalized 177, 179
ground state of 214
lattice 350
modified 18–19
nonlocal 311–316
breathing mode 314–315
competetion in 313
discrete 311
mode switching in 315
optical fibers 198
perturbed 204
quantum analogy 11
quantum lattice 344–347
superlattice system 310
three-dimensional 183
two-dimensional 231
NMR imaging, magnetic 75
Noise, Gaussian and delta correlated 358, 365
Noisy light tower model 429–433, 443
Nonlinear Schrödinger, see NLS
Nonlinear vacuum 3
Nonlinearities
competing 177
logarithmic 178
nonresonant 176
quadratic 5
saturable 177, 178
transiting 177
Norm (number)
classical for dihalomethanes 340
focusing NLS 11, 14
NLS 214, 217, 308
nonlocal NLS 311
plasma 22
QDNLS 345
quantum for dihalomethanes 341
three wave 7
2D lattice NLS 317
Number of particles (see Norm)
Number state method 342–343
Nunez, P.L. 438

Olfactory bulb 428
Optical switching, nonlinear 205–209
Orientation, gyrotactic 415
Ornstein–Uhlenbeck process 358, 359
Orthogonality conditions 54
Oscillations, high frequency plasma 21
Oscillator, van der Pol 48
Ostrovskaya, E.A. 187
Ostrovskii–Sutin equation 382
Ostrovsky, L.A. 49

Parmentier, R.D. 62, 98
Pegmatite 295
Peierls–Nabarro barrier 161, 190, 270, 276, 279, 281
Pelinovsky, D.E. 180, 191
Pelinovsky, E.N. 52
Penetration depth
Josephson 104, 127
London 103, 124, 135
Peptide chains 348
Perring, J.K. 268
Perturbation analysis 47–62
direct 51–52
ISTM based 55–57
non-stationary 54–55
quantum 343
Peschel, U. 261
Pesme, D. 226
Peters, R.D. 325
Petrishchev, V.A. 16
Peyrard, M. 278, 374, 379, 380
Phase locking in JJs 106–108, 125–128
neuronal 432
Phase matching 248, 251
randomization 4–9
Pierrre, G. 335
Pilipetskii, N.F. 182
Pitts, W. 428, 436
Plasma
frequency 21

in JJs 128, 131, 133, 134
generation 234
isotropic 26
Polubarinova-Kochina, P.Ya. 34
Ponderomotive force 25
Prohovsky, E.W. 374
Protein
α-helix in 395, 402, 403, 406
amide-I oscillation 348, 406
helix backbone 395, 396
microtubular assocciated 357
model vs. natural 348
motors 364–368
Pulse
axonal 429
femptosecond 183, 235
rates, axonal 435
splitting, with normal GVD 233

QDNLS (quantum discrete nonlinear Schrödinger)
effective mass 345, 346, 347
energy eigenvalues 346, 347
equation 344–347
Hamiltonian for 341
Hartree approximation for 347
number operator for 341
quasicontinuum band 346
soliton
bands 346
binding energy 347
wave packet 347
translation operator of 345
Quantization of magnetic flux 69, 81, 272
Quantum discrete NLS equation, see QDNLS
Quantum
Dirac's theory of 154, 341
lattice solitons 339–355
number state method 342–343
solitons 347
wave packets 343
Quodon 293, 296, 298
mechanical model of 296, 297
model for high-T_c superconductors 303–304
molecular dynamics models of 298–299
sputtering applications of 303

Raman scattering 183, 195, 209, 284
Ramia, M. 416
Random walk 160
Ranka, J.K. 183, 224
Ratchet(s)
correlated 358–363
Brownian 362
effect in biomolecules 357
Remoissenet, M. 278
Resonance
causing secular growth 48
conditions for 8, 26
lower hybrid 25
Reynolds number 39, 416
Rolls 4
Rosenau, P. 324
Russell, F.M. 293, 333
Russo, M. 83
Rustamov, A.R. 182
Rutherford scattering 295
Rytov, S.M. 49

Sánchez, A. 167
Sørensen, M.P. 191
Sagdeev, R.Z. 32
Sakai, S. 273
Salerno, M. 355, 374, 386, 387
Sasik, R. 167
Saut, J.C. 226
Scaling for focusing NLS 12
Scharf, R. 308
Schnitzer, H.J. 167
Schrieffer, J.R. 381
Scott, A.C. 272, 324, 373, 406, 444
Second harmonic generation 247
Secularity 48, 53, 54
Seeger, A. 268
Segev, M. 191
Self-focusing 182–184
boundary layer 38–41
critical 229
NLS, with anomalous dispersion 214–220
normal dispersion 220–224
nonstationary 20
optical 174
semiconductor lasers 230

two-dimensional 225
Self-similarity 17, 18, 23, 28, 31, 34, 41
strong vs. weak 217, 220
Self-trapping, optical 174
Selvin, P.R. 374, 384
Semiconductor laser, control of 239–243
Semiconductor, optical 178
SG (sine–Gordon) equation
discrete 273
domain walls 324
double 284
magnetic systems 281
mechanical models for 324
model for DNA 383
perturbed 121, 125, 280
relation to FK model 266
solitons, homopolar vs. anti-polar 111
Shapiro, S. 78
Shen, Y.R. 229
Shih, M.F. 183
Shock waves 53, 183
Shrira, V.I. 39
Simulations
Langevin dynamics 166
Monte Carlo 164, 165–166
Spin dynamics 164, 166
Sine–Gordon, see SG
Singularities, weak 34
Skryabin, D. 191
Skyrme, T.H.R. 268
Snyman, J.A. 276
Snyman, J.H. 276
Sobolev inequality 13
Sobolev space 215
Soliton(s) (of) 4
asymmetric 260
binding energy, in ACN 350, 352
birth and death of 62
bistable 178
Davydov, acoustic vs. optical mode interactions 348
dihalomethane 341
first and second harmonic 255
gap 251, 259, 260
generalized KP 34–36
guiding center 203
hydrodynamic 39
interactions 58–59
lattice 59, 293
longitudinal compression 402–404
magnetic 283–284
multidimensional 36, 248
multihump 186–189, 256
one parameter 179–180
optical 199–205, 247–261
black 202
bright 179, 200–201
coupled bright-dark 181
dark 50, 180, 201–202, 232, 256
discrete spatial 189–191
dispersion managed 203–205
loss managed 202–203, 204
photorefractive 186
spatial 173–191, 248
temporal 174
vortex 184
two parameter 180–181
quantum 347
radiation from 59–60
soft 14
spiralling 184, 185
statistics in DNA 380–381
three component 400–405
thorns 41
Toda-like 298
two-wave 180
walking 181, 257
Solitonic gluons 188, 189
Soustova, I.A. 62
Spin waves 138
Spin waves, quasilocal 148, 154
Sputtering, quodon application to 303
Superconducting quantum interference device (SQUID) 72–75
Srolovitz, D.J. 276
Stability, for dark solitons 180
Stability, solitary wave 178–181
Stationary solution(s)
dihalomethane 339, 340
focusing NLS 11
Stokes, G.G. 32
Strickland, D. 183

Sublattices, interacting 274
Super-breather 303
Superconductivity, low- vs. high-T_c 121, 135
Supercritical case 17–18
Superfluid(s) 139
  charged 152
  films of 151
Surface physics
  FK model for 275–276
  singularities 32–34
  sputtering 303
Swarms, cytobacterial 423
Swicord, M.L. 374, 382
Swihart velocity 104, 105, 108, 122, 132, 272
Swimming velocity probability density function (SWPDF)
  definition of 413–414
  measurement of 414
Synapse(s) 427, 429, 437
Synergetics 439, 440
System, three wave(s) 6

Takeno, S. 374
Talanov, V.I. 17
Tanuiti, T. 60
Tass, P. 436
Taylor-Chirikov map 279
Tensor, antisymmetric 143
  gyrocoupling 142
Testing, nondestructive 74
Thermal noise
  additive vs. multiplicative 156
  introduction of 156–157
Thirring model 181, 258
Thorns, formation of 41
Thouless, D.J. 163
Three-wave mixing 181
Tikhonenko, V. 184
Time averages, of neural activity 433–434
Tomboulis, E. 155
Torner, L. 261
Torus, infinite dimensional 5
Transform Hilbert 33
Transformation, scaling 22–23
Transmissions, nonlinear, electrical 323
Tubulin 360–361, 362, 365
Tunneling time 343–344
Turbulence, weak 7, 8, 243

Ustinov, A.V. 273

Völkel, A. 166
Vakhitov–Kolokolov stability criterion 179–180, 180–181, 255, 257
van der Merwe, J.H. 276
Van der Zandt, H.S.J. 273, 283
Van Zandt, L.L. 374, 382
Vaqueiro, P. 304
Variational method 252, 254–255
Velocities, characteristic 61
Vibrations, molecular 339
Virial
  arguments 222
  equations 221
  integral 215
  theorem 37, 50
Viscosity 27, 29, 39
Vlasov, S.N. 16
Volkov, S.N. 379
Vortex
  core radius 139
  diffusion constant 155
  dynamics, above KT transition 163–166
    effects of thermal noise on 155–162
  mass of 143–146
  motion, cycloidal 153
    effective 165–166
  rings 182
  sheets 32
  static 152
  structure, static 158
  trajectory of 149
Vortex-antivortex pairs 138, 144, 161
Vortices in 2D magnets 137–166
  nonplanar 139, 143

Water tank experiments 324
Wave(s)
  acoustic 26
  breaking 29, 30, 33
  equations

asymptotic reduction 60–61
integrable 47
envelope 49
gravity 3, 11, 28, 32
guides
semiconductor 178
optical 173
slab 177
ion-acoustic 28
Langmuir 21, 22
magnetoacoustic 28
modulated 49-50
monochromatic 6, 9, 10, 26
packet
for ACN 352
localized 5
quasi-harmonic 49
sea surface 32
shallow water 50
shock, envelope 49
solitary
modulated 50
multi-component 174
spin 138
Stokes 50
wind driven 3
Webb, S.J. 374, 382
Weinstein, M. 14
White caps 3
Whitham, G.B. 49, 50
Willis, C.R. 155, 281, 286
Wilson, H.R. 436
Wilson–Cowan model 436
Wiseley, D. 423
Wright, E. 244, 355
Wysin. G. 166

Xiao, J.X. 379

Yakushevich, L.V. 373, 374, 375, 381, 383
Yomosa, S. 374, 406

Zabusky, N.J. 393
Zakharov, V.E. 214
Zener diodes 79
Zhang, Ch.T. 373, 374, 379, 383
Zharova, N.A. 223
Zhou, G.F. 373