

K&E Series on Knots and Everything – Vol. 19

IDEAL KNOTS

Editors

A. Stasiak, V. Katritch and L. H. Kauffman



World Scientific
www.ebook3000.com

IDEAL KNOTS

SERIES ON KNOTS AND EVERYTHING

Editor-in-charge: Louis H. Kauffman

Published:

- Vol. 1: Knots and Physics
L. H. Kauffman
- Vol. 2: How Surfaces Intersect in Space
J. S. Carter
- Vol. 3: Quantum Topology
edited by L. H. Kauffman & R. A. Baadhio
- Vol. 4: Gauge Fields, Knots and Gravity
J. Baez & J. P. Muniain
- Vol. 5: Gems, Computers and Attractors for 3-Manifolds
S. Lins
- Vol. 6: Knots and Applications
edited by L. H. Kauffman
- Vol. 7: Random Knotting and Linking
edited by K. C. Millett & D. W. Sumners
- Vol. 8: Symmetric Bends: How to Join Two Lengths of Cord
R. E. Miles
- Vol. 9: Combinatorial Physics
T. Bastin & C. W. Kilmister
- Vol. 10: Nonstandard Logics and Nonstandard Metrics in Physics
W. M. Honig
- Vol. 11: History and Science of Knots
edited by J. C. Turner & P. van de Griend
- Vol. 12: Relativistic Reality: A Modern View
J. D. Edmonds, Jr.
- Vol. 13: Entropic Spacetime Theory
J. Armel
- Vol. 14: Diamond – A Paradox Logic
N. S. K. Hellerstein
- Vol. 15: Lectures at Knots '96
edited by S. Suzuki
- Vol. 16: Delta – A Paradox Logic
N. S. K. Hellerstein
- Vol. 19: Ideal Knots
A. Stasiak, V. Katritch and L. H. Kauffman

 Series on Knots and Everything — Vol. 19

IDEAL KNOTS

Editors

A. Stasiak

Université de Lausanne

V. Katritch

Rutgers University

L. H. Kauffman

University of Illinois, Chicago

 **World Scientific**
Singapore • New Jersey • London • Hong Kong

www.ebook3000.com

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Title of cover: Plato, Kelvin and ideal trefoil

Cover done by Ben Laurie.

IDEAL KNOTS

Copyright © 1998 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-02-3530-5

Printed in Singapore by Uto-Print

Preface

In a strict topological sense all representations of a given type of a knot are equal as they can be converted into each other by a finite number of moves maintaining the original topology of the knot (moving of segments without passing through each other, translation and rotation, change of the scale). However, we intuitively avoid this extreme polymorphism of knots. Thus when we make a quick drawing of a trefoil we usually trace it with just three crossings and not with many additional unnecessary crossings. When we are asked for a really nice drawing of a trefoil we try to give it a very regular symmetrical shape. This is why the founders of the mathematical classification of knots represented different knots in the tables in such a way that each type of the knot was shown with its minimal crossing number and in addition had a regular shape with elements of symmetry (when this was possible). Furthermore, when several persons are asked independently for an extremely nice and elaborated drawing of a trefoil they will produce different drawings where the presented trefoils will usually maintain three-fold symmetry, but while one person will draw a trefoil with elongated blades another will give it almost circular shape. This is of course different from drawings of a triangle or of a tetragon where almost everyone will draw an equilateral triangle and a square where the only difference between individual drawings is their scale.

Can we thus think about an ideal representation of a knot such that everyone would agree that it is the best representation of a given knot, like we all can agree that the square is the ideal representation of a tetragon? This question was raised at the physical knot theory session organized by Jonathan Simon and Gregory Buck at 1996 ASM meeting in Iowa and it turned out that there was no consensus. Background and the particular scientific interest of concerned scientists showed that they had in mind different ideal representations of knots. Some were applying simple geometrical criteria, some were minimizing certain physical or abstract energies, some looked for the simplest parametric descriptions of given knots, some wanted to have high symmetry, some hoped that the ideal shape would correspond to a "mean" trajectory of all possible representations of a given knot, some demanded that ideal knots should be built of a minimal number of straight segments on a particular lattice etc.etc. Despite this lack of consensus one thing become apparent: among an infinite sea of shapes of every type of knot only a small subset or even just one unique representation can satisfy certain criteria and is thus ideal according to these criteria. In this volume different authors characterize their approach to their favourite ideal knots and links.

The following list of chapters with short descriptions of their contents should help to navigate between the chapters and to encourage readers to visit each domain of the ideality.

Ideal knots and their relation to the physics of real knots by A. Stasiak, J. Dubochet, V. Katritch and P. Pieranski. This chapter presents the concept of Platonic, ideal geometric representations of knots and demonstrates their ability to govern the time averaged behaviour of randomly distorted knotted polymers forming corresponding types of knots.

In search of ideal knots by P. Pieranski. This chapter describes a numerical algorithm used to approach ideal geometric representations of knots and shows fascinating transitions of knots on their way toward ideal configurations.

Annealing ideal knots and links: methods and pitfalls by B. Laurie. A simulated annealing algorithm adapted to perform a search for ideal representations of knots and links is presented together with impressive graphics of the obtained configurations and a generous offer to share the code.

Knots with minimal energies by Y. Diao, C. Ernst and E. J. Janse van Rensburg. The authors reach the conclusion that the best energy which could drive knots into their canonical representation is their “thickness”, experiments with real knotted ropes are compared with numerical simulations of the thick knots which are the same as ideal geometric representations of knots.

The writhe of knots and links by E. J. Janse van Rensburg, D. W. Sumners, S. G. Whittington. Certain geometrical properties of random knots in the cubic lattice show surprising correlations with corresponding properties calculated for ideal geometric representations of knots.

Minimal lattice knots by E. J. Janse van Rensburg. An ideal knot on the cubic lattice can be defined as that which requires the smallest number of edges to form a given knot. A theoretical and numerical approach to find minimal lattice knots is presented. Basic motifs presented in more complex knots show striking similarities to the motifs present in ideal geometric representations of these knots.

Minimal edge piecewise linear knots by J. A. Calvo and K. C. Millett. Knots presented in this chapter reach their specific ideality by requiring the smallest number of straight segments to form a given type of knot .

Entropy of a knot: simple arguments about difficult problem by A. Yu. Grosberg. This chapter discusses the available conformational space for different types of knots and concludes that it is related to the length/diameter ratio of ideal geometric representation of a given knot.

Approximating the thickness of a knot by E. J. Rawdon. The ratio between the length and the thickness of the shortest rope forming a given type of knot is a topological invariant, however defining the thickness for polygonal knots used in computer simulations is a complex problem which is resolved in this chapter.

Energy functions for knots: beginning to predict physical behaviour by J. Simon. This chapter and its preface introduces the concept of novel type of knot invariants based on specific physical properties of idealized shapes of knots and discusses whether certain idealized shapes are more ideal than others.

Physically-based stochastic simplification of mathematical knots by R. P. Grzeszczuk, M. Huang and L. H. Kauffman. This chapter describes optimization of a simulated annealing algorithm and compares it with the gradient descent method applied for the process of untangling very complex configurations of knots and unknots.

Knots and fluid dynamics by H. K. Moffatt. This chapter considers relaxation of knotted or linked solenoidal vector fields embedded in a fluid medium and demonstrates that minimum energy equilibrium states are closely related to ideal geometric representations of knots.

Knots in bistable reacting systems by A. Malevanets and R. Kapral. Flows of reagent in chemical reactions may form topologically stabilized patterns with two stable phases where one of the phases forms a knot immersed in the other phase. Relaxation of such patterns leads to assuming a nearly ideal shape by the knotted phase.

New developments in topological fluid mechanics: from Kelvin's vortex knots to magnetic knots by R. L. Ricca. Starting from a discussion of Kelvin's original conjectures on vortex atoms, the author reviews some of his new results on stability of vortex knots and dynamics of inflexional magnetic knots and considers possible implications for energy estimates.

Hamiltonian approach to knotted solitons by A. J. Niemi. This chapter demonstrates that Kelvin's proposal of stable knotted vortexes achieving nice regular shapes is now confirmed by modern simulations involving complex nonlinear equations of motion.

Energy of knots by J. O'Hara. In this chapter equilibrium configurations of self repelling knots are considered where the repulsion is governed by arbitrary energies related to Coulomb's repelling force. Depending on the type of energies the knots could have or not have, a unique configuration minimizing a given energy.

Möbius-invariant knot energies by R. B. Kusner and J. M. Sullivan. Among different arbitrary energies associated with trajectories of knots there is a class of energies which remain constant upon Möbius transformation (reflection in a sphere). Trajectories of knots which reach continuous minima of those energies are presented in this chapter.

An introduction to harmonic knots by A. K. Trautwein. This chapter presents such configurations of knots that the x , y and z coordinates of their trajectories can be expressed by continuous harmonic functions with a possibly small number of harmonic components. These configurations have nice symmetrical shapes.

Fourier knots by L. H. Kauffman. This chapter presents knots which have ideally simple and elegant parametrization of their trajectories via finite Fourier series, and thus have simple and elegant shapes.

Symmetric knots and billiard knots by J. Przytycki. This chapter presents selected aspects of knots symmetry visible in the regularities and symmetries of their algebraic invariants or in symmetric trajectories of the actual knots.

Andrzej Stasiak

Vsevolod Katritch

Louis H. Kauffman

This page is intentionally left blank

CONTENTS

| | | |
|------------|--|-----|
| Preface | | v |
| Chapter 1 | Ideal Knots and Their Relation to the Physics of Real Knots (6 colour plates) <i>A. Stasiak, J. Dubochet, V. Katritch and P. Pieranski</i> | 1 |
| Chapter 2 | In Search of Ideal Knots <i>P. Pieranski</i> | 20 |
| Chapter 3 | Annealing Ideal Knots and Links: Methods and Pitfalls (2 colour plates) <i>B. Laurie</i> | 42 |
| Chapter 4 | Knots with Minimal Energies <i>Y. Diao, C. Ernst and E. J. Janse van Rensburg</i> | 52 |
| Chapter 5 | The Writhe of Knots and Links <i>E. J. Janse van Rensburg, D. W. Summers and S. G. Whittington</i> | 70 |
| Chapter 6 | Minimal Lattice Knots <i>E. J. Janse van Rensburg</i> | 88 |
| Chapter 7 | Minimal Edge Piecewise Linear Knots <i>J. A. Calvo and K. C. Millett</i> | 107 |
| Chapter 8 | Entropy of a Knot: Simple Arguments About Difficult Problem <i>A. Yu. Grosberg</i> | 129 |
| Chapter 9 | Approximating the Thickness of a Knot (4 colour plates) <i>E. J. Rawdon</i> | 143 |
| Chapter 10 | Energy Functions for Knots: Beginning to Predict Physical Behaviour (2 colour plates) <i>J. Simon</i> | 151 |

| | | |
|---------------|--|-----|
| Chapter 11 | Physically-Based Stochastic Simplification of Mathematical Knots <i>R. P. Grzeszczuk, M. Huang and L. H. Kauffman</i> | 183 |
| Colour Plates | | 207 |
| Chapter 12 | Knots and Fluid Dynamics <i>H. K. Moffatt</i> | 223 |
| Chapter 13 | Knots in Bistable Reacting Systems (1 colour plate) <i>A. Malevanets and R. Kapral</i> | 234 |
| Chapter 14 | New Developments in Topological Fluid Mechanics: From Kelvin's Vortex Knots to Magnetic Knots <i>R. L. Ricca</i> | 255 |
| Chapter 15 | Hamiltonian Approach to Knotted Solitons <i>A. J. Niemi</i> | 274 |
| Chapter 16 | Energy of Knots <i>J. O'Hara</i> | 288 |
| Chapter 17 | Möbius-Invariant Knot Energies <i>R. B. Kusner and J. M. Sullivan</i> | 315 |
| Chapter 18 | An Introduction to Harmonic Knots <i>A. K. Trautwein</i> | 353 |
| Chapter 19 | Fourier Knots <i>L. H. Kauffman</i> | 364 |
| Chapter 20 | Symmetric Knots and Billiard Knots <i>J. Przytycki</i> | 374 |

CHAPTER 1

IDEAL KNOTS AND THEIR RELATION TO THE PHYSICS OF REAL KNOTS

ANDRZEJ STASIAK, JACQUES DUBOCHET

Laboratoire d'Analyse Ultrastructurale

Université de Lausanne

1015 Lausanne-Dorigny, Switzerland.

e-mail: Andrzej.Stasiak@lau.unil.ch

Jacques.Dubochet@lau.unil.ch

VSEVOLOD KATRITCH

Department of Chemistry

Rutgers the State University of New Jersey

New Brunswick, New Jersey 08903, USA.

e-mail: seva@dna2.rutgers.edu

PIOTR PIERANSKI

Department of Physics

Poznan University of Technology

Piotrowo 3, 60 965 Poznan, Poland

and

Institute of Molecular Physics,

Smoluchowskiego 17, 60 159 Poznan

e-mail: pieransk@phys.put.poznan.pl

We present here the concept of ideal geometric representations of knots which are defined as minimal length trajectories of uniform diameter tubes forming a given type of knot. We show that ideal geometric representations of knots show interesting relations between each other and allow us to predict certain average properties of randomly distorted knotted polymers. Some relations between the behaviour of real physical knots and idealised representations of knots which minimise or maximise certain properties were previously observed and discussed in Ref. 1-5.

1 Ideal knots

We learnt at school that the circle is the ideal planar geometric figure, since, amongst planar objects, it has the highest area to circumference ratio. Similarly the sphere is the ideal solid, since it has the highest volume to surface area ratio. The concept of ideal geometric objects was developed in Ancient Greece by Plato (427-347 B.C.), who said that we have the ability to intuitively grasp the ideal nature of those objects. Indeed intuition tells us, for example, that the equilateral triangle is the ideal triangle, and that the square is the ideal quadrilateral. Our intuition is corroborated by the fact that equilateral triangles and squares show the highest area to perimeter ratios for all triangles and quadrilaterals. Therefore, we can ask ourselves (as we would have asked Plato, given the chance) what are the ideal representations of different knots? Since knots occupy 3-D space, one should apply to them the principle of maximising their volume to surface area ratio. To attribute volume to knots in the most ideal way, one should consider them to be made of impenetrable cylindrical tubes of uniform diameter along their whole length. Obviously the highest ratio of the volume to surface area will be obtained for that configuration of the knot which allows us to minimise the length/diameter ratio of the tube forming a given type of knot.

One way to obtain ideal configurations is to uniformly increase the diameter of the cylindrical tube forming a knot while keeping the axial length constant and while adjusting the trajectory of the knot to permit the maximal radial inflation^{3, 6}. Equivalently, one can proceed with shortening of the knotted cylindrical tube or rope while keeping its diameter constant until its length is minimised and the knot type is still maintained⁷.

Fig.1 shows the process of rope shortening for the trefoil which initially has an arbitrary configuration. To begin with, axial shrinkage can take place everywhere and the axial trajectory of the knot therefore undergoes a simple scale reduction while the diameter of the rope remains unchanged. This leads to surface collisions of the rope segments, following which further shortening of the rope necessitates that its axial trajectory adjusts toward the ideal. Finally, further adjustments of the knot's trajectory and further shortening of the rope are not possible and we see that practically every radial cross-section of the rope forming an ideal trefoil shows a point of contact with the surface of another segment of the rope. We intuitively feel that such maximally shortened trefoil knot showing a nice symmetry indeed adopts its ideal representation. The situation becomes less intuitive when using simulation approach we try to bring more complicated knots into their ideal configurations. Colour Plates 1-3 show the results of such simulations performed for all prime knots

with up to 8 crossings. The notations accompanying the knots shown correspond to Alexander-Briggs notations⁸ used in standard tables of knots^{9, 10} where the main number indicates the minimum number of crossings possible for this knot type and the index indicates the tabular position amongst the knots with the same minimum crossing number.

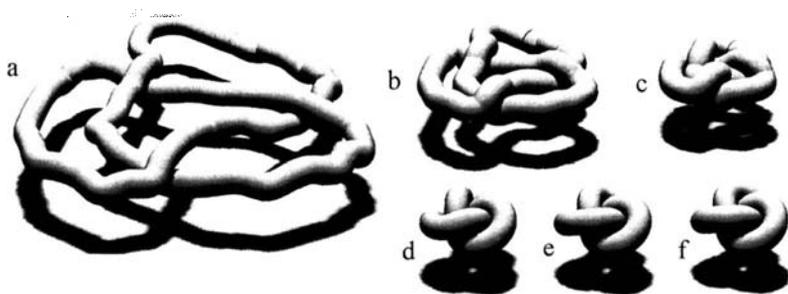


Fig. 1. Evolution of the trefoil knot towards its ideal conformation. The process of the knot tightening was simulated numerically with the SONO algorithm (see chapter by Pieranski). (a) - the initial conformation has an irregular shape. (b) to (e) - collisions between segments of the shrinking rope change its initial trajectory and allow maximal shortening of the whole rope. (f) - ideal conformation : the ratio of volume to surface area reaches its maximum.

It should be stressed here that the configurations presented are obtained by a simulation approach and it is still possible that they represent only a local rather than the absolute minimum in the configuration space of a given knot. We do not yet know how corrugated the conformational space of knots is or how many local minima and saddle points there are. We tested several different simulation algorithms in order to find procedures which are less likely to terminate in local minima. The algorithm used here (see chapter by Pieranski for its description) is less sensitive to local minima than the algorithm used previously to find ideal configurations of knots⁶. For example with the present algorithm we find that the symmetrical configuration of the knot 5_1 ; shown in Ref. 6 represents only a local minimum.

2 Ideal trajectories as knot invariants

Can the process of knot tightening be compared with finding a topological invariant of knots? If yes, then irrespective of the starting configuration of a given

knot, one should be able to obtain the same ideal configuration. Perhaps the best test for the knot tightening approach (and the computer algorithm used to simulate this process) is provided by the so-called Perko pair of knots. These two representations of 10 crossing knots were listed as distinct in knot tables compiled in 1899 by C.N. Little, and were believed by generations of topologists to correspond to different knot types, until, in 1974, K.A. Perko demonstrated that they represent the same knot. When Perko pair diagrams, denoted in Rolfsen tables as knots 10_{161} and 10_{162} , were taken as starting configurations for the process of knot tightening, the two configurations converged toward the same shape shown in Fig. 2.

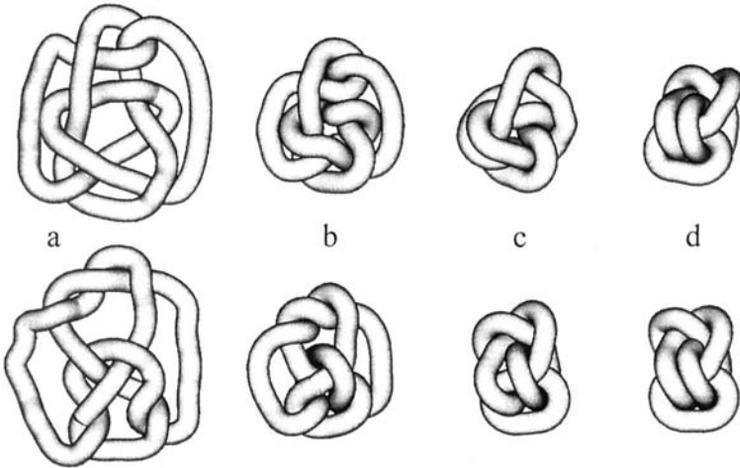


Fig. 2. Convergence of two different tabular representations of the Perko knot, into the same ideal form. (The final configurations of 10_{161} and 10_{162} are seen from slightly different angle)

3 Length/diameter ratio of ideal knots as a scale independent measure of the complexity of knots.

All knots shown in Plates 1-3 are constructed from rope of the same diameter. It is clear that as the knots become more complicated, one needs more rope to form the knot. To express the increasing complexity of ideal knots by a measure which is independent of the actual diameter of the rope, or the scale at which the ideal knot is presented, one can use the ratio of the length to the diameter of the rope forming a given ideal knot. Although we are not yet sure if our simulation approach will

always find the ideal configuration, it should be clear that one or more representation of a given knot exists for which the length to diameter ratio reaches the value of a global minimum. In the worst case there may be an infinite number of representations with the same length to diameter ratio but there will always be a length to diameter ratio value below which it is not possible to form a given knot. This minimum length/diameter ratio thus constitutes the topological invariant of a given knot. Of course without an analytical solution to the problem of ideal knots configurations we are limited in the precision of determination of the minimum length/diameter ratio. This is because our simulation approach is limited to polygonal knots made out of a limited number of segments of uniform size. However, we estimate that the error in the determination of the length to diameter ratio of ideal knots is less than 0.5%. Table I lists the length/diameter ratios obtained for different ideal knots.

The length to diameter ratio has a very simple intuitive meaning. It corresponds to the shortest relative length of very flexible but radially incompressible rope which is sufficient to tie a given type of knot. If one has, for example, a rope which is 1 cm thick, one would need at least 16.4 cm of this rope to form a trefoil (knot 3₁) and 47 cm to form knot 11₁: the more complicated the knot, the longer the piece of rope (of given diameter) required to tie it. A corollary is that the minimum length of rope needed to form a given knot should be a good measure of knot complexity. Knots could thus be classified and ranked according to the increasing length/diameter ratio of their ideal representations. Looking at Plate 3 and Table I we can see that the so-called nonalternating knots, which are placed in standard tables of knots as the last group of knots with a given minimum crossing number, e.g. 8₁₉, 8₂₀, 8₂₁, can be made with a piece of rope having a lower length/diameter ratio than alternating knots with the same minimum number of crossings. This result is easy to explain since in the alternating knots their trajectories are perfectly interwoven so that as we follow the rope forming ideal representations of alternating knots we see that it passes alternatively under and over the segments with which it crosses. In nonalternating knots the interweaving is not perfect and ropes can pass twice over and twice under the segments with which they cross and thus less rope is needed in-between such two crossings. Therefore, using as a criterion the length/diameter ratio of a rope forming the ideal representation of a given knot type, nonalternating knots are less complex than alternating knots with the same minimum crossing number. The fact that nonalternating knots can have shorter trajectories than alternating knots with the same minimum crossing number was observed also for strongly tightened knots made of real ropes (see the chapter by Janse van Rensburg) and for the minimum length knots on the cubic lattice¹¹(see also the chapter: "Minimal lattice knots" by Janse van Rensburg). Interestingly also

| TYPE | L/D | ACN | Wr | TYPE | L/D | ACN | Wr |
|------|-------|-------|-------|------|-------|-------|-------|
| 31 | 16.33 | 4.26 | 3.39 | 98 | 40.19 | 17.16 | 2.24 |
| 41 | 20.99 | 6.47 | 0.00 | 99 | 39.90 | 16.60 | 10.23 |
| 51 | 23.55 | 7.75 | 6.25 | 910 | 39.82 | 16.61 | 8.56 |
| 52 | 24.68 | 8.21 | 4.56 | 911 | 40.63 | 16.87 | 6.16 |
| 61 | 28.30 | 10.15 | 1.13 | 912 | 40.01 | 16.88 | 4.52 |
| 62 | 28.47 | 10.39 | 2.77 | 913 | 40.31 | 16.94 | 8.57 |
| 63 | 28.88 | 10.52 | 0.02 | 914 | 39.98 | 16.93 | 1.79 |
| 71 | 30.70 | 11.36 | 9.10 | 915 | 41.00 | 17.32 | 4.53 |
| 72 | 32.41 | 12.18 | 5.74 | 916 | 40.00 | 16.74 | 10.17 |
| 73 | 31.90 | 12.07 | 7.40 | 917 | 40.15 | 17.19 | 2.20 |
| 74 | 32.53 | 12.61 | 5.82 | 918 | 40.72 | 17.50 | 8.56 |
| 75 | 32.57 | 12.38 | 7.43 | 919 | 41.01 | 17.29 | 0.67 |
| 76 | 32.82 | 12.76 | 3.43 | 920 | 42.86 | 18.00 | 6.53 |
| 77 | 32.76 | 12.80 | 0.66 | 921 | 40.49 | 16.88 | 4.50 |
| 81 | 35.46 | 14.02 | 2.29 | 922 | 40.44 | 17.41 | 2.14 |
| 82 | 35.63 | 14.30 | 5.63 | 923 | 40.57 | 17.48 | 8.54 |
| 83 | 35.50 | 14.00 | 0.00 | 924 | 40.42 | 17.06 | 0.58 |
| 84 | 35.93 | 14.57 | 1.62 | 925 | 40.51 | 17.64 | 4.57 |
| 85 | 36.01 | 14.60 | 5.60 | 926 | 40.36 | 17.14 | 3.48 |
| 86 | 36.15 | 14.50 | 3.97 | 927 | 40.81 | 17.71 | 0.63 |
| 87 | 36.04 | 14.42 | 2.83 | 928 | 40.71 | 17.54 | 3.40 |
| 88 | 36.58 | 14.79 | 1.19 | 929 | 40.71 | 17.80 | 2.23 |
| 89 | 36.17 | 14.52 | 0.04 | 930 | 40.88 | 17.75 | 0.60 |
| 810 | 36.63 | 14.93 | 2.78 | 931 | 40.74 | 17.83 | 3.37 |
| 811 | 38.02 | 15.49 | 3.88 | 932 | 40.68 | 17.58 | 3.46 |
| 812 | 36.94 | 14.91 | 0.00 | 933 | 41.34 | 17.98 | 0.61 |
| 813 | 36.35 | 14.80 | 1.21 | 934 | 41.02 | 18.06 | 0.59 |
| 814 | 36.89 | 15.03 | 4.03 | 935 | 40.05 | 17.29 | 6.99 |
| 815 | 37.11 | 15.38 | 7.91 | 936 | 40.42 | 17.60 | 6.23 |
| 816 | 37.39 | 15.42 | 2.76 | 937 | 40.68 | 17.42 | 0.54 |
| 817 | 37.19 | 15.41 | 0.01 | 938 | 40.89 | 17.93 | 8.58 |
| 818 | 37.40 | 15.50 | 0.08 | 939 | 42.19 | 18.09 | 4.61 |
| 819 | 30.46 | 11.20 | 8.64 | 940 | 40.77 | 17.67 | 3.42 |
| 820 | 31.56 | 11.96 | 1.96 | 941 | 40.63 | 17.76 | 1.82 |
| 821 | 32.72 | 12.61 | 4.66 | 942 | 34.73 | 13.15 | 1.18 |
| 91 | 37.81 | 15.18 | 11.99 | 943 | 35.78 | 14.24 | 5.23 |
| 92 | 39.63 | 16.90 | 6.86 | 944 | 35.79 | 14.48 | 1.32 |
| 93 | 39.20 | 16.82 | 10.25 | 945 | 37.41 | 15.18 | 5.18 |
| 94 | 39.11 | 16.32 | 8.52 | 946 | 34.27 | 13.55 | 2.57 |
| 95 | 39.78 | 16.85 | 6.95 | 947 | 37.49 | 15.25 | 2.78 |
| 96 | 39.96 | 16.89 | 10.25 | 948 | 37.06 | 14.73 | 3.95 |
| 97 | 40.22 | 16.49 | 8.56 | 949 | 36.91 | 14.99 | 7.90 |

Table I Parameters of the most tight conformations found with the SONO algorithm.

the energies of nonalternating knots are smaller than the energies of alternating knots with the same minimal crossing number².

4 Average crossing number of ideal knots.

To characterise a knot one usually tries to find out what is the minimum number of crossings this knot can have in a planar projection. For example, there are 49 different prime knots which can be brought into configurations producing planar projections with the minimum number of crossings equal to 9, and all those knots are called 9 crossing knots. Minimum crossing number is by definition an integer and as such cannot be used for the fine distinction between different knots. Therefore as the minimum crossing number increases there are more and more knots with a given minimum crossing number (165 with 10 crossings). However, if instead of searching for the particular direction where a given knot shows minimum number of crossings one looks at the knot from infinitely many directions, one can calculate the average number of crossings which is a noninteger real number. Such a number, when calculated with high precision, can be uniquely attributed to the observed configuration of a given knot. Of course as long as we do not have unique representations of different knots, the value of the average crossing number will depend on the arbitrarily chosen configuration of a given knot. The situation is different for ideal knots which, at least in the case of not too complex knots tested by us, seemed to converge toward unique representations characteristic for a given type of knot. The average crossing number when calculated with high precision for the ideal configurations of knots is likely to be different for each knot type and as such can be used to distinguish between different knots. Table I lists the average crossing number calculated for the axial trajectory of ideal representations of different knots.

As with the length to diameter ratio of ideal knots, the average crossing number of the ideal configurations of knots is also a topological invariant. At the moment the only way to approach this value is by means of numerical simulations of limited precision, though elements of the analytical estimation have already been found (see chapter by Pieranski).

5 Relation between length/diameter ratio and the average crossing number of ideal representations of knots

When the values of the average crossing number of the ideal forms of 20 relatively simple knots (up to 11 crossings in their standard tabular representations) were plotted against the length/diameter ratio of the corresponding ideal forms, a quasi linear relation was obtained⁶. This prompted the question: is this relation strictly linear or is it a more complex one which only appears linear when relatively simple knots are taken into a consideration? Obviously the best way to answer this question is to take more complex knots, with minimum crossing numbers of several

dozens, and try to obtain their ideal configurations. However, as the knot becomes more complicated, the computer simulation gets more difficult and this increases the chances of obtaining final configurations which terminate in one of the local minima of the complex conformational space instead of the global minimum corresponding to the ideal configuration. In addition even if several independent simulation runs for the same knot type end in the same configuration we can not eliminate the possibility that there exists a better solution. Despite this uncertainty, we decided to simulate ideal trajectories of several more complex knots and to check their average crossing number and the length/diameter ratio. We have chosen torus type of knots for the simplicity of generation of their starting configurations.

Torus knots are defined as those which can be placed without self intersections on the surface of a regular torus. We took this class of torus knots which encircle twice the central straight axis of the torus and at the same time wind around the torus a given odd number of times. Since there are no tables of knots with more than 13 crossings other designations than the standard Alexander-Brigs notations are needed to describe more complicated knots. For torus type of knots the notation includes two numbers where the first one tells how many times the knot encircles the central straight axis of the torus and the second number tells how many times the knot winds around the torus (See (2,33)-torus knots on Fig. 3). Torus knots can be generated in nice symmetric configurations using a modified Lissajous approach where orthogonal oscillations are centred around a circular trajectory. Such symmetric torus knots evolve in an interesting way when they are used as the starting axial configurations for the simulation procedure (See Plate 4. and for more details see chapter by Pieranski). Initially the knots behave as if the tori about which knot axes wind were just getting thicker and shorter while the knots maintain a perfect cyclic group symmetry (C_{33} for the (2,33)-torus knot).

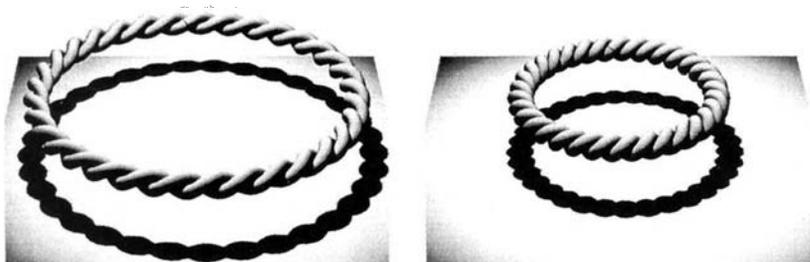


Fig. 3 The initial stage of evolution of the (2,33)-torus knot tightened with the SONO algorithm. For the next stages of the evolution see Plate 4.

Looking at the ropes forming symmetric torus knots one can see that they form a regular double helical arrangement whereby opposing ropes wind around a perfect circle.

During uniform shortening of the rope, axial paths of symmetric representations of torus knot change so that the opposing segments tend to become perpendicular to each other and then reach the optimum form (highest volume to surface area ratio) for this symmetric double helical arrangement (Fig 3b.) Further reduction in the rope length (maintaining fixed diameter) is only possible when the perfect double helical symmetry is broken and one of the opposing ropes becomes the central one. See Plate 4 and Fig. 4a. In this new arrangement the second, external rope can tightly wind around the central one. The symmetry breaking causes the whole knot to lose its toroidal shape upon which it folds into an irregular discoidal form whereby the opposing ropes switch between internal and external position in a non regular way.

In one of our long-duration simulation runs, a regular elongated form of torus knots emerged from the irregular discoidal form.

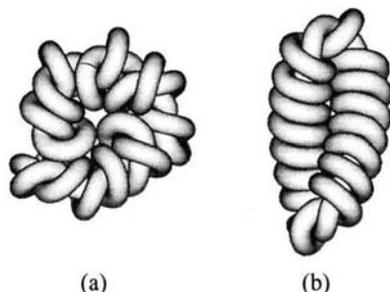


Fig. 4 Irregular discoidal and nice elongated form of (2,33)-torus knot obtained when starting from different initial configurations. See text and Plate 5.

In such elongated form there is only one place where the central and external rope exchange their positions (torus knots in contrast to torus links need to have a site where the central and external rope exchange). This site of rope exchange is located at one of the apexes of the elongated form (Fig. 4b and Plate 5). The elongated regular forms of torus knots with high crossing number demonstrated the shortest length, for a given rope diameter, of all the forms we have obtained. We assume therefore that these forms represent ideal configurations of torus knots. As seen in Plate 5, searching for the minimal rope length the elongated form becomes spontaneously twisted. Interestingly, for minimum length torus knots in the cubic lattice one also observes a tight winding of one strand around an essentially straight central strand (see Fig. 3 in the chapter "Minimal lattice knots" by Janse van Rensburg). As already indicated, our simulation procedure only rarely resulted in the generation of elongated regular forms when starting from the symmetric configurations of torus knots. However, when starting from configurations which were already constructed with one switch point between the external and central

rope, the simulations quickly reached a very regular elongated form with the shortest length for a given diameter (see the chapter by Pieranski). Therefore we were able to bring into this configuration very complex torus knots including $(2,63)$ -torus knots and calculate their average crossing number and length/diameter ratios. Fig 5 shows the relation between the length/diameter ratio and the average crossing number of ideal representations of $(2,n)$ -torus knots starting from the trefoil knot $(2,3)$ -torus knot and ending with $(2,63)$ -torus knot. It is visible that initial slope for less complex torus knots is lower than this for more complex torus knots. Therefore it seems that there is no universal linear relation between length/diameter ratio and average crossing number of ideal representations of all knots.

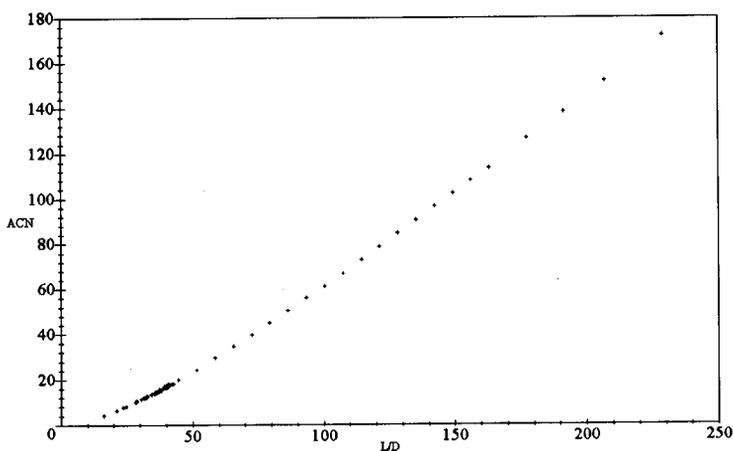


Fig 5. Relation between length/diameter ratio and average crossing number of the ideal geometrical representations of different knots, including the data obtained for large torus knots. Note a substantial change of the slope upon progression from simple toward more complicated knots.

However, as already mentioned, we are not sure that the elongated forms of large torus knots actually represent ideal forms of these knots. Furthermore we do not yet know if other large knots will follow the pattern of torus knots or if each family of knots will have its own characteristic relation between length/diameter ratio and the average crossing number of their ideal representations.

6 Writhe of ideal representations of knots.

Topologists are frequently interested in determining the chirality of different knots. Writhe is a measure of the chirality, and is quantified in a similar way as the mean crossing number. A given configuration of a knot is observed from thousands of directions, equisampling the sphere, and perceived crossings are counted.

However, in contrast to the mean crossing number, the chirality of the observed crossings is important. Thus, right handed crossings score as +1 and left handed as -1, and the writhe corresponds to the average score contributed by crossings perceived from a random direction, whereby the absolute writhe value is always smaller than the average crossing number. For achiral knots, the writhe of ideal configurations turned out to be practically equal to zero. Table I lists the writhe values calculated for ideal configurations of different knots. For chiral knots we provided writhe values of right-handed representations.

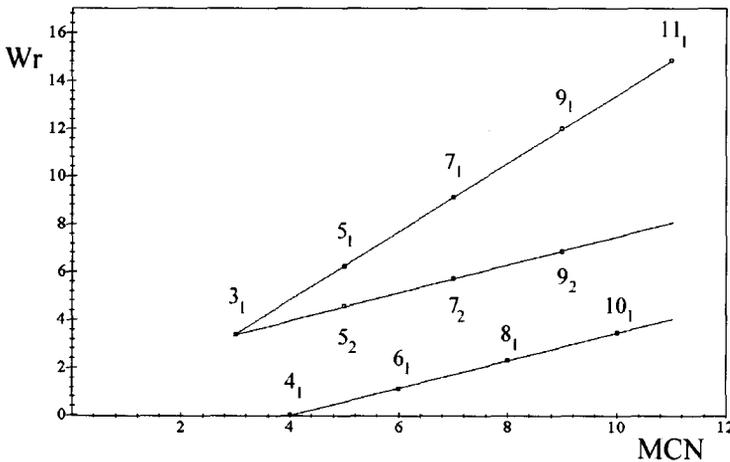


Fig 6. Different families of knots show different slopes of the linear increase of writhe Wr with increasing minimal crossing number (MCN). Knots $3_1, 5_1, \dots, 11_1$ belong to the family of the $(2,n)$ torus knots. Knots $4_1, 6_1, 8_1$ and 10_1 belong to the family of twist knots with the even number of crossings, while $3_1, 5_2, 7_2$ and 9_2 belong to the family of twist knots with the odd number of crossings.

Fig. 6 shows that when the writhe values of ideal representations of different knots are plotted against their minimum crossing numbers, one observes a linear relationship between the knots belonging to the same families. So, for example, torus knots fell on one line with a characteristic slope while twist knots with even and odd numbers of crossings in their tabular standard representations are found on two parallel lines with a lower slope. Thus ideal representations of knots naturally divide themselves into families of knots and this division can be perceived without analysing the overall topology of the knots but just by measuring their writhe.

7 Ideal composite knots.

Composite knots are formed when two or more knots are tied successively on the same string. Composite knots can always be redistributed on a string in such a way that a plane pierced by the knot trajectory in only two places can separate the knot into two parts, whereby each part upon simple closure can form a nontrivial knot.

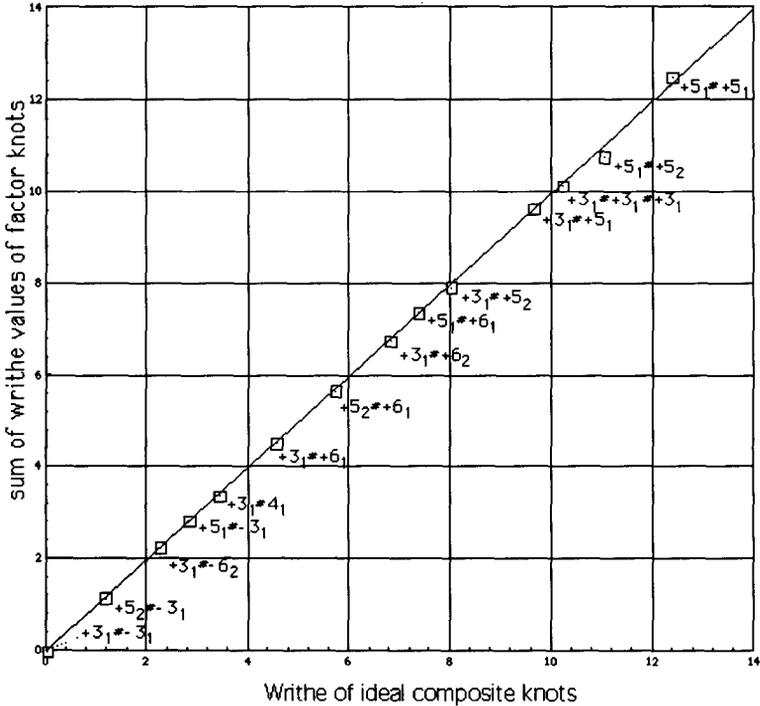


Fig. 7. Additivity of the writhe in composite knots.

Knots which constitute a composite knot are called factor knots. Of course the same criteria of ideality can be applied to composite knots as to the prime knots considered earlier. Plate 3 (lower part) presents the ideal configurations of some composite knots. Numerical notations indicate what factor knots are joined together to form a given composite knot, so for example $+3_1 \# -3_1$ indicates that this composite knot is formed from one left and one right-handed trefoil (knot 3_1). Interestingly, the writhe of the ideal composite knot $+3_1 \# -3_1$ turned out to be exactly zero as if positive writhe of a right-handed trefoil exactly compensated the negative

writhe of a left-handed trefoil (Table 2). Composite knot $+3_1\#-3_1$ is achiral¹² therefore our previous observation that the writhe of achiral prime knots in their ideal configurations is zero is apparently also valid for composite knots. When checking the writhe values for different composite knots in their ideal form we noticed that these values were always equal to the sum of the writhe values of the ideal forms of the factor knots constituting a given composite knot (Fig. 7). The observed additivity of writhe is a very interesting property of ideal knots since, usually, the writhe values are not additive¹³ i.e. when one arbitrarily connects two closed curves in space by removing fragments of the original trajectories and replacing them by almost straight connectors, the writhe of the newly formed curve usually differs from the sum of the writhes of the two starting curves. It appears that factor knots, in the ideal forms of composite knots, adjust their trajectory and their relative position in respect to each other in such a way that upon joining of two or more ideal knots their total writhe does not change.

8 Relation between ideal and real trajectories of knots.

Upon discussing geometrical and topological relations between the ideal representations of different knots it is good to consider the connections between ideal and real knots. Perhaps we should recall here some thoughts of Plato, who said that although real objects do not have the properties of ideal objects, some reflection of ideality should still be present in real objects.

To look for the reflection of ideality in real knots it may be useful to consider the average shapes of thermally agitated knotted polymers in solution. In order to obtain reliable statistics it would be best to have a direct physical technique providing us with precise information about writhe and crossing number of the configurations actually adopted by millions of knotted molecules in solution. Although, at present, there is no such experimental technique, the progress in computer simulation techniques permits reliable generation and analysis of millions of expected configurations of knotted polymers in solution¹⁴. Certainly the most studied polymer is DNA, and we understand DNA properties sufficiently well to be able to use the Metropolis Monte Carlo simulation technique to model DNA behaviour under given conditions¹⁵. In order to have the shape of knotted DNA molecules unaffected by the intrinsic stiffness of DNA, we have chosen to model long DNA molecules so that the local curvature needed to form a given knot is usually smaller than the average local curvature induced by thermal motion. If modelled DNA molecules were short they would tend to minimise the curvature and would therefore systematically deviate from the equilibrium of stochastic representations satisfying the topology of a given knot. To eliminate the possible effects of torsional stress in double-stranded DNA we modelled DNA with interruptions in one of the strands so that torsional stress can dissipate. The

simulation results presented below refer to knotted double-stranded DNA molecules placed in a solution, though a qualitatively similar behaviour is expected for any torsionally relaxed knotted polymers suspended in a solution. Plate 6 shows a comparison between the axial trajectories of ideal knots and some modelled trajectories of knotted DNA molecules undergoing thermal motion. The modelled trajectories correspond to 5400 base pair-long DNA molecules forming a 3_1 and 4_1 knot, respectively. It is clear that the trajectories of the knotted molecules undergoing thermal motion bear little resemblance to the ideal trajectories of the corresponding knots. When the writhe is measured for momentary configurations of thermally shaken knotted molecules, the measured values very rarely approach those of the ideal configurations of the given knots. However, when the writhe is measured for millions of configurations, the mean value then becomes practically equal to the writhe of the ideal configuration. As shown in Fig. 8a this is true for the average writhe of every knot modelled, and is also true for different lengths of modelled DNA molecules.

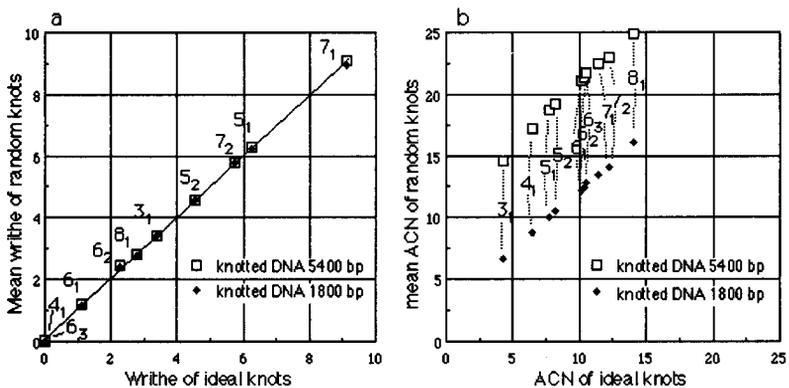


Fig. 8. Relations between ideal configurations of knots and knotted DNA chains undergoing thermal motion.

a. A 1:1 correspondence between the writhe of the ideal forms of knots and the mean writhe value of ensembles of ca. 9'000'000 configurations simulating the behaviour of knotted double-stranded DNA molecules (5400 and 1800 bp long) randomly distorted by thermal motion.

b. Linear correlation between the average crossing number of the ideal configurations and the mean of the average crossing numbers of ensembles of simulated configurations of knotted DNA chains with 1800 and 5400 base pairs.

The fact that the ideal configuration of a given knot has a writhe equal to the mean writhe of the floppy, randomly distorted polymers forming the same type of knot may initially be surprising. However, if one considers a simple circular polymer undergoing thermal motion one would expect that its momentary configurations can

be, with the same probability, twisted in right-handed and left-handed ways. Therefore the mean writhe, which is a measure of handedness or chirality should remain equal to 0, and a writhe of 0 is that of an ideal circle. Thus, for reasons of symmetry, floppy, circular polymers will maintain a mean writhe equal to that of an ideal circle. For the same reason floppy knots of different types will maintain their mean writhe equal to the writhe of the ideal configurations of these knots as long as the size of the knotted chains or of knotted domains is sufficiently large so that bending stress does not systematically change the shape of the knotted chains. Random knots built in the cubic lattice are free from bending stress and are ideal to test whether the mean writhe of a population of randomly distorted knotted chains forming a given knot corresponds to that of the ideal geometric representation of a given knot. Janse Van Rensburg *et al.*¹⁶ calculated the mean writhe for randomly generated populations of different knots in the cubic lattice and noticed that for a given type of knot the mean writhe of the population did not depend on the number of segments in the walk. For example, populations of trefoils built of 24 segments or of 250 segments showed essentially the same writhe, 3.40 ± 0.025 , which corresponds to the writhe observed for ideal trefoils presented here. For more details concerning the mean writhe of random knots on the cubic lattice see the chapter by Janse van Rensburg, Summers and Whittington.

Let us now look at the average crossing number of populations of knotted DNA molecules undergoing thermal motion. Fig. 8b shows that there is still a linear relationship between the average crossing numbers of ideal representations of different knots and the average crossing numbers calculated for populations of different knots with the same axial length. The slope of the observed linear relationship seems to be 1, and the longer the chain of the knotted polymers, the more additional crossings it has. Explaining the relationship between the average crossing numbers of the ideal representations of knots and those of a random population of knots one needs to point out the basic difference between the average crossing number and the writhe. Writhe is a measure of chirality so that additional crossings can have positive or negative values and, as discussed already, the accidental additional crossings have the same chance of adopting left or right handed configurations so that, on average, the contribution to writhe of additional crossings tends to 0. In contrast to the writhe, the average crossing number increases by +1 for each perceived crossing and thus additional crossings will never cancel each other but simply accumulate. The longer the polymeric chain undergoing thermal motion, the higher the number of additional crossings and this is exactly what is observed in Fig. 8b.

9 Gel electrophoresis as a physical method to test the relationship between ideal and real knots.

Upon observing 1:1 correspondence between the writhe of ideal knots and the mean writhe of modelled polymer chains forming corresponding knots one starts to appreciate that the ideal configurations predict or govern the real physical behaviour of knotted polymers in solution.

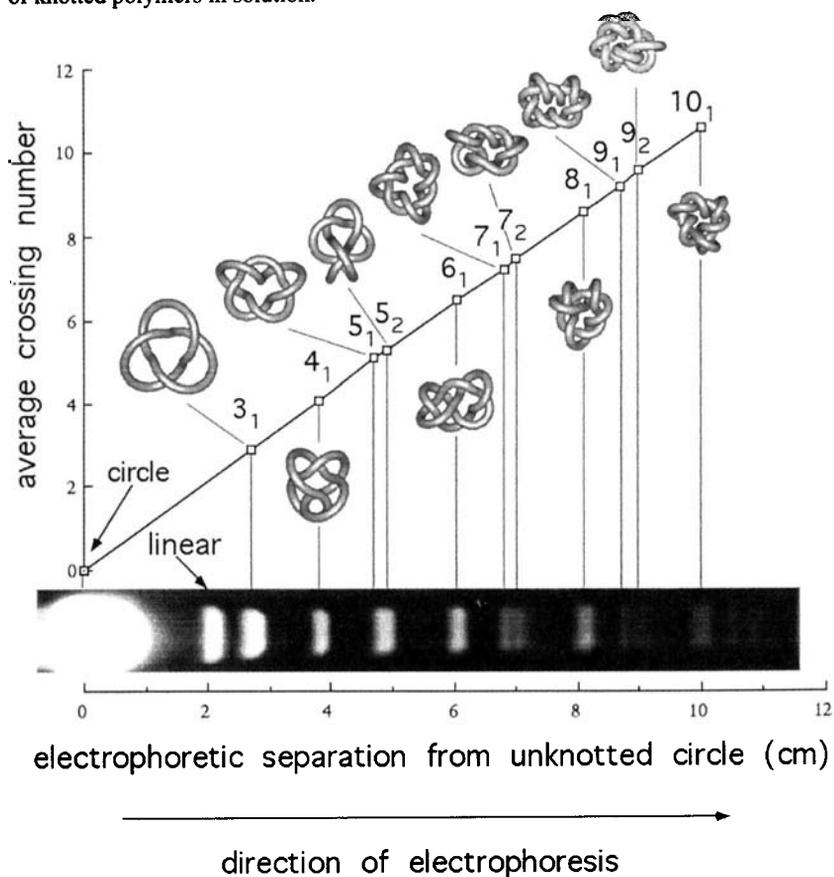


Fig. 9. Agarose gel electrophoresis separates DNA knots according to the overall compaction of ideal forms of the corresponding knots. The type of knot in each band was verified by electron microscopy. The ideal configuration of the corresponding knots were scaled to the same length and to facilitate visual tracing of the knots the diameter of the tubes was shrunk to 1/3 whilst maintaining the same axial path. The figure is adapted from Ref. 18

However, one may argue that the relationships described up to now are just between computer generated ideal configurations and computer generated random configurations of knots. A real physical test is needed. This test is provided by gel electrophoresis, a technique which can separate different DNA knots. In the laboratory of Nicholas Cozzarelli (University of Berkeley) this technique was brought close to perfection and Fig. 9 shows a gel where DNA molecules of the same length, though forming different well characterised knots, are analysed¹⁷.

Ideal forms of different knots with the same length can be obtained by appropriate scaling of ideal knots with the same diameter but different length (as those shown on Plates 1-3). Upon correct scaling to bring different ideal knots to the same length we see that as knots become more complex their overall dimensions decrease (see corresponding representations of the knots in Fig. 9). The same applies to knotted DNA molecules of the same length, the more complicated the knot the more compact is the average configuration of this molecule. For molecules with the same charge, the higher the compaction the quicker the migration on the gel. Since the average crossing number is a good measure of knots compaction it is thus natural to plot the migration distance of different knots versus the mean crossing number of their ideal representations. Fig. 9 shows that this relation is linear. This demonstrates that the ideal configurations of knots can be used to predict the physical behaviour of real knots. It should, however, be mentioned here that knotted molecules on the gel do not adopt ideal configurations: but only their average shape has a compaction proportional to the compaction of ideal configurations of these knots.

10 Conclusions.

Ideal configurations of knots showed interesting geometrical and topological relationships between different knots. These configurations provide a basis for a natural classification of knots. Most importantly, however, is the fact that the physical behaviour of real physical knots can be predicted by the properties of the ideal representations of those knots. This was demonstrated here for polymers in solution but may also apply to other systems including superstrings on the lowest scale and cosmic strings on the highest.

Idealised representations of knots which minimise or maximise certain properties were shown to have interesting relations to the expected behaviour of real physical knots formed by magnetic flux lines¹, polymers^{2,3}, bistable chemical systems⁴ or knotted fields⁵. (See also the chapters by some of the authors of these papers).

Several elements presented in this chapter were previously described in the Ref. 6,19-21.

Acknowledgements.

This work was supported by Swiss National Science Foundation grant 31-42158.94, Foundation Herbette, University of Lausanne, US public Health Service grants GM34809 and by Polish Scientific Committee grant 8T11-F010-08P04.

We thank Robert McNeel & Associates for providing us with the Rhinoceros Beta Release software with which most of the graphics presented in the chapter were created.

References

1. H. K. Moffatt, 1990, The energy spectrum of knots and links., *Nature* **347**, 367-369.
2. J. Simon, 1996, Energy functions for knots: beginning to predict physical behaviour., in *Mathematical Approaches to Biomolecular Structure and Dynamics* (eds. J. P. Mesirov, K. Shulten and D. W. Sumners) 39-58 (Springer-Verlag: New York).
3. A. Y. Grosberg, A. Feigel and Y. Rabin, 1996, Flory-type theory of a knotted ring polymer., *Phys. Rev. A* **54**, 6618-6622.
4. A. Malevanets and R. Kapral, 1996, Links, knots, and knotted labyrinths in bistable systems., *Phys. Rev. Lett.* **77**, 767-770.
5. L. D. Faddeev and A. J. Niemi, 1997, Stable knot-like structures in classical field theory, *Nature* **387**, 58-61.
6. V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet and A. Stasiak, 1996, Geometry and physics of knots., *Nature* **384**, 142-145.
7. P. Pieranski, 1996, Search of ideal knots, *ProDialog* **5**, 111-120 (in Polish).
8. J. W. Alexander and G. B. Briggs, 1927, On types of knotted curves, *Ann. of Math.* **28**, 562-586.
9. D. Rolfsen, 1976, *Knots and links* (Publish or Perish Press.: Berkeley, CA.).
10. C. C. Adams, 1994, *The Knot Book* (W.H. Freeman and Company: New York).
11. E. J. Janse van Rensburg, 1996, Lattice invariant for knots, in *Mathematical Approaches to Biomolecular Structure and Dynamics* (eds. J. P. Mesirov, K. Shulten and D. W. Sumners) 11-20 (Springer-Verlag: New York).
12. C. Liang and K. Mislow, 1994, On amphicheiral knots, *J. Math. Chem.* **15**, 1-34.
13. F. B. Fuller, 1978, Decomposition of the linking number of a closed ribbon: a problem from molecular biology, *Proc. Natl. Acad. Sci. USA* **75**, 3557-3561.
14. V. V. Rybenkov, N. R. Cozzarelli and A. V. Vologodskii, 1993, The probability of DNA knotting and the effective diameter of the DNA double helix, *Proc. Nat. Acad. Sci. U.S.A.* **90**, 5307-5311.

15. A. V. Vologodskii, S. D. Levene, K. V. Klenin, M. Frank-Kamenetskii and N. R. Cozzarelli, 1992, Conformational and thermodynamic properties of supercoiled DNA, *J. Mol. Biol.* **227**, 1224-1243.
16. E. J. Janse van Rensburg, E. Orlandini, D. W. Sumners, M. C. Tesi and S. G. Whittington, 1997, The writhe of knots in the cubic lattice., *Journal of Knot Theory and its Ramification* **5**, 31-44.
17. N. J. Crisona, R. Kanaar, T. N. Gonzalez, E. L. Zechiedrich, A. Klippel and N. R. Cozzarelli, 1994, Processive recombination by wild-type Gin and an enhancer-independent mutant. Insight into the mechanisms of recombination selectivity and strand exchange., *J. Mol. Biol.* **243**, 437-457.
18. A. V. Vologodskii, N. Crisona, B. Laurie, P. Pieranski, V. Katritch, J. Dubochet and A. Stasiak, 1998, Sedimentation and electrophoretic migration of DNA knots and catenanes, *J. Mol. Biol.* **278**, 1-3.
19. A. Stasiak, V. Katritch, J. Bednar, D. Michoud and J. Dubochet, 1996, Electrophoretic mobility of DNA knots, *Nature* **384**, 122.
20. A. Stasiak, 1997, Nœuds idéaux et nœuds réels, *Pour la Science, special issue: La Science des Nœuds* **April 1997**, 106-111.
21. V. Katritch, W. K. Olson, P. Pieranski, J. Dubochet and A. Stasiak, 1997, Properties of ideal composite knots, *Nature* **388**, 148-151.

CHAPTER 2

IN SEARCH OF IDEAL KNOTS

PIOTR PIERAŃSKI

Department of Physics

Poznan University of Technology

Piotrowo 3, 60 965 Poznan, Poland

and

Institute of Molecular Physics,

Smoluchowskiego 17, 60 159 Poznan

E-mail: pieransk@phys.put.poznan.pl

A particular version of the knot tightening algorithm is described. It is shown that the algorithm is able to remove empty loops and nugatory crossing leading to the simplification of the conformation of any knot. The problem of finding the ground state conformation of knots is discussed. Results of tightening various types of knots are presented and analysed.

1 Introduction

Knot is a closed, selfavoiding curve. From the topological point of view all conformations of the curve accessible via transformations during which no self-crossings occur are equivalent. From the point of a physicist, who thinks of the knot as a material object, whose points interact with forces stemming e.g. from a certain potential field, different conformations of a knot may have different energies. In such a case finding the ground state of the knot, i.e. the particular conformation at which the internal interaction energy reaches its minimum, is a challenging task. Our interest in the problem was stimulated by the seminal paper on the energy spectrum of knots and links by Moffat¹. A rigorous treatment of various formulations of the energy function and references concerning the history of the problem can be found in chapters by O'Hara and Diao et al. in this volume. Here, we shall limit ourselves but to the simplest case of the energy defined by the knot thickness. Putting aside all subtleties of the definition, one can state that the scale invariant parameter which determines in this case the knot energy (in what follows we shall refer to it as the *thickness energy*) is the ratio L/D , where L is the length of the rope (or tube) used to tie a particular conformation of the knot and D is the diameter of the rope.

The conformation at which the L/D parameter reaches its global minimum is the *ground state* of the knot. Knots in their ground state conformations are

sometimes called *ideal* - term coined by Stasiak². As it happens with ideal things, to reach ideal conformations of knots is not a trivial task: as we have found out, some of the ideal knots presented in Ref. 2 are in fact not ideal - we managed to find conformations for which values of the L/D parameter are smaller. But, are the values we managed to find *the minimal ones*? We cannot be sure. Thus, talking below less about the ideal knots we shall talk more about the tool we created to search for them.

2 The knot inflation process

Let K be a given knot type. Let C_K be a particular, smooth conformation of the knot at each point r of which the tangent unit vector t is well defined. Imagine now a process in which the curve C_K is slowly inflated into a tube T_K . At each point r of C_K the crosssection of the tube with a plane perpendicular to the tangent vector t is a disk of diameter D . If the curvature all along C_K is finite and smaller than a certain critical value, depending on the actual value of D , the tube the inflation process starts without problems. Depending on the actual shape of C_K , overlaps between different segments of the tube, or at its strongly curved parts, will soon appear. If by an appropriate procedure modifying the shape of C_K (keeping its length L fixed) the overlaps are removed, the process of inflation can continue until the conformation of the knot reaches a limit state above which the newly created overlaps can no more be removed.

The knot inflation process was introduced independently by a few authors^{1,3,4,5}. As easy to guess, instead of inflating the tube while keeping its length fixed, one can consider the complementary process in which the length of the tube is reduced while its diameter is kept fixed. Below we present the Shrink-On-No-Overlaps (SONO) algorithm we developed during our studies of braids⁶.

Our approach to the knot tightening process is experimental. It was not our aim to create a universal, autonomous algorithm able to find on its own the global ground state conformation of any knot. What we were aiming at was rather the creation of a virtual laboratory within which various experiments on knots could be performed, among them - tightening. The SONO algorithm is one of the tools found in the virtual laboratory. Other tools allow one to study various properties of the knots such as their curvature and twist maps, writhe and the average crossing number.

It is interesting that laboratory experiments with tightening knots tied on a real rope of have been performed not by physicists but mathematicians. The experiments are described in Ref. 4.

3 Shrink-On-No-Overlaps (SONO) ALGORITHM

Shrink-On-No-Overlaps algorithm is built from a number of procedures. Below we describe them one by one. Developing the SONO algorithm we tried to make it as fast as possible, thus simplicity of the applied procedures was of primary importance.

3.1 *ControlLeashes (CL) procedure.*

For obvious reasons, in the numerical experiments described below knots are discretised. Thus, let P_i , $i=1..N$, be equidistant points belonging to C_K . In what follows we shall refer to them as the *nodes*. $l=L/N$ denotes the length of *leashes*, which tie neighbouring nodes along the knot. The length of all leashes should be equal l . To keep an eye on it we defined a ControlLeashes procedure. It checks the distance $d_{i,i+1}$ between i and $i+1$ node and corrects its length to the proper L/N value.

If $d_{i,i+1} \lessgtr l$ then nodes i and $i+1$ are symmetrically moved (away or towards their centre of mass) along the line which passes through them to positions at which $d_{i,i+1}=l$.

The procedure starts at a randomly chosen node and runs, accordingly to another random choice instruction, up or down around the chain of nodes. Obviously, the CL procedure defined in such a manner is not selfconsistent; after its single application the leashes are not of equal length. On the other hand, it is also obvious that its frequent application should reduce the dispersion of the leash length. It does.

The CL procedure returns the lengths of the longest and shortest leashes found within the chain. Monitoring the values one can estimate the actual dispersion of the leash lengths.

3.2 *FindNeighbours(FN) and RemoveOverlaps (RO)procedures*

Nodes are points. To simulate a tube of diameter D we assume that each of the points is surrounded by a hard sphere of diameter D . As the number N of the nodes is large (or L is small) and the leash length l is smaller than D , the spheres surrounding consecutive nodes must be allowed to overlap. To achieve the aim, the hard core repulsion between spheres is defined in a particular manner: the spheres repel each other only if their index distance is larger than a certain *Skipped* integer.

The *index distance* between an i -th and k -th node is defined as the (smaller of the) number of leashes which separate nodes in question along the chain. A proper choice of the *Skipped* parameter is crucial. It should be larger than $\text{round}(D/l)$. One

should take into consideration situations in which the simulated tube makes the tightest U-turn

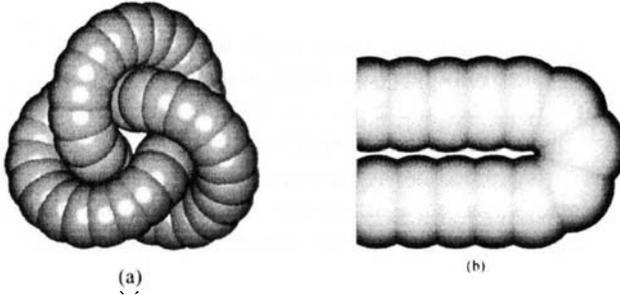


Fig. 1 (a) The discrete form of the 3.1 knot as seen by the SONO algorithm. The number of nodes $N=49$ is about 3 times smaller than the normalised standard value $\text{round}(10L/D)$. (b) The tight U-turn.

The sphere located at the entrance to such a U-turn should be blind to overlaps with all spheres found within the turn and it should start repelling all spheres which left the turn. Since the length of the tightest U-turn equals $\pi D/2$, the proper *Skipped* value should be close to $\text{round}(\pi D/2l)$.

Removing overlaps which appear in the knot during its tightening process is the most time consuming task. If the procedure which performs it is defined without imagination, its time consumption grows with N^2 . This happens if at each step of the tightening process possible overlaps of each node with each other node are checked. If the knot tightening process is slow, the evolution of the knot conformation is smooth and uniform and for a given node only a few nodes have a chance to overlap with it. (By a node we mean here the sphere which surrounds it.) Thus, before the overlap removing procedure is called, one should find for each node all nodes which have a chance to overlap with it. This task is performed by a FindNeighbours (FN) procedure which updates an integer array $\text{nn}[1..N, 1..m]$, whose i -th row contains indexes of all nodes which are found within a distance smaller than $(D+\epsilon)$, where ϵ is a small parameter adjusted experimentally. At the entrance to the FN procedure the nn array is zeroed. Then, consecutive rows are filled in. If the length m of the rows is chosen large enough, at the end of the FN procedure the rows are found to be only partially filled in: they end with sequences of zeros.

Assume that the nn array has been updated. Then, the RO procedure starts detecting and removing overlaps. The check starts at a randomly chosen nodes and, according to another randomly chosen parameter, it runs up or down the chain of nodes. For a given i -th node only the nodes whose indexes are found within the i -th row of the nn array are checked for overlapping. If the spheres surrounding the i -th and j -th nodes are found to be overlapping, the nodes are shifted apart,

symmetrically, along the line which joins their centres, to a distance $(D+\delta)$, where δ is another experimentally adjusted parameter. Putting $\delta>0$ leaves some extra free space between the shifted apart spheres. As found experimentally, this speeds up the tightening process.

The FN procedure is called much less frequently than the CL and RO. Typically it is called only every 200 iteration steps. As in the case of the CL procedure, the RO procedure is not selfconsistent. Removing overlaps between two nodes may create overlaps with other neighbours of the moved nodes, however, repeated use of the procedure asymptotically removes all overlaps (if the knot is not too tight already). The state of overlaps found within the chain is monitored by finding during the run of the RO procedure the values of the maximum and average overlaps.

3.3 *ShiftNodes (SN) procedure.*

As stated at the beginning, in the numerical experiments knots are represented by knotted, discrete chain of spherical beads threaded on C_K . Since the beads are spherical, the surface of the chain is not smooth. To avoid jamming we introduced into the SONO algorithm an additional procedure which shifts the beads along the C_K thread by an incommensurate fraction of the internode (leash length) distance, left or right. New positions of the beads are calculated via a simple, linear interpolation. The primary aim of the shifting procedure is to prevent jamming. Its sideeffect, which proved to be of great importance, is smoothening and cutting corners of C_K . The latter leads to some additional effects such as a rotation of the knot as a whole, which happens in the case of strongly chiral knots, e.g. the 3.1 knot.

3.4 *ReduceNodeNumber (RNN), DoubleNodeNumber (DNN) and NormalizeNodeNumber (NNN) procedures.*

Tightening a knot aimed at finding its global ground state is a difficult experiment. It aims at minimising $\Lambda=L/D$ vs. C , where denotes the actual conformation of the knot. Unfortunately, for most knots the $\Lambda(C)$ function displays a whole set of local minima. Any procedure aimed at finding the global minimum must be able to get out of them. In the thermal bath based Monte Carlo algorithms such as that used in Ref. 2, thermal fluctuations perform the task. In the mechanical knot tightening process simulated with the SONO algorithm the mechanism is different. Experiments prove that reduction of the number of nodes and increasing the δ parameter in the RO procedure allows one to cross some of the minima. We defined simple procedures which reduce twice (RNN) or double (DNN) the number of

nodes. The procedures are called, when necessary, by the operator performing the tightening experiment.

To obtain comparable accuracy of parameters for different types of knots, we defined also a procedure (NNN) which normalises the node number to a common standard value depending on Λ . In all experiments described below we assumed that the standard number of nodes is equal round($10 \cdot \Lambda$). This particular choice of the standard node number is of course arbitrary.

In addition to the procedures described above the numerical knot tightening workbench contains: procedures finding the total length L of the knot, its writhe Wr and the average crossing number ACN , the procedure finding and plotting the curvature map τ and, finally, the procedure which plots the actual conformation of the knot. Images of the knot conformation are stored on the hard disk what at any step of the tightening process allows the operator to call an animation procedure which displays the images in form of a short movie. Analysing the movie the operator may estimate the effectiveness of the tightening process and check if the knot did not change its type.

3.5 *The core of the SONO algorithm.*

The general flow diagram of the SONO algorithm is as follows.

1. The (x,y,z) coordinates of the nodes of a chosen knot are taken either from the procedure in which the knot is drawn with the mouse on the screen (the over- and undercrossings marked during the drawing process with mouse keys) or it is read from a file within which the coordinates were stored at one of the previous runs.
2. The nn array containing indexes of the near neighbours is filled in with by the FN procedure.
3. The lengths of the internode leashes are corrected by the CL procedure.
4. Overlaps are removed with the RO procedure.
5. If the average overlap value is smaller than a certain (defined by the operator) threshold, the knot is tightened: the (x,y,z) coordinates of the nodes and the leash length l are multiplied by a scaling factor $s < 1$, while the diameter D determining the overlapping remains unchanged.

Steps 3, 4 and 5 are repeated NumOfIt times, NumOfIt being of order $10^2 - 10^3$, before the FN procedure used in step 2 is called again.

4 TESTS OF THE SONO ALGORITHM

4.1 Untagling an entangled unknot.

Any procedure aimed at searching for the global ground states of knots should be able to perform such basic tasks as removing empty loops and nugatory crossings from even the most entangled conformation of a knot. Fig. 2 demonstrates how the SONO algorithm performed the task in the case of an entangled conformation of the trivial knot.

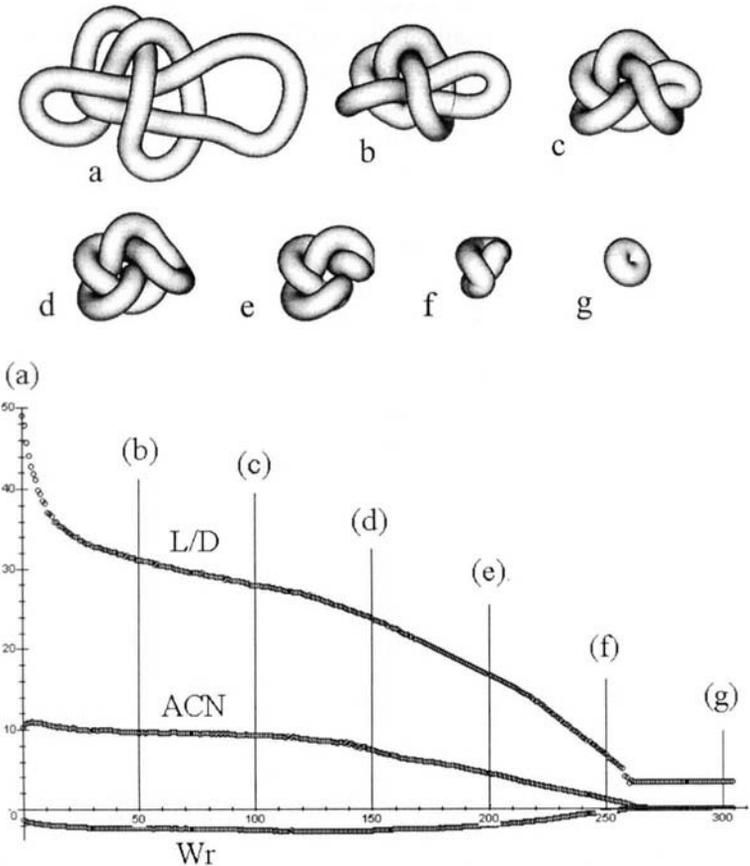


Fig. 2 Consecutive stages of the tightening process performed by the SONO algorithm on an entangled trivial knot. The run presented in the figure lasted about 5 min. on a PC Pentium 100.

In spite of its apparent effectiveness the SONO algorithm performs the untangling task in a rather awkward manner. Rather than tediously pulling out the empty loops it should just detect and cut them. Such a CutEmptyLoops procedure may be added to the tools it uses.

4.2 The Moffat test

Another test of knot tightening algorithms was suggested by Moffat. As he put it in his comment⁷ to Ref. 1: *"It would be interesting to test the algorithm on the simpler $T_{3,2}$ and $T_{2,3}$ configurations of the trefoil; it is not clear to me, how $T_{3,2}$ could flow to $T_{2,3}$ through the process described."*

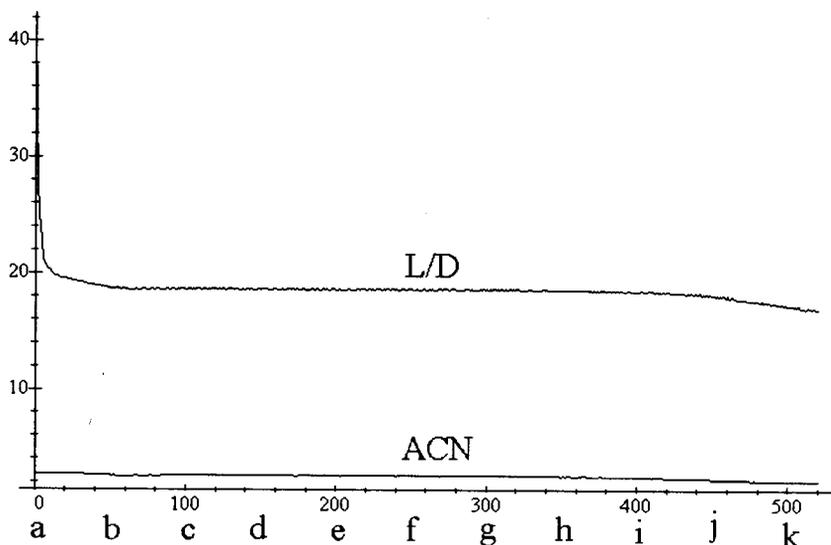
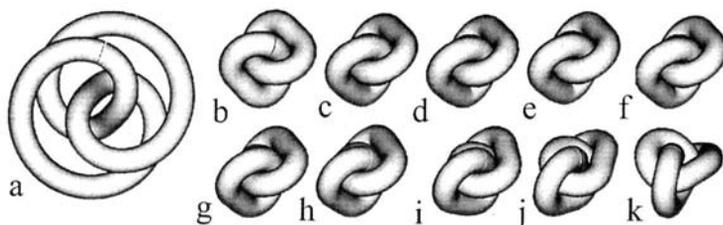


Fig. 3 Moffat test of the SONO algorithm. (a) the initial, loose conformation of the $T_{3,2}$ knot generated numerically. (k) the ground state configuration.

Fig. 3 presents result of the Moffat test we performed with the SONO algorithm. A symmetrical $T_{3,2}$ initial configuration (a) was generated numerically. Then, the configuration was read by the SONO algorithm and tightened. As clearly seen in the figure, after the very rapid initial stage, in which the loose initial configuration became tight (b), the evolution practically stops. The L/D and ACN parameters remain on average constant: a local minimum or saddle of the knot energy has been reached. However, as seen in the figure as well, the L/D value fluctuates. The fluctuations are induced mostly by the ShiftNodes procedure; let us remind that the discrete representation of the knot is not smooth, see Fig. 1. As a result of the fluctuation the symmetry of the $T_{3,2}$ configuration becomes broken (h), and the knot slips into the ideal $T_{2,3}$ configuration. Let us note that the time needed to break the symmetry of the $T_{3,2}$ configuration strongly depends on the number of nodes. A considerable, temporary reduction of the number allows the operator to initiate it. During the run presented in the figure all parameters of the SONO algorithm remained fixed. We performed the Moffat test for other torus knots. In all studied cases the SONO algorithm managed to break the n -fold symmetry of the $T_{m,n}$ knot.

4.3 The Perko Pair test

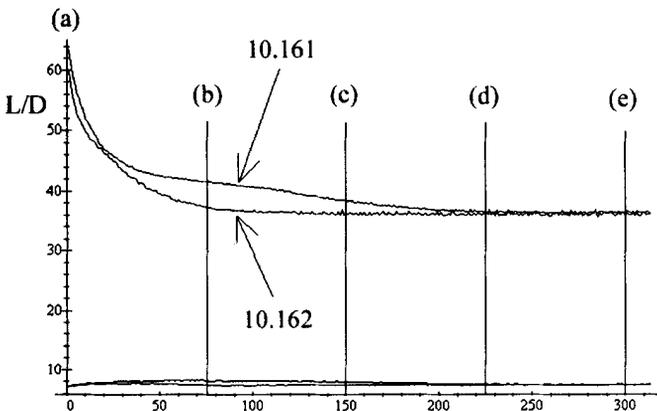


Fig. 4 Evolution of knots from the Perko pair to the identical final conformation.

Another test, which any algorithm aimed at finding the ideal conformations of prime knots should pass is the ability to bring knots 10_{161} and 10_{162} , the Perko pair, to a single, minimum energy conformation. The test was passed by the tools used in Ref. 2. In chapter by Stasiak et al. we presented how the task of bringing the Perko pair knots to the same conformation was performed by the SONO algorithm.

Initial conformations (a) of the Perko pair knots were defined by redrawing on the screen the conformations found in „*The Knot Book*” by C. C. Adams, p.32. The conformations were tightened separately with the SONO algorithm. Fig. 2 presented in chapter by Stasiak et al. shows consecutive stages of the knots evolution. As checked by an additional knot comparing procedure, although differently oriented, the final conformations of the knots were identical.

Fig. 4 presents the evolution of L/D for both knots. As seen in the plot, it is rather the 10_{161} knot which has a problem with finding the proper shape; evolution of the 10_{162} knot is rapid and leads immediately to the final conformation.

5 In search of ideal prime knots

First pictures of the ideal prime knots were presented in Ref. 2. As mentioned above, tests we performed with the SONO algorithm revealed that some of the configurations presented were not ideal.⁸ Below we present a numerical experiment which illustrates well the problems encountered during the search for the ideal conformations.

The first, most spectacular case is the 5_1 knot. Its conformation presented in Ref. 2 has a 5-fold symmetry axis and $L/D=24.2$. SONO algorithm managed to break the symmetry arriving at a conformation for which $L/D=23.5$. Fig. 5 presents the symmetry breaking process observed in a single run lasting on a PC 100MHz about 2 min. The run started from a symmetrical $T_{2,5}$ conformation generated numerically (a). Initially, (a) - (c), the evolution was very fast until the knot became tight (c). Then, the evolution stopped (c), (d), (e) - the knot entered a local energy minimum. The stable tight conformation preserves the 5-fold symmetry axis present in the initial conformation. To induce the symmetry breaking, the δ parameter which determines the distance to which the overlapping nodes are shifted by the RO procedure was temporarily increased to 0.1 (previously it was equal 0.00001). This resulted in strong fluctuations of the knot conformation visible within the L/D plot. Figures (f), (g), (h), (i) present evolution of the knot shape towards the asymmetrical conformation (j).

Obviously, the knot conformation (j) to which the SONO algorithm arrived in the single run described above is not of a good quality. First of all, the number of nodes, $N=46$, of which it is built is too small to provide reliable values of its L/D , ACN , Wr parameters. In a standard procedure we apply to clean knots, its node number is normalised to $\text{round}(10 \cdot L/D)$ and the overlaps are carefully removed. We do not describe the cleaning procedure in more detail, although its application is quite essential if reliable data are to be obtained.

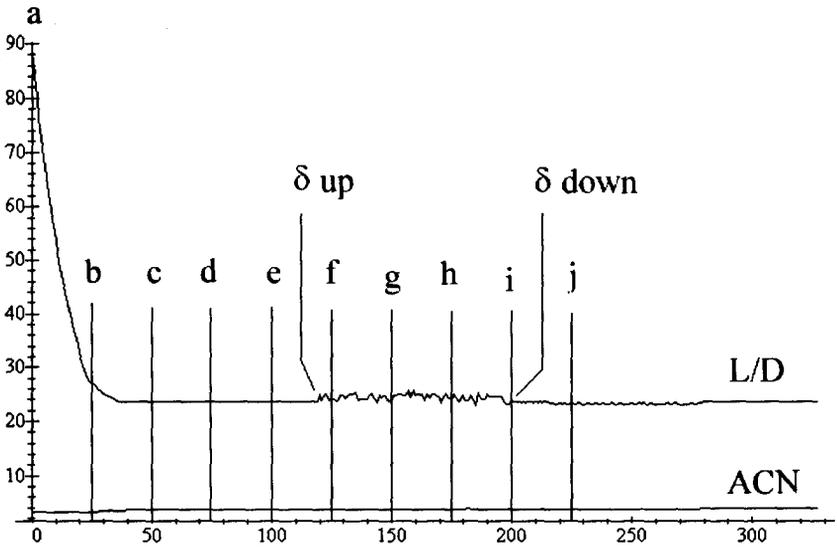
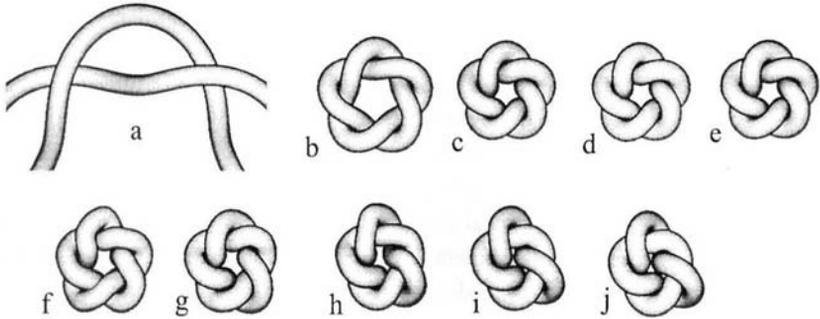


Fig. 5 Symmetry breaking of the 5_1 torus knot observed during the tightening of the knot with the SONO algorithm. See text.

The local minimum encountered during tightening of the 5_1 knot has been found also in the case of larger $T_{2,m}$ knots (m_1 knots in the Rolfsen classification). The Pascal code we are using allows us to study $T_{2,m}$ knots up to $m=63$.

The case discussed above clearly demonstrates that the knot inflation process may encounter problems when getting into the local minima within the thickness energy function. On the other hand, the case demonstrates also that the tools built into the SONO algorithm may in some cases allow the operator to force the knot to leave them. In the end, however, we cannot be sure if the conformation to which the knot eventually arrives is the final one, i.e. if the global energy minimum was

reached. At the present state of the art of tightening knots we must admit that we have no certainty if the conformations to which we arrived are the ideal ones. On the other hand, due to the clear definition of the thickness energy the conformations obtained can be compared: the lower L/D , the better.

5.1 Prime knots up to 9 crossings.

The race for the lowest thickness energy conformations was initiated with the Monte Carlo annealing procedure whose results were presented in Ref. 2. The best conformations to which we arrived applying the SONO algorithm to all knots up to 8 crossings are shown in the Plates 1-3. L/D , ACN and Wr parameters of the conformations are listed in Table I found in chapter by Stasiak et al. Here we present two plots covering a slightly larger range of knots.

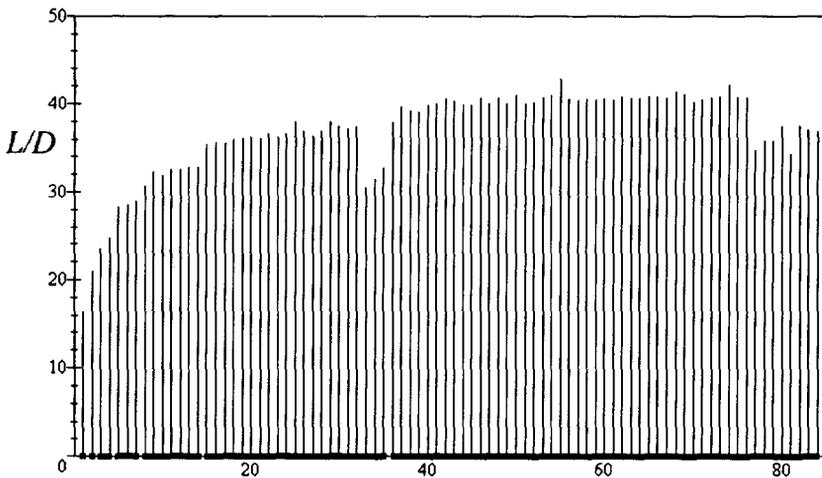


Fig. 6 The thickness energy for consecutive knots up to 9 crossings. Thickness energy is plotted in units of the rope diameter D .

The first one, Fig. 6, analogous to Fig. 2 found in Ref. 4, presents the L/D values of the consecutive prime knots up to 9 crossings in the order they are listed in the Rolfsen table. In the terminology of Ref. 3, the plot presents the *closed thickness energies* in units of the rope diameter D .

Note the drop, visible in Fig. 6, in the L/D value as the nonalternating knots are reached in classes of both 8 and 9 crossing knots.

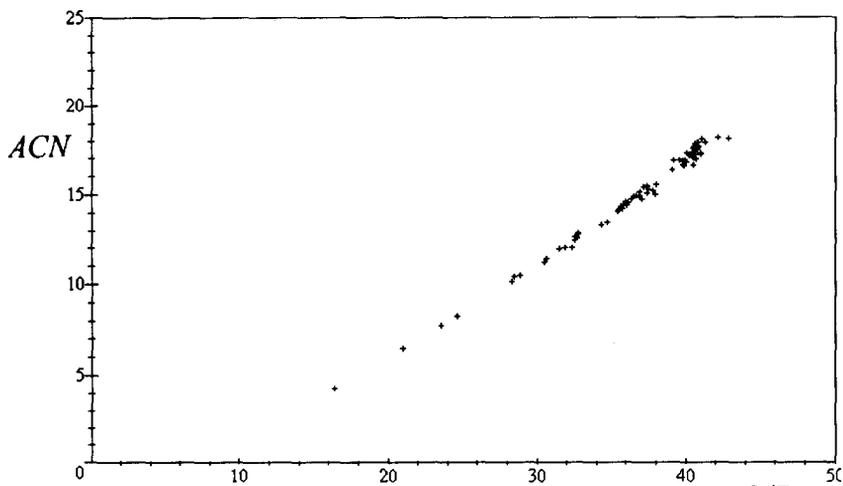


Fig. 7. Average crossing number ACN of prime knots up to 9 crossings vs. L/D

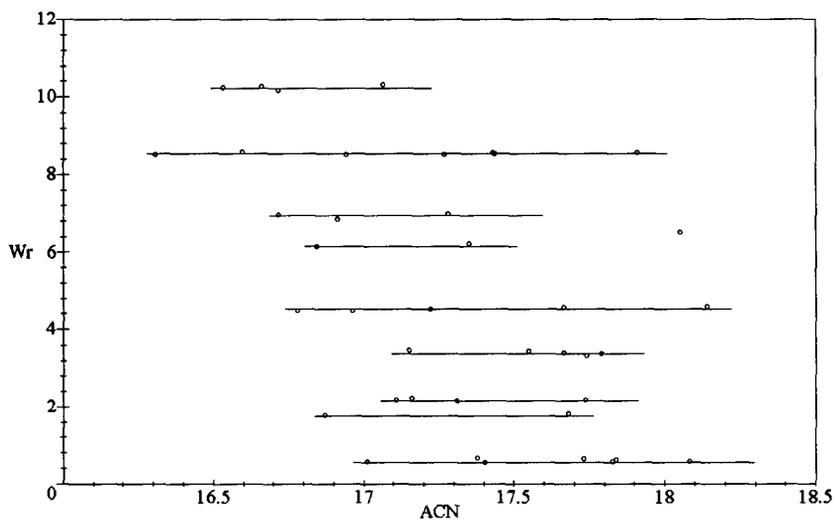


Fig. 8. Writhe vs. average crossing number for all 9 crossings knots.

The second plot, Fig. 7, follows the idea presented in Ref. 2 and correlates the L/D values with the average crossing number ACN . Note, that the L/D vs. ACN relation, apparently linear at the beginning, is clearly nonlinear in a larger interval.

Trying to find the quantitative description of the relation we studied larger torus knots $T_{2,m}$, which in the Rolfsen notation are denoted as m_1 knots, m - odd. The nonlinearity of the ACN vs. L/D relation is clearly visible in Fig. 5 found in chapter by Stasiak et al., where we plotted data of all knots we studied.

Another essential parameter which characterises optimised conformations of knots is their writhe. Fig. 8 presents the variable correlated with the average crossing number. As clearly seen in the figure, the writhe is a variable, which within a family of knots with the same minimal crossing number takes values localised near a few, well defined levels. Reasons for this quasiquantization of writhe is not clear to us. For a discussion of the writhe vs. minimal crossing number correlations see chapter by Stasiak et. al.

Another problem, which puts in doubt the idea of using the parameters of the ideal conformations as the knot type identifiers, is the existence of knots which being of different type have almost identical values of their L/D , ACN and Wr parameters. This is well seen in the class of knots with $MCN=9$. Table 1 below presents values of the parameters for two pairs of knots for which the differences are smallest.

| Knot type | | L/D | | ACN | | Wr | |
|-----------------|-----------------|-------|-------|-------|-------|-------|-------|
| 9 ₆ | 9 ₁₆ | 39.97 | 40.00 | 16.87 | 16.77 | 10.26 | 10.19 |
| 9 ₁₈ | 9 ₂₃ | 40.71 | 40.58 | 17.48 | 17.47 | 8.57 | 8.56 |

Table 1 L/D , ACN and Wr parameters of two pairs of the most similar knots.

The differences are below the estimated accuracy of 1% with which the L/D , ACN and Wr parameters are determined. (In the table the values are given with excessive accuracy.) We are thus forced to conclude, that in practice the set of L/D , ACN and Wr values does not determine in an unambiguous manner the knot type.

5.2 Tightening of the torus knots $T_{2,m}$

To check the ACN vs. L/D dependence in a larger interval of L/D we tightened with the SONO algorithm the sequence of $T_{2,m}$ knots, m odd, which initiate in the Rolfsen notation all classes of knots with m crossings.

Initial configurations of the knots were generated numerically according to equations:

$$\begin{aligned}
 z(t) &= r(t) \cos(2\pi v_1 t), \\
 x(t) &= r(t) \sin(2\pi v_2 t), \\
 y(t) &= r(t) \cos(2\pi v_2 t), \\
 &\text{where}
 \end{aligned}$$

$$r(t) = R_0 + R_1 \sin(2\pi \nu_1 t);$$

R_0 and R_1 are the radii of the torus onto the surface of which the knot trajectory is defined. Putting $\nu_1 = 2$ and $\nu_2 = m$ one obtains circular conformation of the $(2, m)$ torus knots. In what follows we shall refer to them as the $T_{2,m}$ knots.

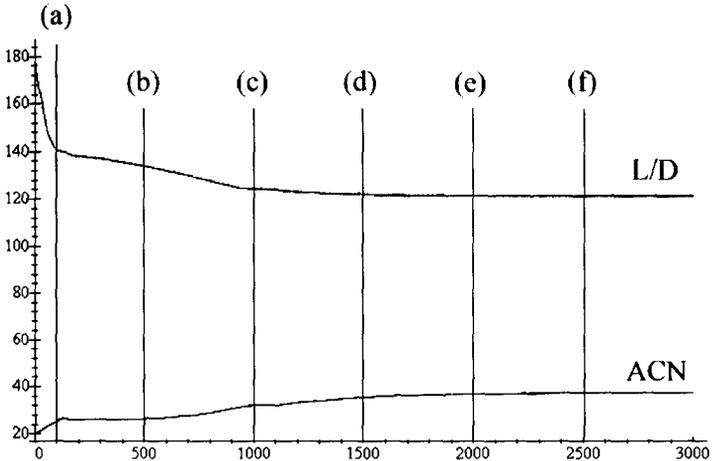
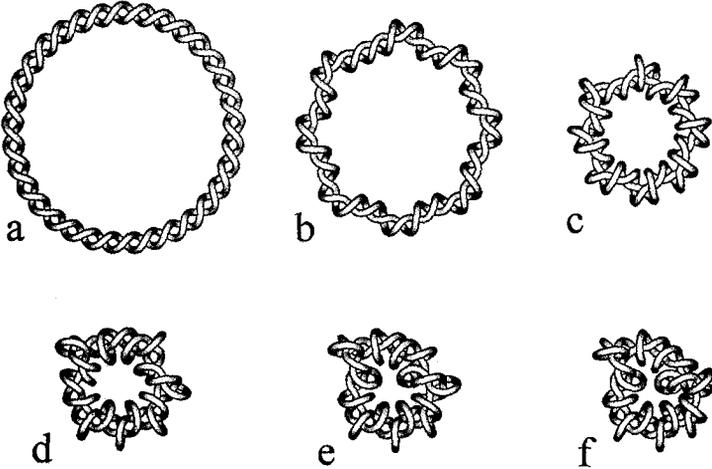


Fig.10 Evolution of the $T_{2,33}$ knot during the tightening process. See also Plate 4.

Evolution of the L/D and ACN parameters of the knot are presented in Fig. 10. For the sake of the clarity the tube of which the knot is made is drawn has a diameter

equal $D/2$. The initial conformation of the knot is not shown. (a) presents the symmetrical conformation to which the initial conformation is rapidly tightened. As seen in (b) the SONO algorithm breaks it easily. In 8 places the arrangement of the tube becomes different. Fig 10. presents a fragment of the configuration in which 3 such places are well visible. The overall shape of the knot becomes almost octagonal; note that one side of the octagon is longer. This cannot be avoided, since the number of crossings is odd. This symmetry breaking leads to (c) - another typical conformation of the knot. The existence of the longer side leads to a another conformation of the knot (d), which eventually stabilises (e) and (f). Conformation of this type is characteristic to larger $T_{2,m}$ knots. Is it the ideal shape, or just one of the local minima? As stated in Chapter by Dubochet et al., we have found that there exists a better conformation. We discuss it in more detail in the next subsection.

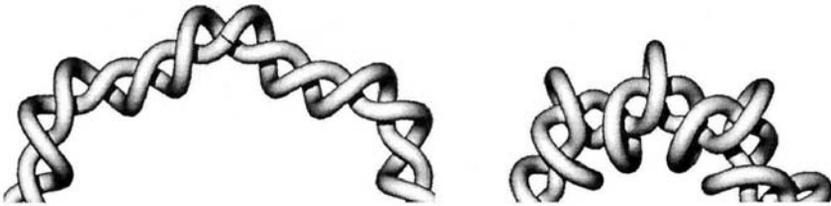


Fig.9 A typical arrangement of the tube in the $T_{2,33}$ torus knot in which the symmetry was broken.

5.3 Towards the ground state conformation of the $T_{2,m}$ knots

Tightening of large $T_{2,m}$ knots starting from their toroidal conformations leads to their compact, globular conformations. Experiments we performed prove that there exist better, with a lower L/D value, conformations of the knots. Fig. 11 present such a conformation of the $T_{2,33}$ knot.

As seen in the figure, structure of the initial conformation, from which the tightening procedure starts, is utterly different from the typical conformation of torus knots. In the conformation the tube is divided into two parts: the first one, bent into the a safety pin shape, forms the core around the second part of the tube is helicoidally wound.

The initial stage of the evolution process, during which all loose parts of the tube are shortened, is very rapid. In the next, much slower stage, the tightly packed safety-pin conformation becomes twisted. The rate of the twist is approximately the same in all torus knots and leads to a kind of a double-helix conformation. As noted by Stasiak⁹, the core part of the tube is not straight: under the pressure induced by

the outer, helicoidal part it becomes helicoidally deformed as well. Analysis of the phenomenon lead us to formulate the following problem:

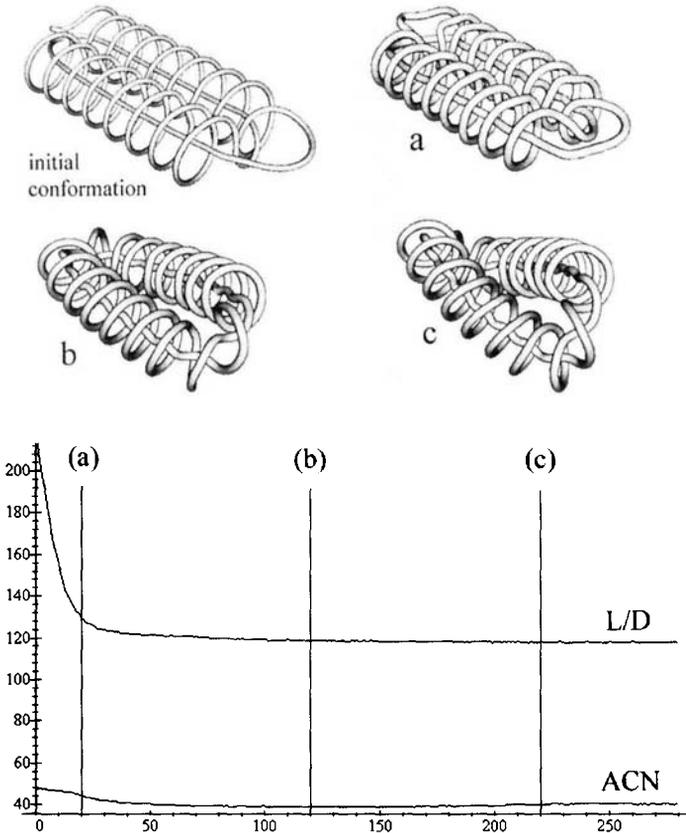


Fig.11 Evolution of the $T_{2,33}$ knot from the safety-pin initial conformation. The initial conformation is presented in a slightly different scale.

Which is the periodic conformation of two twisted together tubes, at which the L/D ratio reaches a minimum?

Answering the question may shed some light on the nature of the ground state conformation of the $T_{2,m}$ torus knots. In a different wording it asks how much of the tube we need to produce two crossings. The problem is being studied by Sylwester Przybyl¹⁰. Fig. 12 presents four different conformations he considered.

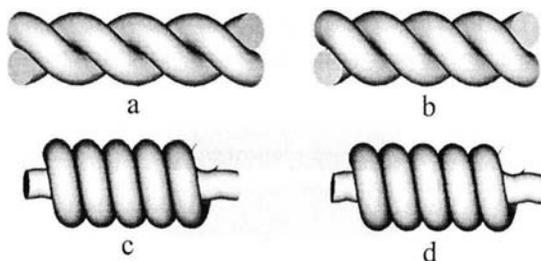


Fig. 12 Four closely packed conformations of twisted tubes. L/D per one period equals, respectively: (a) 8.886, (b) 8.5013, (c) 7.2261, (d) 7.2013. See text.

Let us describe in a few words the differences between them. All of the conformation can be described parametrically by two sets of equations:

$$\begin{aligned}
 x_1(t) &= Pt & x_2(t) &= Pt \\
 y_1(t) &= R_1 \sin(2\pi t) & y_2(t) &= -R_2 \sin(2\pi t) \\
 z_1(t) &= R_1 \cos(2\pi t) & z_2(t) &= -R_2 \cos(2\pi t)
 \end{aligned}$$

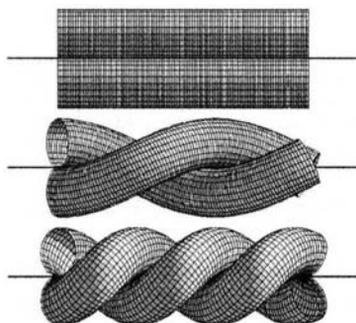


Fig. 13 The simplest way of twisting two tubes together. (a) Take two straight tubes and put them in parallel so that they touch along a common line. (b) Start twisting them together so that they remain in touch along the initial contact line. (c) Stop twisting, when the pitch of the structure equals π - overlaps are just about to appear.

In the first conformation, Fig. 13, the tubes are wound around and touch each other along the line lying between them: $R_1=R_2=0.5D$. The pitch of the periodic

structure $P=\pi D$. The thickness energy per period $L_p/D=8.8858$. By L_p we denoted here the length of the tubes found within one period of the structure.

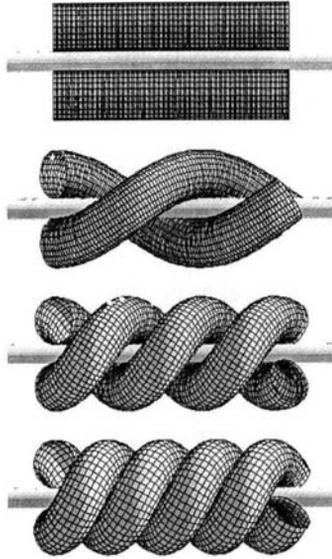


Fig. 14 The more clever way of twisting two tubes together. (a) Take two straight tubes and put them in parallel so that they do not touch. Insert a virtual cylinder between them. (b) Start twisting them together so that they remain in touch with the virtual cylinder. (c) Do not stop twisting when the pitch reaches the π value. (d) Stop twisting, when the tubes come in touch with each other. (Do not forget to remove the virtual cylinder.)

In the second conformation, Fig. 14, the radii of the spirals defining axial curves running inside the tubes are slightly larger than the radii of the tubes: $R_1=R_2=0.5229D$. As a result, a small hole appears along the x axis. The pitch of the structure is smaller than π : $P=2.6967D$. The tubes touch each other along two spiral curves running between the tubes. The thickness energy per period $L_p/D=8.5013$.

In the third conformation, Fig. 15, the radii of the spirals are different, but their sum is equal D : $R_1=0.97326D$, $R_2=0.0267D$. Pitch of the structure is much smaller than previously: $P=1.0136D$. The thickness energy per period $L_p/D=7.2261$. Note, that the value of R_2 is limited by condition that the curvature of the inner spiral cannot be larger than $2/D$.

In the fourth and it seems best, conformation (we do not show its construction details because visually it differs too little from the third one) $R_1=0.9670D$, $R_2=0.0377D$, thus, their sum is larger than 1. The minimal pitch $P=1.0139D$. The thickness energy per period $L_p/D=7.2013$.

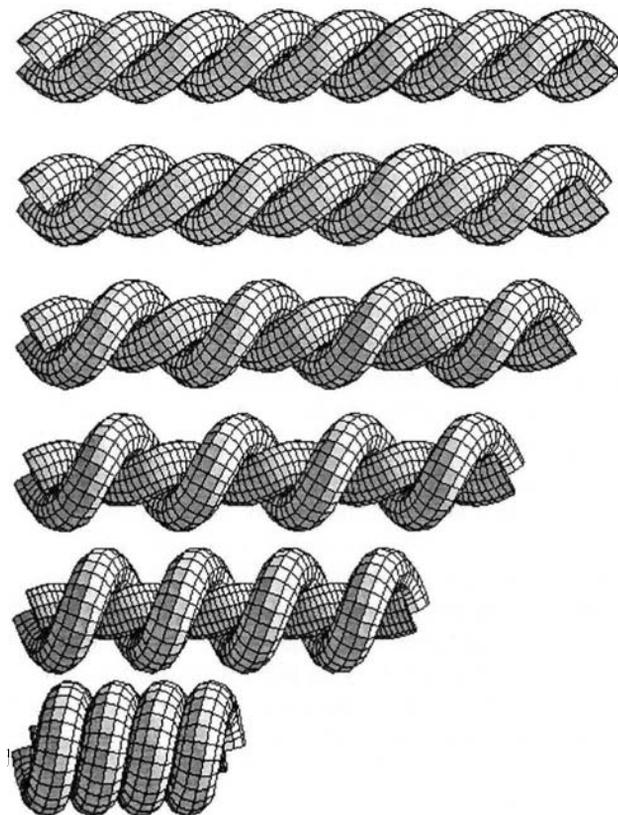


Fig. 15 Take the close packed structure obtained by procedure described in the caption of Fig. 13 and, remaining all time at the limit above which overlaps between the tubes appear, start increasing the radius of the spiral formed by one of the tubes decreasing at the same time the radius of the spiral formed by the other one. Stop the process, when the first spiral becomes close packed.

Looking at the numbers cited above, one can immediately see the reasons for which the m -fold symmetry of the initial conformations of the $T_{2,m}$ knots must be broken.

Conclusions

The simplest experimental procedure one can think of when trying to find the most tight conformations of a knot is to tie it on a rope and try to tighten it as much as possible by pulling the ends of the rope. To stay in agreement with what in mathematics is meant by a knot, one cannot forget at the end of the process to cut the spare parts and join the ends of the rope together.

The technique of numerical simulations provides us with a chance to perform such experiments in a much more clean manner:

- using the perfectly flexible and slippery, but at the same time perfectly hard in its circular crosssection rope,
- shortening the rope without the necessity of cutting it.

The SONO algorithm we described performs the simulation task in the most simple and thus effective manner. Using it we managed to find more tight conformations for a few from the knots considered in Ref. 2. Unfortunately, we cannot be by no means sure if the conformations we found are the most tight ones. What makes things even worse is that for the more complex knots results of the simulations depend on the initial conformations. We would be in a much better position if the ideal conformations of at least a few knots were known rigorously. (So far we know the rigorous solution but for a single knot – the trivial one.) We hope that mathematicians will be able to make some progress in this direction. Without a rigorous understanding of the nature of the ideal conformations the accuracy of results of any numerical work will remain unknown.

Acknowledgements

I thank Arne Skjeltop for introducing me to the realm of knots. My work on the algorithm searching for the ground state of knots gained its proper intensity due to the friendly attitude of the Lausanne group: Jacques Dubochet, Andrzej Stasiak and Vsevolod Katritch. I thank LAU for the financial support.

The Pascal code of some of the procedures described in the paper was optimised due to help of Ben Laurie and Zdzisław Michalski.

I thank Robert McNeel & Associates for providing me with the beta version of the *Rhinoceros* software, with the use of which most knot images presented within the text were created.

This work was carried out under KBN Project 8T11 F010 08p04.

References

- ¹ H. K. Moffat, 1990, The energy spectrum of knots and links., *Nature* **347**, 367-369.
- ² V. Katritch, J. Bednar, D. Michoud, R. G. Sharein, J. Dubochet and A. Stasiak, 1996, Geometry and physics of knots., *Nature* **384**, 142-145.
- ³ P. Pieranski, 1996, Search of ideal knots., *ProDialog*, **5**, 111 (in Polish).
- ⁴ Y. Diao, C. Ernst, E. J. Janse Van Rensburg, this volume.
- ⁵ A. Yu. Grosberg, A. Feigel and Y. Rabin, 1996, Flory-type teory of knotted ring polymers., *Phys. Rev. E* **54**, 6618.
- ⁶ P. Pieranski, S. Clausen, G. Helgesen and A. T. Skjeltorp, 1996, Nonlinear phenomena in systems of magnetic holes., *Phys. Rev. Lett.* **77**, 1620.
- ⁷ H. K. Moffatt, 1996, Pulling the knot tight, *Nature*, **384**, 114.
- ⁸ The tests were performed in close collaboration with the Lausanne group. I am gratefully obliged to Vsevolod Katritch who, at the time I started to work on prime knots, was already aware that some of the configurations he found were not the best one and whose remarks allowed me to clean the code of the SONO algorithm from numerous bugs.
- ⁹ Private communication.
- ¹⁰ S. Przybył, P. Pieranski, 1998, Search of ideal knots. III Application of Maple V:4 to the problem of two ropes tightly twisted together, *Pro Dialog* **6**, 18 (in Polish).

CHAPTER 3

ANNEALING IDEAL KNOTS AND LINKS: METHODS AND PITFALLS.

BEN LAURIE

A.L. Digital Ltd.

*Voysey House, Barley Mow Passage,
London, W4 4GB, England*

Email: ben@algroup.co.uk

Simulated annealing is a powerful method for finding global minima. The application of annealing to the search for ideal knots and links is explored, and some optimisations and problems discussed.

1 Simulated Annealing

The method of simulated annealing is often used to search for the global minimum of multidimensional functions, and is particularly suitable when the function has many local minima. The ratio of a link's^a axial diameter to its length is clearly such a function.

I will not describe simulated annealing in depth; it is well covered elsewhere. *Numerical Recipes*¹ is a good source for this and many other algorithms.

Simulated annealing is well suited to finding ideal knots, as the energy function has many flat plateaux, and local minima, against which simulated annealing has considerable resilience.

2 General Method

At each step of the annealing process, a vertex is chosen at random. This vertex is then randomly moved using a square distribution^b based on a parameter s . The “energy” of the new configuration is calculated, and the new configuration accepted or rejected according to the usual rules of simulated annealing:

^a I will use the term “links” to mean both knots and links.

^b As I was writing this, I was referred to an article suggesting that this is not the best distribution to use.

$$P = \begin{cases} e^{-\frac{k\Delta E}{T}} & \text{if } \Delta E > 0 \\ 1 & \text{if } \Delta E \leq 0 \end{cases} \quad (1)$$

P is the probability the new configuration will be accepted, ΔE is the change in energy and k is a constant (I use $k=1$). If the new configuration is accepted, the parameter s is increased slightly, otherwise it is reduced. This ensures that around 50% of moves are accepted, which is thought to be a good thing². s is initialised to .1 of the initial maximal diameter, D .

3 Measuring Distance

An ideal link is simulated as a polygon. The co-ordinates of the vertices are all that is needed to represent the knot. In order to “idealise” the knot, it is necessary to determine the distance between each part of the knot and every other part of the knot. A little thought makes it clear that the minimum of these distances is the diameter of the maximal axial tube that can be threaded along the knot without intersecting itself. The ratio of the link’s length to this minimal distance is what we must minimise to obtain an ideal knot.

There are three different ways to measure the distance between parts of the link.

3.1 Vertex to Vertex distance

The simplest measure is simply to take distances between vertices.

Let \mathbf{a} and \mathbf{b} be the position vectors of two vertices of the link. Then the distance $d_{vv}(\mathbf{a}, \mathbf{b})$ is given by:

$$d_{vv}(\mathbf{a}, \mathbf{b}) = |\mathbf{a} - \mathbf{b}| \quad (2)$$

The disadvantage of this approach is that a link can change topology without its energy going through infinity. In fact, it can change topology without the energy even getting particularly high. Although we can disallow such changes (see Avoiding Change of Type, p. 1), it is better to not have to.

3.2 Vertex to Edge distance

Next, we can take the distance from each vertex to each edge. This is a little more complicated to calculate.

Let \mathbf{a} be the position vector of a vertex of the link, and $(\mathbf{b}_0, \mathbf{b}_1)$ be the position vectors of the ends of an edge, then:

$$\text{Let } \alpha = \frac{(\mathbf{b}_1 - \mathbf{a}) \bullet (\mathbf{b}_0 - \mathbf{b}_1)}{(\mathbf{b}_0 - \mathbf{b}_1)^2} \quad (3)$$

$$\text{Then } d_{ve}(\mathbf{a}, (\mathbf{b}_0, \mathbf{b}_1)) = \begin{cases} |\mathbf{b}_1 - \mathbf{a}| & \text{if } \alpha < 0 \\ |\mathbf{b}_0 - \mathbf{a}| & \text{if } \alpha > 1 \\ |\alpha \mathbf{b}_0 + (1 - \alpha) \mathbf{b}_1| & \text{otherwise} \end{cases} \quad (4)$$

3.3 Edge to Edge distance

The final approach is to calculate the distance between each edge and each other edge:

Let $(\mathbf{a}_0, \mathbf{a}_1)$ and $(\mathbf{b}_0, \mathbf{b}_1)$ be the position vectors of the ends of two edges of the link.

$$\text{Let } t = (\mathbf{a}_0 - \mathbf{a}_1) \bullet (\mathbf{b}_0 - \mathbf{b}_1)^2 - (\mathbf{a}_0 - \mathbf{a}_1)^2 (\mathbf{b}_0 - \mathbf{b}_1)^2 \quad (5)$$

$$\text{Let } \alpha = \left[\frac{(\mathbf{b}_0 - \mathbf{b}_1) \bullet (\mathbf{a}_0 - \mathbf{a}_1)(\mathbf{a}_1 - \mathbf{b}_1) \bullet (\mathbf{b}_0 - \mathbf{b}_1)}{- (\mathbf{a}_1 - \mathbf{b}_1) \bullet (\mathbf{a}_0 - \mathbf{a}_1)(\mathbf{b}_0 - \mathbf{b}_1)^2} \right] / t \quad (6)$$

$$\text{Let } \beta = \left[\frac{(\mathbf{a}_0 - \mathbf{a}_1)^2 (\mathbf{a}_1 - \mathbf{b}_1) \bullet (\mathbf{b}_0 - \mathbf{b}_1)}{- (\mathbf{a}_0 - \mathbf{a}_1) \bullet (\mathbf{b}_0 - \mathbf{b}_1)(\mathbf{a}_1 - \mathbf{b}_1) \bullet (\mathbf{a}_0 - \mathbf{a}_1)} \right] / t \quad (7)$$

Then $d_{ee}((\mathbf{a}_0, \mathbf{a}_1), (\mathbf{b}_0, \mathbf{b}_1)) =$

$$\begin{cases} \min(d_{ve}(\mathbf{a}_1, (\mathbf{b}_0, \mathbf{b}_1)), d_{ve}(\mathbf{b}_1, (\mathbf{a}_0, \mathbf{a}_1))) & \text{if } \alpha < 0 \text{ and } \beta < 0 \\ \min(d_{ve}(\mathbf{a}_1, (\mathbf{b}_0, \mathbf{b}_1)), d_{ve}(\mathbf{b}_0, (\mathbf{a}_0, \mathbf{a}_1))) & \text{if } \alpha < 0 \text{ and } \beta > 1 \\ \min(d_{ve}(\mathbf{a}_0, (\mathbf{b}_0, \mathbf{b}_1)), d_{ve}(\mathbf{b}_1, (\mathbf{a}_0, \mathbf{a}_1))) & \text{if } \alpha > 1 \text{ and } \beta < 0 \\ \min(d_{ve}(\mathbf{a}_0, (\mathbf{b}_0, \mathbf{b}_1)), d_{ve}(\mathbf{b}_0, (\mathbf{a}_0, \mathbf{a}_1))) & \text{if } \alpha > 1 \text{ and } \beta > 1 \\ d_{ve}(\mathbf{a}_1, (\mathbf{b}_0, \mathbf{b}_1)) & \text{if } \alpha < 0 \text{ and } 0 < \beta < 1 \\ d_{ve}(\mathbf{a}_0, (\mathbf{b}_0, \mathbf{b}_1)) & \text{if } \alpha > 1 \text{ and } 0 < \beta < 1 \\ d_{ve}(\mathbf{b}_1, (\mathbf{a}_0, \mathbf{a}_1)) & \text{if } 0 < \alpha < 1 \text{ and } \beta < 0 \\ d_{ve}(\mathbf{b}_0, (\mathbf{a}_0, \mathbf{a}_1)) & \text{if } 0 < \alpha < 1 \text{ and } \beta > 1 \\ |\alpha \mathbf{a}_0 + (1 - \alpha) \mathbf{a}_1 - \beta \mathbf{b}_0 - (1 - \beta) \mathbf{b}_1| & \text{otherwise} \end{cases} \quad (8)$$

Although this is the most complex to calculate, it is also, in some ways, the best. A link cannot change type without passing through infinite energy, and the method also yields a precisely correct answer for a polygonal link.

There is a mathematically more elegant way to do this calculation, using matrices, but it turns out to be about 30% slower³.

4 What Does “Local” Mean?

Clearly we should only measure the distance between parts of the knot that are not “next to” each other. How do we define this? One method often used is to say that two segments of a link are “next to” each other if they are on the same component and are within some fraction of the total length of the component of each other. This method is clearly unsatisfactory – we must pick the fraction for each link by some ad hoc method. Consider n trefoils composed; in this case the fraction we choose must be less than $1/n$ (substantially less, in fact), so, in the limit, the fraction must be zero. But zero obviously doesn’t work for any non-infinite link.

A more satisfactory method is to consider the ratio of the distance along the link between two knot segments and the direct distance between the same two segments⁴. If they are close to each other, this ratio will be close to 1. The further apart they are, the larger the ratio becomes. In practice, this turns out to be a very sensitive test, and is independent of knot type^c. In order to determine an appropriate ratio for the cut-off between “local” and “nonlocal”, we can consider a circle. The opposite side of the circle is clearly nonlocal. The direct distance is d (the diameter of the circle), and the distance along the knot is $\Pi d/2$. The ratio in this case is, therefore, $\Pi/2$. This gives us an upper bound for the appropriate ratio of ~ 1.55 .

A lower bound can be obtained by considering the (common) case of two parts of a link, one of which forms a spiral, and the other is the axis of that spiral. The distance from the spiral to its axis is the one we should be measuring, rather than from the spiral to itself. If we simplify the situation by assuming that the spiral is locally planar (that is, a circle), then it can be seen that the ratio should be derived from an equilateral triangle with one vertex at the centre of the circle, and the other two on the circumference. In this case, the direct distance is $d/2$ and the distance around the circle is $\Pi d/6$, giving a ratio of $\Pi/3$, or ~ 1.047 .

^c With the exception of the unknot. Although we can minimise an unknot correctly (in the sense that it takes up the correct shape), it is difficult to get the correct length to diameter ratio, which should be Π . This is because in the unknot every segment except the one precisely opposite should be considered “local”.

If we were not using approximations, the correct value to use would clearly be $\pi/2$. But if we were to use this value, the unknot would not work correctly. So we must use a somewhat lower value. In practice 1.2 or 1.3 work well.

5 Avoiding Change of Type

The simplest way to avoid a change of type is to prevent the link from “uncrossing”. A simple test for uncrossing is to check, each time we move a vertex \mathbf{v}_n to a new position, \mathbf{v}_n' , whether any edges pass through either of the triangles $(\mathbf{v}_{n-1}, \mathbf{v}_n, \mathbf{v}_n')$ and $(\mathbf{v}_n, \mathbf{v}_n', \mathbf{v}_{n+1})$, other than the edges $(\mathbf{v}_{n-2}, \mathbf{v}_{n-1})$, $(\mathbf{v}_{n-1}, \mathbf{v}_n)$, $(\mathbf{v}_n, \mathbf{v}_{n+1})$ and $(\mathbf{v}_{n+1}, \mathbf{v}_{n+2})$. These triangles correspond to the space swept by the edges $(\mathbf{v}_{n-1}, \mathbf{v}_n)$ and $(\mathbf{v}_n, \mathbf{v}_{n+1})$ as the vertex moves smoothly from \mathbf{v}_n to \mathbf{v}_n' . The test to see whether an edge $(\mathbf{v}_k, \mathbf{v}_{k+1})$ has passed through a triangle $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is as follows:

$$\text{Let } \mathbf{n} = (\mathbf{a} - \mathbf{b}) \times (\mathbf{a} - \mathbf{c}) \quad (9)$$

$$\text{Let } \alpha = \frac{\mathbf{n} \cdot (\mathbf{a} - \mathbf{v}_{k+1})}{\mathbf{n} \cdot (\mathbf{a} - \mathbf{b})} \quad (10)$$

$$\text{Let } \mathbf{p} = \alpha \mathbf{v}_k + (1 - \alpha) \mathbf{v}_{k+1} \quad (11)$$

$$\text{Let } d = a_1 b_0 - a_0 b_1 - a_1 c_0 + b_1 c_0 + a_0 c_1 - b_0 c_1 \quad (12)$$

$$\text{Let } k = -\frac{p_1 a_0 + p_0 a_1 + p_1 c_0 - a_1 c_0 - p_0 c_1 + a_0 c_1}{d} \quad (13)$$

$$\text{Let } l = \frac{p_1 a_0 - p_0 a_1 - p_1 b_0 + a_1 b_0 + p_0 b_1 - a_0 b_1}{d} \quad (14)$$

$$\text{The edge passes through the triangle if } 0 \leq k \leq 1 \ \& \ 0 \leq l \leq 1 \ \& \ k + l \leq 1 \quad (15)$$

In practice, since we always know how close the knot is to itself, we can avoid doing this test when the step size is sufficiently small. On the face of it, it would seem that a step size which is less than the minimal distance would be sufficient, but this is not so. The problem is that the knot can cross itself locally (i.e. within the area of the knot that we don't use to measure the minimal distance). This would suggest that we should always perform the test, but experimentation has shown that a cutoff at a step size 1/10 of the minimal distance works well. This is because step size is related to simulated temperature – when it is sufficiently small,

changes which may lead to crossing correspond to such large energy deltas that the chance of them being accepted is vanishingly small.

Because there is a tiny chance that an uncrossing change will be accepted when this test is not being performed, it is important to inspect the annealed knot to ensure it has not changed type^d.

6 The Distance Cache

During the annealing process, a great deal of time is spent calculating distances. But at each step only a single vertex is moved. In order to reduce the computational overhead, I implemented a cache of distances between pairs of line segments (or vertices, depending on the measure used). The obvious approach to using such a cache is to recalculate all the entries affected by each vertex move. But this is still wasteful, as, typically, each line segment is only at or near the minimal distance to a single other line segment. The other distances are usually much larger than this minimal distance. So, I augment the cache to include an “error” term, as well as a distance. Each time a vertex v_n is moved by a distance k , each cache entry with v_n as one of the pair has k added to the error term. Then, as we look for the minimal distance (i.e. the smallest entry in the cache), we take the cached distance c and subtract the error term k . Only if $c-k$ is less than the smallest distance found so far do we recalculate the actual distance. Of course, when recalculated, the new distance is cached, and the error term set to 0.

7 Results

7.1 Writhe and Average Crossing Number

One of the interesting aspects of ideal knots is that the writhe and average crossing number are related to L/d . In the case of links, the writhe is not a well-defined quantity, as it has multiple values depending on which direction you go around the link components when calculating it. The average crossing number (ACN), however, does not suffer from this problem. The writhe and ACN can be calculated as follows:

^d If the link is one we have worked on before, the L/d is a useful assurance that the type hasn't changed (we expect it to be close to previously obtained values).

For each pair of points p_{i_c, j_n} and p_{j_c, i_n} , where i_c, j_c are component numbers, and i_n, j_n are the indices of points on those components, calculate:

$$T_i = p_{i_c, j_n+1} - p_{i_c, j_n} \quad (16)$$

$$T_j = p_{j_c, i_n+1} - p_{j_c, i_n} \quad (17)$$

$$d_{ij} = (T_i \times E_{ij}) \bullet T_j \quad (18)$$

$$w_{ij} = \frac{|T_i||T_j|d_{ij}}{E_{ij}^2} \quad (19)$$

$$a_{ij} = \frac{|T_i||T_j||d_{ij}|}{E_{ij}^2} \quad (20)$$

The writhe is the sum of the w_{ij} and the ACN is the sum of the a_{ij} . (21)

Table I presents the results of annealing various 2 component links (and the trefoil). In this chart, N is the number of nodes in the polygonal approximation, and R is the ratio of the component lengths. These links are shown in plates 1 and 2.

| Link | N | L/d | ACN | R |
|----------|-----|---------|---------|---------|
| 3^1_1 | 144 | 16.3989 | 4.28005 | - |
| 2^1_2 | 160 | 12.6633 | 2.50905 | 1.01379 |
| 4^1_2 | 160 | 20.0542 | 6.02241 | 1.32176 |
| 5^1_2 | 160 | 24.9804 | 8.49914 | 1.87351 |
| 6^1_2 | 160 | 27.3297 | 9.53802 | 1.00719 |
| 6^2_2 | 160 | 28.4778 | 10.1166 | 1.0104 |
| 6^3_2 | 160 | 29.8303 | 10.7249 | 1.96305 |
| 7^1_2 | 160 | 32.2939 | 12.335 | 1.32443 |
| 7^4_2 | 160 | 32.8594 | 12.7081 | 1.72998 |
| 8^1_2 | 160 | 34.5898 | 13.4111 | 1.07362 |
| 9^1_2 | 160 | 39.6763 | 16.56 | 1.02934 |
| 10^1_2 | 160 | 41.7828 | 17.9126 | 1.01076 |
| 11^1_2 | 160 | 46.9882 | 21.0256 | 1.1147 |
| 12^1_2 | 160 | 49.3442 | 22.6926 | 1.13258 |
| 14^1_2 | 160 | 57.1055 | 26.3581 | 1.84719 |

Table I

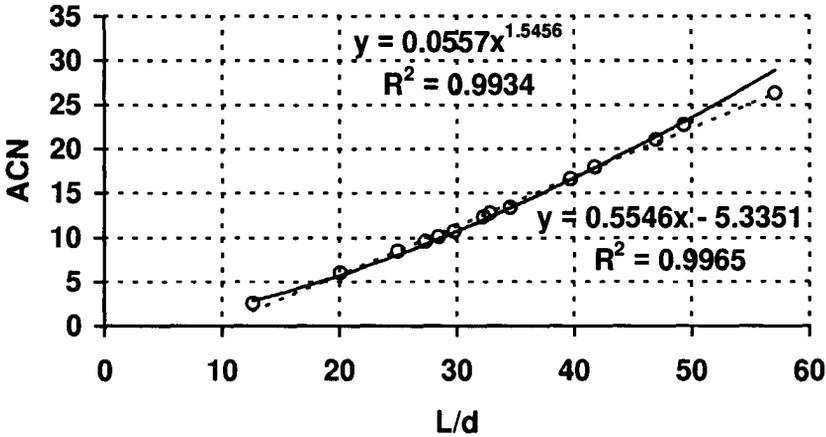


Figure 1

The ACN was originally thought to have a linear relationship to L/d , both in knots⁵ and in links^{6,7}. However, recent mathematical analysis^{8,9} suggests that this may be an artefact of the types of knots and links analysed, in that they happen to adopt roughly planar conformations when tightened. It is also possible that we were seduced by the good fit a linear relationship gives. As can be seen in figure 1, a distinctly non-linear curve is nearly as good a fit as the linear one, though the exponent (1.5456) is rather higher than the expected maximum of $4/3$. Also, a preliminary study of Hopf tori (Hopf tori are a bundle of n tori simply linked with another bundle of n tori) exhibits a power law with exponent ~ 1.44 ($R^2 = .999$).

7.2 Smoothness

Although the resulting links look smooth, detailed examination shows that they are not. Using visualisation tools to show the curvature, we can see that it varies unsmoothly. More interestingly, if we study the relationship between the individual segments, we can see that the points of closest approach between the segments behave in a very complex way. They do not progress in a simple way from one segment to the next. Instead, each segment is influenced by its immediate neighbours, and their relationship to other segments. Some segments do not even achieve the minimal distance, and it seems they cannot. There is an extraordinarily

complex mesh of interrelationships that define the final shape of the ideal link. An interesting question is whether this is purely a result of approximating the link with a polygon, or whether these complex relationships persist even for completely smooth curves.

8 Limitations of the Method

Experimentation with this method has revealed several limitations.

8.1 Performance

Even with the caching scheme described above, a 160 vertex link takes around 14 hours to anneal on an SGI O2 workstation. For good results, I'd like to be using at least 10 times as many vertices, but since the time taken seems to be $O(n^3)$ this is clearly out of the question without some radical improvement.

8.2 Migration of Components

Of course, the whole point of using simulated annealing is to find global minima. A good test of whether this is, in fact, happening, is to see whether the loops on a trefoil with two loops linked to it can move from one "lobe" of the trefoil to another. Using simulated annealing, they have never been observed to do so. This can be understood easily: when the temperature is sufficiently high to permit such a movement, the link is so tangled it is virtually impossible for the loops to squeeze through the gaps in the trefoil, and at lower temperatures moving a loop through the centre of a lobe is more or less prohibited, since the energy is guaranteed to at least double as a consequence.

9 Future Directions

Szu and Hartley¹⁰ claim that the speed of annealing can be hugely improved by choosing the distribution of the perturbations carefully. Investigation of this procedure is planned.

Pieranski¹¹ uses a method which simulates the physical tightening of string. The steps used deterministically in Pieranski's method could be applied as perturbations for annealing.

10 Availability of Code

C++ source code implementing the method described here is available on application to me by email. The source includes a general purpose class for representing links, software to visualise links on SGI platforms (or others that support OpenInventor¹²) and software to produce raytraced images using POVray¹³. It also introduces a file format for the platform-independent interchange of knot and link coordinates, and associated information, known as the Portable Knot Format (PKF). The only conditions imposed on use of the code are:

- The authors should be acknowledged in any derived works, or descriptions of results obtained using the code.
- Improvements should be sent to the authors for possible inclusion in future versions.

11 References

1. Press, Teukolsky, Vetterling and Flannery, 1988, *Numerical Recipes in C* (Cambridge: Cambridge University Press).
2. Vsevolod Katrich, 1996, *personal communication*.
3. Myk Soar, 1997, *personal communication*.
4. Camilla von Massenbach, 1997, *personal communication*.
5. Vsevolod Katrich et al., 1996, *Geometry and Physics of Knots*, Nature 384, 142-5.
6. Ben Laurie et al., 1998, *Geometry and Physics of Catenanes Applied to Study DNA Replication*, Biophysical Journal 74, No. 6.
7. Alexander Vologodskii et al., 1998, *Sedimentation and Electrophoretic Migration of DNA Knots and Catenanes*, Journal of Molecular Biology 278, 1-3.
8. Jason Cantarella et al., *Tight knot values deviate from linear relations*, Nature 392, 237-8.
9. Gregory Buck, *Four-thirds power law for knots and links*, Nature 392, 238-9.
10. Szu and Hartley, 1987, *Fast Simulated Annealing*, Physics Letters A 122, 157-162.
11. Piotr Pieranski, 1998, *In Search of Ideal Knots*, this book.
12. <http://www.sgi.com/Technology/Inventor.html>
13. <http://www.povray.org/>

CHAPTER 4

KNOTS WITH MINIMAL ENERGIES

Y. DIAO[†], C. ERNST[‡] and E.J. JANSE VAN RENSBURG[‡]

[†]*Department of Mathematics*
University of North Carolina at Charlotte
Charlotte, NC 28223, USA

[‡]*Department of Mathematics*
Western Kentucky University
Bowling Green, KY 42101, USA

[‡]*Department of Mathematics and Statistics*
York University
North York, Ontario, M3J 1P3, Canada

ABSTRACT

Energy functions on knots are continuous and scale-invariant functions defined from knot conformations into non-negative real numbers. The infimum of an energy function is a knot invariant which defines (not necessarily unique) “canonical conformations” of knots in three space. In this paper we examine strategies of how to find these canonical conformations and compute or measure their minimal energy. Furthermore, we discuss properties that energy functions should have if one wants to compute canonical conformations of knots with minimal energy. Two types of energies are discussed in more detail, the first type consists of energies of C^2 or C^1 knots defined using the concept of thickness of a knot; the second type is a polygonal energy which should be well suited to numerical computation since it has all the properties discussed earlier.

Keywords: Knots, Polygonal Knots, Energy of Knots, Properties of Energy Functions, Thickness of Knots

1. INTRODUCTION

The energies of knots were introduced by O’Hara [1] in an attempt to study both the complexity and canonical forms of knots. Energy functions are scale invariant functions defined from knot- conformations into non-negative

real numbers. The energy of a given knot type \mathcal{K} is the infimum of the knot energies taken over the set of all conformations of knot type \mathcal{K} . Ideally, the energy infimum is attained at a tame conformation which provides us with a *canonical form* of the knot type. However, different energy functions can yield different canonical forms of the same knot type. Intuitively, the connection between the complexity of a knot conformation and its energy is simple - the more complicated the conformation the higher the energy.

A large number of energy functions are known and discussed in the literature [2,3,4,5]. At a first glance it seems that one can define new energy functions at will, and therefore obtain many “ideal shapes” of the same knot. In reality, we have found that it is surprisingly difficult to define “good” knot energies, i.e. knot energies for which one can establish properties designed to ensure that the canonical conformations with minimal energies reflect the complexity of the knot in an intuitive manner. An example of such a property in the case of polygonal knots is the following: If one increases the number of edges of a polygonal knot, then the angles between consecutive edges must become “uniformly small” in a polygonal configuration with minimal energy. Clearly such a property is necessary if one hopes that the polygon should approximate the canonical form of a smooth knot as the number of edges increases. Many knot energies in the literature do not have these properties and therefore are not good at revealing a nice canonical conformation of a knot [7,10].

Other properties that should help finding canonical forms of a knot are the following: The energy function should have a global infimum on the unit circle (in the case of polygonal knots, on the regular polygon). In addition, the knot energy should diverge when it is computed from any sequence of knots which approach a singular conformation. This will avoid the event that a minimal energy conformation is singular, and it will present an energy barrier between knots of different types. The last property is standard in the numerical approximation of minimal energy conformations. Many well-known energy functions fail the above tests, see again [7, 10]. In the next section, we will give precise definitions of these properties with the concept of *ideal* and *semi-ideal* knot energies.

In this paper we examine strategies of how to find canonical conformations of knots and links and how to compute or measure their minimal energy. Many of the results quoted in this paper have appeared elsewhere, and the proofs have been omitted.

2. DEFINITIONS AND PROPERTIES OF ENERGY FUNCTIONS

One can imagine an energy function on knots to be a scale-invariant function from the space of knot conformations into non-negative real numbers. It seems desirable that energy functions are continuous, but we will see that some functions derived from the thickness of a knot are in fact not continuous. To probe the continuity of knot energies, we define a metric on the space of all knot conformations of unit length (since energy functions are scale invariant, we need to only consider these).

Definition 2.1. Let $\alpha(s)$ and $\beta(s)$ be two simple closed curves of the same length parametrized by arc length. If the curves are k -times differentiable, then the k -distance between them is defined as

$$d^k(\alpha(s), \beta(s)) = \min_{A \in O} \left\{ \max_{0 \leq s \leq L, 0 \leq t \leq L} \sum_{i=0}^k |A\alpha^{(i)}(s+t) - \beta^{(i)}(s)| \right\}, \quad (1)$$

where L is the length of $\alpha(s)$ and $\beta(s)$, A is a rigid motion (the composition of a translation and a rotation), O is the group of all rigid motions in \mathcal{R}^3 and $\alpha^{(i)}(s)$, $\beta^{(i)}(s)$ are the i -th derivatives of $\alpha(s)$ and $\beta(s)$.

Using the above metric, we can define continuous functions on the space of all knot conformations of unit length:

Definition 2.2. Let $C^k(T)$ be the set of all simple closed curves of unit length that are k times differentiable. Then $f : C^k(T) \rightarrow \mathcal{R}$ is said to be k -continuous on $C^k(T)$ if $\forall \epsilon > 0, \exists \delta > 0$, such that $|f(\alpha) - f(\beta)| < \epsilon$ whenever the k -th distance between α and β is less than δ .

Remark: If $f : C^k(T) \rightarrow \mathcal{R}$ is k -continuous, then f is invariant under rigid motion, i.e., $f(A\alpha) = f(\alpha)$ for any $\alpha \in C^k(T)$ and $A \in O$. Also, f will be m -continuous on $C^m(T)$ if $m \geq k$ since $C^m(T) \subset C^k(T)$. On the other hand, one can find examples in which f is k -continuous but not m -continuous for some $m < k$.

We express scale invariance in definition 2.3:

Definition 2.3. Let $C^k(T')$ be the set of all simple closed curves (of any length) that are k times differentiable. Let $f : C^k(T') \rightarrow \mathcal{R}$. Then f is said to be scale invariant if we have $f(\lambda\alpha(s)) = f(\alpha(s))$ for any real number λ and $\alpha(s) \in C^k(T)$.

Definition 2.4. Let $f : C^k(T') \rightarrow \mathcal{R}^+$. Then f is called an energy function on $C^k(T')$ if it is scale invariant and f is continuous when restricted to $C^k(T)$.

Some desirable properties of an energy function are given in the following definition. Let f be an energy function on $C^k(T')$. Then for a given knot type \mathcal{K} , the energy of \mathcal{K} denoted $f(\mathcal{K})$ is defined to be the infimum of f over all knots K of knot type \mathcal{K} in $C^k(T')$.

Definition 2.5. Let $f : C^k(T') \rightarrow \mathcal{R}^+$, where $k > 0$, be an energy function. Then

- (a) f is called a *basic* energy function if $f(\alpha)$ is an absolute minimum if and only if α is a circle;
- (b) f is called a *strong* energy function if for any given positive number a , there are only finitely many knot types \mathcal{K} such that $f(\mathcal{K}) \leq a$;
- (c) f is called a *charge* energy if f goes to infinity when α approaches a closed curve with self-intersections;
- (d) f is called a *tight* energy if $f(\alpha_i)$ goes to infinity for any sequence of embeddings α_i of equal length and of the same knot type K , where α_i contains a knotted arc of knot type K in a ball B_i and the diameter of the ball B_i converges to zero.

We will then call f a *semi-ideal* energy function if it satisfies conditions (a) to (c) above and an *ideal* energy function if it satisfies conditions (a) to (d) above.

Condition (b) is desirable since it ensures that only a finite number of knot types have energies below any given threshold, thus a strong energy can serve as a measure of the complexity of a knot. Conditions (c) and (d) are designed to ensure nice canonical conformations. Note that there are knot energies which are infinite at singular knot embeddings, but not tight (see [7] for examples). If one has a knot energy function which is not tight and one wants to find a canonical form of a knot by minimizing the knot energy using numerical simulations with a large number of edges, then it may happen that the knot energy is minimized at a tiny knotted ball pair on an almost regular polygon. In such a case all knotted curves (any knot type !) might approach nearly identical conformations (namely the unit circle).

A family of (almost) ideal energy functions can be defined as the reciprocal of the thickness of a knotted curve. In the following two sections we will discuss a variety of different approaches and definitions of thickness of a space curve.

3. ENERGIES DEFINED BY KNOT THICKNESSES

The disk thickness. Let K be a C^2 knot of length L . A number $r > 0$ is said to be “nice” if for any x, y on K , we have $D(x, r) \cap D(y, r) = \emptyset$, where $D(x, r)$ and $D(y, r)$ are the discs of radius r centered at x and y which

are normal to K . The *disk thickness* of K is defined to be the supremum over all the nice r 's and will be denoted by $t_D(K)$. It follows that for any $0 < r < t_D(K)$, $K(r)$ has a natural circle bundle structure $D^2 \times S^1$. Since each $D(x, r)$ is disjoint from $D(y, r)$ (for any $x \neq y$), $K(r)$ can be retracted onto K by retracting each $D(x, r)$ to x . Let $D(K)$ be the smallest distance between points x and y on K , where the tangents at x and y are perpendicular to the line segment connecting x with y . $D(K)$ is called the double critical self distance of K . The disk thickness has been well characterized in reference [6] where the following theorem is shown:

Theorem 3.1. $t_D(K) = \min\{\frac{1}{2}D(K), \frac{1}{\kappa}\}$ where κ is the maximum curvature of K .

The advantage of $t_D(K)$ over other thicknesses is that it is 2-continuous (see [8]) hence $L/t_D(K)$ would define an ideal energy function (see theorem 3.2). But since $t_D(K)$ depends on the curvature of K , it becomes undefined if K is only differentiable or is piecewise linear. Further, since it is restricted by the reciprocal of the maximum curvature of K , it has a tendency to lose the big picture of the knot if the knot contains a very short arc with large curvature. The two definitions after theorem 3.2 are attempts to define a thickness of a curve which does not depend on curvature and can be used for C^1 curves.

Theorem 3.2. Let K be a C^2 knot of length L , then $\frac{L}{t_D(K)}$ is an ideal energy function. \square

A proof for theorem 3.2 can be found in reference [7].

The maximal thickness. The c -neighborhood of a simple closed C^1 -curve K is the set of all points with distances less than or equal to c from K . We denote the c -neighborhood of K by K_c . Also, for any $x \in K$, let Σ_x denote the plane that intersects K perpendicularly at x . If c is small enough, then for any $x \in K$, the component of $K_c \cap \Sigma_x$ containing x is a meridian disk of K_c which only intersects K_c at x . For small c one can show that K_c is homotopic to K via strong deformation retract, see [8] for a detailed proof. A number c with this property, that K_c is homotopic to K via strong deformation retract, is called a *nice* number and the *thickness* of K is defined by $t(K) = \sup\{c \mid t \text{ is nice } \forall t \in (0, c)\}$. The advantage of this definition is that it captures the intuitive meaning of thickness; the neighbourhood of the curve can be made thicker until it becomes constricted by itself. It is the maximal possible thickness since any larger neighborhood of K would contain a neighborhood of K that has a different homotopy type from K . Unfortunately the maximal thickness is not continuous in a C^1 metric and thus $L/t(K)$ is not an ideal energy. An additional disadvantage of $t(K)$ is that it is difficult

to calculate or simulate on a computer since it is not analytically defined and no formulas such as in theorem 3.1 are known. The following third thickness is an attempt to define a C^1 thickness analytically, so that it can be calculated easier.

The ϵ -thickness. Let K be a closed curve, and let x and y be two points on K , separated by an arc-length distance $s(x, y)$ along the curve. Let the Euclidean distance between x and y be $d(x, y)$. Define T_x to be the tangent vector to K at x , and let $\theta(x, y)$ be the smaller angle between T_x and the line-segment from x to y . Let K be a C^1 curve and $2\pi/3 \geq \epsilon \geq 0$, then the ϵ -thickness of K is defined as

$$t_\epsilon(K) = \inf_{x, y \in K} \left\{ \frac{d(x, y)}{2 \sin \theta(x, y)} : \frac{2 \sin \theta(x, y) \cdot s(x, y)}{d(x, y)} \geq \epsilon \right\}. \quad (2)$$

If K is a C^2 curve and x and y are points on K with small arc-length distance $s(x, y)$, then we can approximate O , the center of curvature at x , by O' , where O' is the point on the intersection line of the plane that bisects the line segment from x to y and Σ_x that is closest to the middle point between x and y . If R is the distance between x and O and R' is the distance between x and O' , then $1/R'$ is an approximation to $1/R$, the curvature of K at x . For C^2 curves one can show that if $y \rightarrow x$ then $1/R' \rightarrow 1/R$. Solving for R' we get $R' = \frac{d(x, y)}{2 \sin \theta(x, y)}$, [8]. Thus the quotient in (2) has a geometric meaning similar to the radius of curvature. The inequality in (2) guarantees that $t_\epsilon(K)$ is not affected by short arcs contained in K (of course the length of such arcs are related to the value of ϵ). Thus one hopes that the ϵ -thickness is better at capturing the big picture of the knot than the disk thickness. Since tangents and normal planes are defined for a C^1 curve as well, this approximation of the center of curvature can be computed for every x and y on K . One can verify that $t_\epsilon(K)$ is a non-decreasing function of ϵ for each fixed K . Furthermore, it can be shown ([8]) that $t_\epsilon(K) \leq t(K)$ and therefore it is indeed a thickness of K , i.e. K_c is homotopic to K via strong deformation retract for all $c < t_\epsilon(K)$. Also $t_0(K) = t_D(K)$ if K is a C^2 curve and thus the ϵ -thickness generalizes the disk thickness. Unfortunately, $t_\epsilon(K)$ fails to be a 1-continuous function. See [8] for details. Thus, the function defined by $L/t_\epsilon(K)$ is not quite an energy function.

We define the thickness of a knot type \mathcal{K} to be the infimum

$$t(\mathcal{K}) = \inf_{K \in \mathcal{K}} t(K), \quad (3)$$

and similarly for $t_D(\mathcal{K})$ and $t_\epsilon(\mathcal{K})$. An open question is: For a given knot type \mathcal{K} , what is the relation between $t(\mathcal{K})$, $t_D(\mathcal{K})$ and $t_\epsilon(\mathcal{K})$? We conjecture the following:

Conjecture. Let \mathcal{K} be any knot type, then there exists a knot $K \in \mathcal{K}$ such that $t(K) = t(\mathcal{K}) = t_D(\mathcal{K}) = t_\epsilon(\mathcal{K})$.

Currently this is only a conjecture. However the conjecture is consistent with numerical simulations. We approximated the thickness of the trefoil knot by a computer simulation using a Metropolis Monte Carlo simulation with simulated annealing (see references [8, 10, 14] for details). The definition of $t_\epsilon(K)$ was adapted to equilateral polygonal knots (this introduces some error in the approximation, but this error is not significant if the polygonal knot has a large number of edges). In figure 1 we illustrate the thickest trefoil knot with 128 edges and $\epsilon = 0.5$. We found that $t_\epsilon \approx 0.0305$ in this case. Simulations with other values of ϵ (we took ϵ to be 1.0, 1.5, 2.0 and 2.09) all gave a thickness of 0.0302. Simulations with trefoils with 64 edges and 32 edges were also performed, in both cases we found that the maximum thickness is insensitive to the value of ϵ (in the case of 32 edges we obtained $t_\epsilon \approx 0.0316$ and for 64 edges we obtained $t_\epsilon \approx 0.0305$). The consistency of our results indicates that $t_\epsilon(3_1) \approx 0.03$, where 3_1 is the trefoil knot supporting our conjecture.

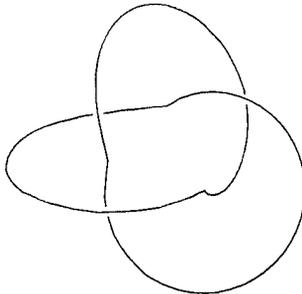


Figure 1: A *polygonal trefoil* with 128 edges and t_ϵ - thickness 0.0305, with $\epsilon = 0.5$. We produced this *polygonal knot* by simulated annealing.

4. OPEN THICKNESS ENERGY

It is extremely difficult to compute the thicknesses of the last section for different knot types. The approach of this section is to design a thickness that can be measured by a physical experiment using ropes. This thickness will be defined for knotted or linked arcs instead of closed curves and can be called the *open thickness* of a knot (or link). First we will define this thickness for a

knotted arc and then we will give a definition of thickness of two knotted or linked arcs, which will form a link when properly closed up. It will be shown that the energy function derived from these definitions has all the properties of an ideal energy function. Most of the results involving knotted arcs in this section can be found in [9]. All the work is done in Euclidean 3-space where each point has Cartesian coordinates (x, y, z) .

Open Thickness of Knots. Let \mathcal{K} be a knot type and $A_{\mathcal{K}}$ be a C^2 curve with arc length parametrization $A_{\mathcal{K}}(t)$, $t \in [0, L]$, and L is the (finite) length of $A_{\mathcal{K}}$. Moreover, assume $A_{\mathcal{K}}$ to have the following properties:

- (1) The endpoints of $A_{\mathcal{K}}$ are $A_{\mathcal{K}}(0) = (0, 0, 0)$ and $A_{\mathcal{K}}(L) = (0, 0, L_{\mathcal{K}})$.
- (2) $A_{\mathcal{K}} \setminus \{A_{\mathcal{K}}(0), A_{\mathcal{K}}(L)\}$ is strictly bounded by the planes $z = 0$ and $z = L_{\mathcal{K}}$.
- (3) The tangents $\frac{d}{dt}A_{\mathcal{K}}(t)$ is parallel to the z -axis for $t = 0$ and $t = L$.
- (4) $(A_{\mathcal{K}}, B)$ is a knotted arc-ball-pair of knot type \mathcal{K} , where B is the solid cylinder with height $L_{\mathcal{K}}$, radius L and axis defined by the endpoints $A_{\mathcal{K}}(0)$ and $A_{\mathcal{K}}(L)$. Note that, except for its endpoints, $A_{\mathcal{K}}$ is disjoint with the boundary of B .

As in the case of the disk thickness of a closed curve, a number $r > 0$ is said to be “nice” if for any x, y on $A_{\mathcal{K}}$, we have $D(x, r) \cap D(y, r) = \emptyset$, where $D(x, r)$ and $D(y, r)$ are the discs of radius r centered at x and y which are normal to $A_{\mathcal{K}}$. The *open thickness* of $A_{\mathcal{K}}$ is defined to be the supremum over all the nice r 's and will be denoted by $t_o(A_{\mathcal{K}})$.

Definition 4.1. The *open thickness energy* of the curve $A_{\mathcal{K}}$ is given by $E^o(A_{\mathcal{K}}) = (L - L_{\mathcal{K}})/t_o(A_{\mathcal{K}})$. The open thickness energy of the knot type \mathcal{K} is defined by $E^o(\mathcal{K}) = \inf\{E^o(A_{\mathcal{K}})\}$.

Theorem 4.2. $E^o(A_{\mathcal{K}})$ is an ideal energy function of knots.

The proof that $E^o(A_{\mathcal{K}})$ is ideal will be omitted, because the proof that $E^o(A_{\mathcal{K}})$ has all the properties of an ideal energy function (except continuity) can be found in [9]. In addition the arguments are similar to the ones used in the proof of Theorem 4.4 in this section. That $E^o(A_{\mathcal{K}})$ is continuous follows from theorem 3.2.

Open Thickness of Links. We can define the open thickness of a 2-component link in a similar manner and prove that an energy derived from this thickness is also ideal.

Let \mathcal{L} be an unsplittable link type and $A_{\mathcal{L}}$ and $B_{\mathcal{L}}$ be C^2 curves with arc length parametrization $A_{\mathcal{L}}(t)$ and $B_{\mathcal{L}}(t)$. L_A is the (finite) length of $A_{\mathcal{L}}$, L_B is the (finite) length of $B_{\mathcal{L}}$ and the total length $L = L_A + L_B$. Moreover, assume $A_{\mathcal{L}}$ and $B_{\mathcal{L}}(t)$ to have the following properties:

(1) The endpoints of $A_{\mathcal{L}}$ are $A_{\mathcal{L}}(0) = (0, 0, 0)$ and $A_{\mathcal{L}}(L_A) = (\epsilon, 0, 0)$ for some real number ϵ . The endpoints of $B_{\mathcal{L}}$ are $B_{\mathcal{L}}(0) = (x_0, y_0, -L_{\mathcal{L}})$ and $B_{\mathcal{L}}(L) = (x_1, y_1, -L_{\mathcal{L}})$, where x_i and y_i are real numbers and $L_{\mathcal{L}}$ is a positive parameter that is less or equal to $L/2$.

(2) $A_{\mathcal{L}} \cup B_{\mathcal{L}} \setminus \{A_{\mathcal{L}}(0), A_{\mathcal{L}}(L), B_{\mathcal{L}}(0), B_{\mathcal{L}}(L)\}$ are strictly bounded by the planes $z = 0$ and $z = L_{\mathcal{L}}$.

(3) The tangents $\frac{d}{dt}A_{\mathcal{L}}(t)$ and $\frac{d}{dt}B_{\mathcal{L}}(t)$ are parallel to the z -axis for $t = 0$ and $t = L_A$ or $t = L_B$ respectively.

(4) Let C be a solid cylinder with height $L_{\mathcal{L}}$, radius L and axis with endpoints $(0, 0, 0)$ and $(0, 0, L_{\mathcal{L}})$. Note that, except for their endpoints, $A_{\mathcal{L}}$ and $B_{\mathcal{L}}$ are disjoint with the boundary of C . So $(C, \{L_A, L_B\})$ is a two string tangle. We require that the numerator of this tangle results in the link \mathcal{L} .

As in the case of the disk thickness of a closed curve, a number $r > 0$ is said to be “nice” if for any x, y on $A_{\mathcal{L}} \cup B_{\mathcal{L}}$, we have $D(x, r) \cap D(y, r) = \emptyset$, where $D(x, r)$ and $D(y, r)$ are the discs of radius r centered at x and y which are normal to $A_{\mathcal{L}} \cup B_{\mathcal{L}}$. The *open thickness* of $A_{\mathcal{L}} \cup B_{\mathcal{L}}$ is defined to be the supremum over all the nice r 's and will be denoted by $t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$.

Definition 4.3. The *open thickness energy* of the arcs $A_{\mathcal{L}}$ and $B_{\mathcal{L}}$ is given by $E^o(A_{\mathcal{L}} \cup B_{\mathcal{L}}) = (L/2 - L_{\mathcal{L}})/t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$. The open thickness energy of the link type \mathcal{L} is defined by $E^o(\mathcal{L}) = \inf\{E^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})\}$.

We will show that the open thickness link energy is charge, tight and strong in the same sense that the open thickness energy of knots are.

Theorem 4.4. $E^o(\mathcal{L})$ is a charge, tight and strong energy function of links.

Proof. $E^o(\mathcal{L})$ depends on the difference $L/2 - L_{\mathcal{L}}$ and on L itself. Appending line segments parallel to the z -axis at endpoints of $A_{\mathcal{L}}$ and $B_{\mathcal{L}}$ outwards normal to the B does not change $E^o(\mathcal{L})$. Uniform dilations of \mathcal{R}^3 do not change $E^o(\mathcal{L})$, since this amounts to scale changes in both $L/2 - L_{\mathcal{L}}$ and $t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$, so the ratio is invariant. Hence $E^o(\mathcal{L})$ is scale invariant. It is also easy to see that $E^o(\mathcal{L})$ is a charge and tight energy function. Continuity of $E^o(\mathcal{L})$ follows from the continuity of the disk thickness, see [8]. So we will need to show that it is strong.

Assume that for $A_{\mathcal{L}} \cup B_{\mathcal{L}}$, we have $E^o(A_{\mathcal{L}} \cup B_{\mathcal{L}}) = (L/2 - L_{\mathcal{L}})/t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}}) \leq a$ for some $a > 0$. Project $A_{\mathcal{L}} \cup B_{\mathcal{L}}$ onto the z -axis by the map $p : (x, y, z) \rightarrow (z)$. p maps $A_{\mathcal{L}} \cup B_{\mathcal{L}}$ onto the interval $[0, L_k]$. Let

$$P_2 = \{(x, y, z) \in A_{\mathcal{L}} \cup B_{\mathcal{L}} | p^{-1}(z) \cap (A_{\mathcal{L}} \cup B_{\mathcal{L}}) \text{ are 2 single points}\}$$

and let

$$P_{>2} = \{(x, y, z) \in A_{\mathcal{L}} \cup B_{\mathcal{L}} | p^{-1}(z) \cap (A_{\mathcal{L}} \cup B_{\mathcal{L}}) \text{ are more than 2 single points}\}.$$

Notice that P_2 is an open subset of $A_{\mathcal{L}} \cup B_{\mathcal{L}}$. A component of $P_{>2}$ consists of all points of $P_{>2}$ whose image under the projection map p is a connected subset of the z -axis. $P_{>2}$ may contain more than one component. We prove that if the projection of such a component has length less than $6t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$, then that component is unknotted and unlinked in the following sense. Intersecting $(A_{\mathcal{L}} \cup B_{\mathcal{L}}, C)$ with two planes parallel to the xy -plane slightly below and above the said component creates a 2-string tangle. Saying that a component of P_2 is unknotted or unlinked means that the tangle is equivalent to the 0-tangle or ∞ -tangle.

Let T be a component of $P_{>2}$ and assume that T forms a nontrivial tangle with the bounding planes. A projection of T into the xy -plane has at least one point of self intersection resulting in a non trivial crossing of T . The two open balls of radius $t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$ centered on different points on the curves $A_{\mathcal{L}} \cup B_{\mathcal{L}}$ which projects to a such a self intersection cannot intersect since $A_{\mathcal{L}} \cup B_{\mathcal{L}}$ has thickness $t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$. So we see that the projected length of T on the z -axis greater or equal to $2t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$ and it follows that the length of the arcs in T is greater or equal to $6t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$.

In particular if the component T is not trivial, then $(L/2 - L_{\mathcal{L}}) \geq t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$, for any $A_{\mathcal{L}} \cup B_{\mathcal{L}}$, and $E^o(\mathcal{L}) \geq 1$. Since $E^o(A_{\mathcal{L}} \cup B_{\mathcal{L}}) \leq a$, we note there are at most a nontrivial components in $P_{>2}$. Let these nontrivial components be C_i , $i = 1, 2, \dots, m$, where $m \leq a$. Let $L(i)$ be the total length of C_i and let the separation between the maximum and minimum z -component of C_i be $L_{\mathcal{L}}(i)$. Then

$$\frac{L(i)/2 - L_{\mathcal{L}}(i)}{t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})} \leq a, \forall i.$$

But we see that $L(i) \geq 3L_{\mathcal{L}}(i)$ since C_i is a nontrivial component of $P_{>2}$, so $L(i) \leq 2(L_{\mathcal{L}}(i) + at^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})) \leq 2(L(i)/3 + at^o(A_{\mathcal{L}} \cup B_{\mathcal{L}}))$. Thus $L(i) \leq 6at^o(A_{\mathcal{L}} \cup B_{\mathcal{L}})$. If we now show that only finitely many tangle types can be represented by C_i for each i , then we are done, since then $A_{\mathcal{L}} \cup B_{\mathcal{L}}$ can be knotted in only a finite number of link types if its open thickness energy is bounded by a . Let $n = \lfloor L(i)/t^o(A_{\mathcal{L}} \cup B_{\mathcal{L}}) \rfloor + 1 \leq 6a + 1$ and divide C_i into n pieces of equal arc-length. Connect the vertices in this division by line segments, to obtain a piecewise linear approximation to C_i which is isotopic in the $t^o(\mathcal{L})$ -neighborhood of C_i to the ball-pair defined by C_i and its bounding planes. The n line-segments can only have at most n^2 crossings in any projection, and so C_i can have only a finite number of knot types. \square

The advantage of an open thickness of a knot (over its thickness) is that the open thickness can be measured in a physical experiment. We first describe

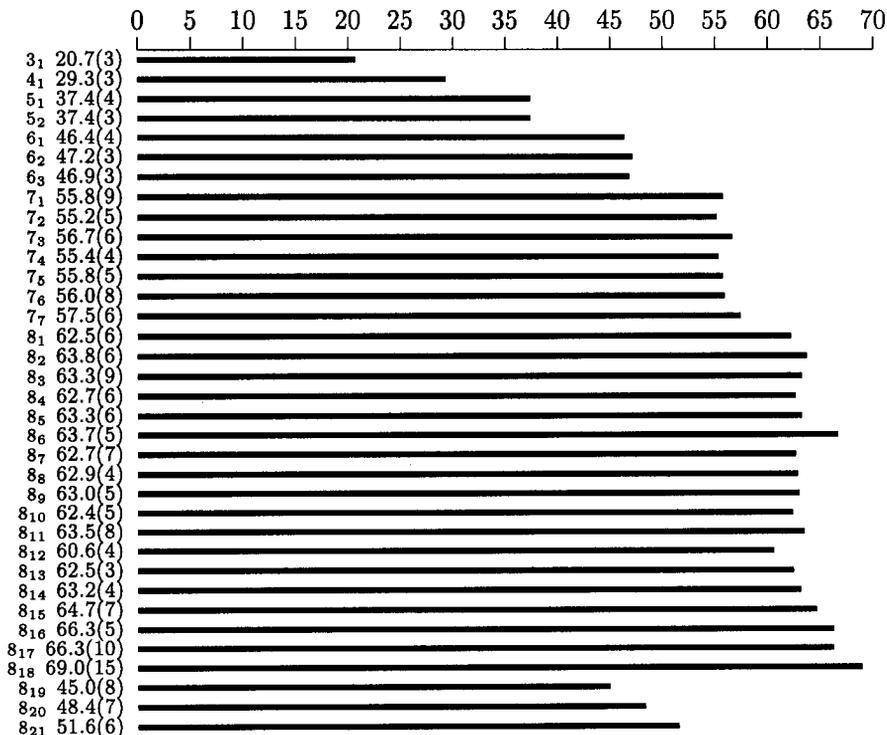


Figure 2: *The open thickness energies of knots to eight crossings. In this bar graph the length of each bar indicates the length of rope used when a given knot is tied, in units of the radius of the rope. We indicate a statistical error bar of one standard deviation in brackets: for example, 20.7(3) is to be interpreted as 20.7 ± 0.3 .*

the experiment in detail for the knotted arcs and then outline the experiment of the linked arcs.

Measuring the Open Thickness of Knots and Links. Take a rope of a known thickness and length and tie a knot of a given type in it. The presence of the knot lessens the end-to-end length of the rope, and by measuring this reduction in length, we can measure an upper bound on the open thickness energy of the knot. If the knot is tied in an optimal way (to minimize the amount of rope used), and if it is pulled tight enough, then it seems not unreasonable to assume that the measurement will approximate the actual open thickness energy of the knot.

We obtained four different mountaineering ropes in two different locations (two per location). The ropes were labeled by the manufacturer as having diameters 3 mm, 5.5 mm, 7 mm and 9.0 mm, but we measured the diameters with a caliper to be 2.46 mm, 4.9 mm, 7.02 mm and 8.0 mm respectively when the ropes were stretched during the experiment by a fixed force to pull the knot tight. One end of the rope was tied to a fixed object with the other hanging freely. The free end of the rope was then connected to a heavy weight (a bucket of sand proved sufficient in one location, while a couple of heavy metallic weights were used at the other). We lubricated the rope with Vaseline and tied each knot five times in each rope to generate statistical data on the open thickness.

There were small systematic differences between the results from different ropes, and we attribute these to the different qualities of the ropes: Knots were more difficult to pull tight in the thicker ropes, while thinner ropes may stretch more under the applied weights. We pooled all the data and computed the average open thicknesses of each knots type. We measured all knots up to 7-crossings by four ropes, while the 8-crossings knots were measured on 2 of the 4 ropes (the 2.46 mm and 7.02 mm ropes), with exception of 8_1 , 8_{19} and 8_{20} , which were measured also by 4.9 mm rope. The data was converted to the open thickness energy.

The converted energies are listed in figure 2. Note that the units of this energy is rope radius, for example, the energy of the trefoil knot is 20.7 ± 0.3 , which means that the length of rope needed to capture a trefoil is about 20.7 times its radius. The number in brackets following each energy is a statistical error bar composed of one standard deviation (computed from the pooled data). The knots 8_{15} to 8_{18} tend to be hard to measure, and 8_{19} was remarkably hard to untie once it was pulled tight.

The general trend in figure 2 is that energies increase with crossing number. In addition, with the exception of the non-alternating knots 8_{19} , 8_{20} and 8_{21} , all the other energies are remarkably well separated by their crossing numbers. Note that this bar graph can be used to identify knots from a complicated diagram: Tie the knot in a rope, and pull it tight. Then measure the amount of rope used in tying the knot; dividing this by the rope radius should give a good where to start searching for the knot in the standard tables.

In a numerical study, Katritch *et al* [11] estimated the thickness of *closed* knotted curves. At a qualitative level, their results are in good agreement with ours; they found a general decrease in the thickness of knotted conformations with increasing crossing number, while the non-alternating knots have anomalously high thickness. It is not clear how to make a quantitative com-

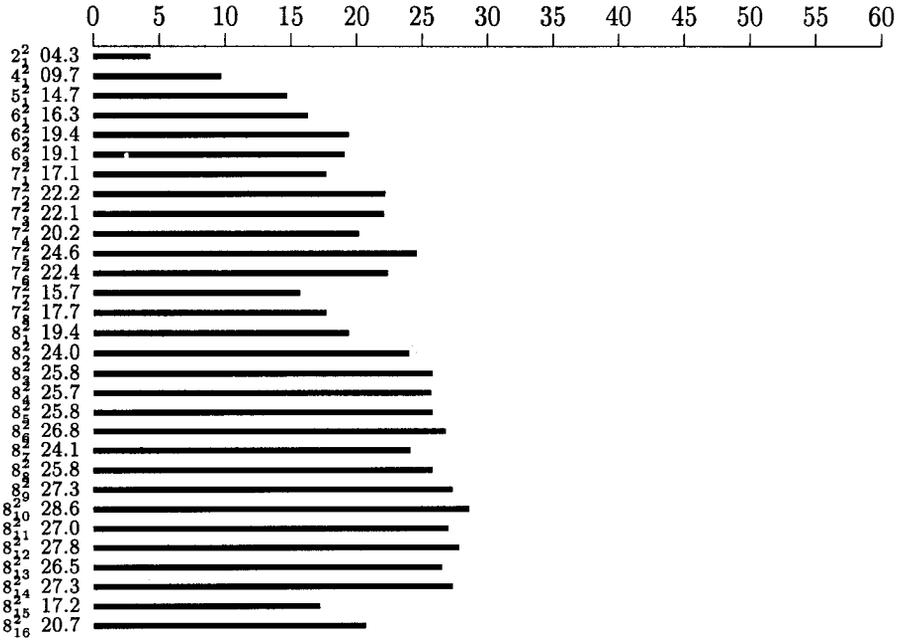


Figure 3: *The open thickness energies of links to eight crossings. In this bar graph the length of each bar indicates the length of rope used when a given link is tied, in units of the radius of the rope. The Hopf link has thickness energy of at most $2\pi - 2$.*

parison to their results. Their data also shows a linear relationship between inverse thickness and average crossing number for many small knots (see also [12]).

The data in figure 2 also compares well with the Möbius energies of knots obtained numerically [13]. With the exception of the non-alternating knots, the energies in Table 2 are within narrow bands determined by the crossing numbers of the knot types. The Möbius energy appears to follow that pattern as well, but the bands are now much wider: the standard deviation of 8-crossing knots in Table 2 from the mean energy (63.7) is 1.9; most of the energies lies in a narrow band of half-width 3% the value of the mean energy. In the case of the Möbius energy we obtain roughly 7% (using data from reference [13]). The non-alternating knots also have lower Möbius energies, compared to other 8-crossing knots, but that is not as dramatic as in figure 2. We note that the knot 8_{19} has open thickness energy lower than that of knots with 6 crossings,

but Möbius energy higher than that of any 6-crossing knot.

The open thickness of tangles were measured similarly to the case for knots. Two ropes were used. For the first rope both ends were tied to a fixed object. The second rope was entangled with the first to form a link and then both its ends were attached to a weight. The ropes were lubricated and the link was tied several times and the results averaged. The open thickness energy of the unlink can not be measured experimentally, since the second rope with the attached weight would simply fall to the floor. To obtain a reference marker to which all measurements can be compared, all measurements were measured relative to configuration of the two ropes forming the Hopf Link. We illustrate our results in figure 3. We conjecture that the open thickness of the Hopf Link is $2\pi - 2$, and this value was added to all measurements in figure 3. As in the knot case one can see that the non-alternating energies have exceptional low energies, while in general for alternating links the energy goes up with the crossing number. Finally we note that the data in figure 3 is less accurate than data in figure 2 because the open link measurements result from measurements using only one type of rope.

5. ENERGIES OF POLYGONAL KNOTS

Several polygonal knot energies have been studied numerically in the literature, notably the cases in references [13,14,5,3]. Some of these energies are assumed to be discrete counterparts of energies defined on smooth knots and of these some are semi-ideal. For a knot energy that has a counterpart defined on smooth knots it seems natural to expect that the polygonal energy will have a well-behaved limit. That is, if the number of edges in the polygonal knot is increased to infinity so that the corresponding sequence of polygonal knots converge pointwise to a smooth knot (while the length of the polygons are kept fixed), the polygonal knot energy will converge to the energy defined on the smooth knot. This requirement is difficult to check in general, and we pose a weaker condition: if a sequence of polygons converges to a smooth knot, then the sequence of the corresponding energies of the polygons should converge as well, and as the number of edges approaches infinity, the limit of the minimal polygonal conformation of a knot should be smooth and non-singular (if it exists). Neither of these conditions is satisfied by many polygonal knot energy functions. Consequently, minimizing these energies would not yield desirable knot conformations.

In this section we like to pose two conditions that are necessary for a polygonal energy function to yield nice knot conformations reflecting the complexity of a knot type. We then define an energy function for polygonal knots

satisfying these conditions. Finally a computer simulation will then produce canonical forms for knotted polygons. For a more detailed treatment see [10].

We restrict our attention to equilateral polygons, and we will refer to equilateral polygons as *polygons*. We denote a polygon with n edges by P_n and the set of all polygons of n edges will be denoted by \mathcal{P}_n . If \mathcal{K} is a knot type, then $P_n(\mathcal{K})$ is a polygon of n edges which is also a knot of type \mathcal{K} . Similarly, the set of all such polygons is denoted by $\mathcal{P}_n(\mathcal{K})$. Note that $\mathcal{P}_n(\mathcal{K})$ contains only non-singular polygons while \mathcal{P}_n includes the singular polygons as well. We will also use \mathcal{P} for the set of all polygons and $\mathcal{P}(\mathcal{K})$ for the set of all nonsingular polygons of knot type \mathcal{K} . To ensure that the minimal polygon appears like a smooth knot as $n \rightarrow \infty$, we would like the excluded angles between adjacent edges to get smaller uniformly as n increases. We call such an energy function *asymptotically smooth*. (Thus, if a sequence of polygons converges uniformly to a knot which is not differentiable at a given point, then the corresponding sequence of energies diverges.) In order to carry out computations when a large number of edges is used, the energy of $\mathcal{P}_n(\mathcal{K})$ should not go to infinity as n increases. We call the energy function with this property *asymptotically finite*. The precise definitions are given below.

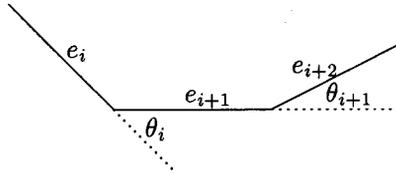


Figure 4: *The Angles Defining the Curvature of a Polygon.*

Definition 5.1. Let f be an energy function of equilateral polygons and let $f_n(\mathcal{K}) = \min\{f(P_n(\mathcal{K}))\}$. f is called an *asymptotically finite* energy function if $\sup_n f_n(\mathcal{K})$ is finite for any knot type \mathcal{K} .

Definition 5.2. Let f be an energy function of equilateral polygons. Let θ_i be the excluded angle between e_i and e_{i+1} of P_n (as shown in figure 4) and let $\theta = \max_i \theta_i$. f is called an *asymptotically smooth* energy function if $\theta \leq \frac{Mf(P_n)}{n}$ for some positive constant M which is independent of n .

Similar to the definitions in section 2 one defines the concept of an ideal polygonal energy. In the definition of basic the role of the circle will be played by a regular polygon. For details see [10]. An energy function of polygonal knots is called *asymptotically ideal* if it is semi-ideal, asymptotically smooth and asymptotically finite. We have to point out that the term “ideal energy”

| Knot | 8 | 16 | 32 | 64 | 128 |
|-------|---------|--------|--------|--------|-------|
| 3_1 | 688.0 | 332.9 | 282.2 | 267.4 | 266.6 |
| 4_1 | 4163.2 | 843.2 | 475.7 | 461.5 | 444.6 |
| 5_1 | 33212.1 | 1015.1 | 727.7 | 655.4 | - |
| 5_2 | 62604.6 | 2217.2 | 782.3 | 710.6 | - |
| 6_1 | - | 2031.1 | 1030.8 | 882.9 | - |
| 6_2 | 11820.4 | 4288.6 | 1012.3 | 901.3 | - |
| 6_3 | - | 3548.9 | 1274.8 | 1058.3 | - |

Table 1: *Polygonal Knot Energies*

is somewhat misleading since minimizing an asymptotically ideal energy may still not produce an ideal smooth conformation representing the given knot type. But on the other hand, it does give us a better chance. In fact, many known polygonal knot energies are not asymptotically ideal (see [10] for examples), thus we can eliminate these energy functions when trying to find an ideal smooth conformation to represent a knot type. The following definition provides us an asymptotically ideal energy function.

Definition 5.3. Let P_{2n} be an equilateral polygon of $2n$ edges $e_1, e_2, e_3, \dots, e_{2n}$ with total length L and define $d_{i,k}$ to be the shortest (Euclidean) distance between e_i and e_{i+k} (with $e_{i+k} = e_{i+k-2n}$ if $i+k > 2n$). The arclength between e_i and e_{i+k} is defined as the shortest arclength between any point on e_i and any point on e_{i+k} and is denoted by $s_{i,k}$. Define

$$E(P_{2n}) = \max_{1 \leq i \leq 2n, 2 \leq k \leq n} L^2 \left(\frac{1}{d_{i,k}^2} - \frac{1}{s_{i,k}^2} \right). \quad (4)$$

Theorem 5.4. The energy function $E(P_{2n})$ defined above (over the set of all equilateral polygons of even number of edges) is an asymptotically ideal energy function.

The proof of Theorem 6.4 is long and technical and is presented in [10]. Computer simulations to estimate the minimal energies of given polygonal knots were performed and reported also in reference [10]. The minimal energies are tabled in table 1 for $2n = 8, 16, 32, 64$ and 128. In figure 5 we display stereographic images of equilateral polygonal knots (all are trefoils). Cross your eyes to superimpose the side-by-side images for depth-perception.

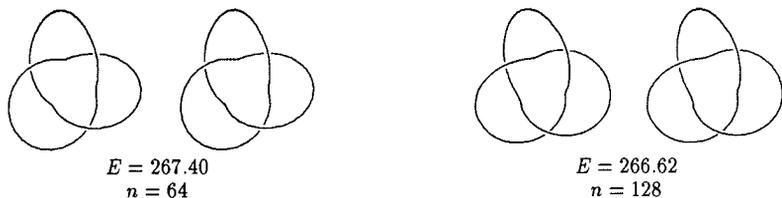


Figure 5: *Stereographic projections of the equilateral polygonal trefoil knots.*

6. CONCLUSION

In this paper we have outlined strategies of finding canonical conformations of knots and their associated minimal energies using energy functions which have nice theoretical properties. The success of our approaches however remains quite limited. To compute energy functions analytically seems virtually impossible. Numerical simulations are very involved, convergence of these computations is slow and the theory that these conformations of polygons reflect conformations of smooth knots is incomplete. The experiments using ropes yield surprisingly consistent data, however it seems to be impossible to increase the precision of these measurements significantly or to use these methods to measure knot energies of knots with higher crossing number.

Acknowledgments: The authors want to thank Douglas Humphrey in the Physics department at Western Kentucky University, Gregory Slone, an undergraduate at Western Kentucky University and the Physics Department at York University, for providing space and help for the rope experiments.

References

- [1] J. O'Hara, *Energy of a Knot*, *Topology*, **30** (1991), 241-247.
- [2] Moffatt, *The Energy Spectrum of Knots and Links*, *Nature*, **347** (1990), 367-369.
- [3] G.R. Buck and J. Orloff, *Computing canonical conformations for knots*, *Topology Appl.*, **51** (1993), 246-253.
- [4] M. H. Freedman, Z-X He and Z. Wang, *Möbius Energy of Knots and Unknots*, *Annals of Mathematics*, **139** (1994), 1-50.

- [5] J. K. Simon, *Energy Functions for Polygonal Knots*, Journal of Knot Theory and its Ramifications, Vol **3**, No **3** (1994), 299-320.
- [6] R.A. Litherland, J. Simon, O. Durumeric and E. Rawdon, *Thickness of Knots*, preprint.
- [7] Y. Diao, C. Ernst and E.J. Janse Van Rensburg, *Properties of Knot Energies*, To appear in Topology and Geometry of Polymers (eds. S.G. Whittington, D.W. Sumners and T. Lodge), IMA Volumes in Mathematics and its Applications.
- [8] Y. Diao, C. Ernst and E.J. Janse Van Rensburg, *Thicknesses of Knots*, Preprint.
- [9] Y. Diao, C. Ernst and E.J. Janse Van Rensburg, *Knot Energies by Ropes*, to appear in J. Knot Theory and its Ramification.
- [10] Y. Diao, C. Ernst and E.J. Janse Van Rensburg, *In Search of a Good Polygonal Knot Energy*, to appear in J. Knot Theory and its Ramification.
- [11] V. Katritch, J. Bednar, D. Michoud, R.G. Scharein, J. Dubochet and A. Stasiak, *Geometry and Physics of Knots*, Nature, **382** (1996), 142-145.
- [12] A. Stasiak, V. Katritch, J. Bednar, D. Michoud, and J. Dubochet, *Electrophoretic Mobility of DNA Knots*, Nature, **382** (1996), 122.
- [13] R.B. Kusner and J.M. Sullivan, *Möbius Energies for Knots and Links*, *Surfaces and Submanifolds*, Preprint.
- [14] T.J. Ligocki and J.A. Sethian, *Recognizing Knots using Simulated Annealing*, J. Knot Theory and its Ramification, **3** (1994), 477-496.

CHAPTER 5

THE WRITHE OF KNOTS AND LINKS

E J JANSE VAN RENSBURG

Department of Mathematics and Statistics, York University, North York, Canada

D W SUMNERS

Department of Mathematics, Florida State University, Tallahassee, USA

S G WHITTINGTON

Department of Chemistry, University of Toronto, Toronto, Canada

Abstract

We discuss the writhe of linked and knotted simple closed curves embedded in the simple cubic lattice, Z^3 . We show that the writhe of a simple closed curve in Z^3 can be computed as the average of its linking numbers with certain pushoffs, and use this result to establish a lower bound on the rate of increase of the mean absolute writhe. We present Monte Carlo results on the distribution of writhe for particular knot types, and compare the mean values with values for ideal knots. Similar results are presented for links and we show that the mean writhe of $(2, 2k)$ torus links increases linearly with crossing number.

1 Introduction

A time-honoured model of the configurational properties of long linear polymer molecules in dilute solution in a good solvent is the self-avoiding walk. This is a walk (a sequence of edges and vertices) on a regular lattice, such as the simple cubic lattice, Z^3 . No two vertices of the walk can occupy the same lattice point, and all self-avoiding walks with the same number (n) of edges have the same probability of occurrence. The self-avoiding condition is designed to mimic the excluded volume effect in polymers, which says that no two monomers can occupy the same region of space. These lattice models have the advantage that they can be treated by combinatorial methods and, by using a computer, one can make use of efficient data structures because of the discrete nature of the space Z^3 .

In the same way, ring polymers can be modelled by (self-avoiding) poly-

gons. A polygon is an embedding of a simple closed curve in Z^3 and two polygons are considered distinct if they can not be superimposed by translation. If we write p_n for the number of polygons with n edges then $p_n = 0$ if n is odd, $p_4 = 3$, $p_6 = 22$, $p_8 = 207$, etc. In this paper we shall be concerned with the geometric properties of polygons and especially with their writhe, which is a measure of the geometrical entanglement complexity of space curves, and which we define in Section 2. Natural questions which one can ask are: "What is the distribution of the values of the writhe?", "How does the mean writhe, or mean square writhe, depend on n ?", etc.

Calculating the writhe of a simple closed curve in R^3 is usually done by a stochastic numerical approximation, but we shall see in Section 2 that there is a considerable simplification when the curve is embedded in a lattice. Indeed, for polygons on the simple cubic lattice, four times the writhe is always an integer, and an exact calculation for the writhe can easily be performed. The approach described in Section 2 is an important ingredient in a proof, outlined in Section 3, that the expectation (over all polygons with n edges) of the absolute value of the writhe increases at least as fast as \sqrt{n} .

In Section 4 we describe a Monte Carlo method for estimating the distribution of the writhe when the polygons are conditioned to have a specified knot type, and in Section 5 we describe some results obtained in this way. It seems that the distribution of the writhe increases in width as n increases, but that the mean writhe is more or less independent of n . We shall see that these mean values are very close to the values estimated by Katritch *et al*¹ for the writhe of "ideal" knots. In addition we note that the mean writhe is a linear function of crossing number for knots in certain families (such as torus knots, twist knots with odd crossing number, twist knots with even crossing number, etc.), and seems to be additive under the process of connect sum when composite knots are formed from their prime components. This additivity under connect sum was also observed by Katritch *et al*². In addition, our results suggest that linearity of writhe for certain homologous families of knots and links, and additivity of writhe, are not dependent on particular features of ideal representations of knot and link types.

Similar questions can be asked about the writhe of links, and we address these questions in Section 6. Perhaps the most striking result is the linear dependence of average writhe on crossing number in the sequence of $(2, 2k)$ torus links. This has been observed experimentally in circular DNA, and numerically in a particular model of DNA links³. Our results suggest that the feature is a more general one, and is not dependent on the particular features of DNA molecules.

2 Computing the writhe of lattice polygons

The writhe of a simple curve in 3-space is a geometric quantity intended to measure the signed non-planarity of a curve. The writhe of a space curve ω ($Wr(\omega)$) is defined in terms of orthogonal projections of ω onto 2-dimensional hyperplanes. We will restrict attention to simple closed curves ω . A projection direction is an oriented line in R^3 ; think of the projection plane as being below the curve ω . Given a knot diagram $\hat{\omega}$ determined by a regular orthogonal projection of the simple closed curve ω , orient the planar curve $\hat{\omega}$, and assign either +1 or -1 to each self-crossing of $\hat{\omega}$, according to the usual crossing sign convention for any diagram of a pair of oriented skew lines in 3-space (see figure 1). The sum of the signed crossings is the *signed crossing number* of that knot diagram; it is an integer-valued function of the space curve and the projection direction. For a fixed space curve, one averages this function over all projection directions (points on a reference 2-sphere) obtaining the writhe of the curve.

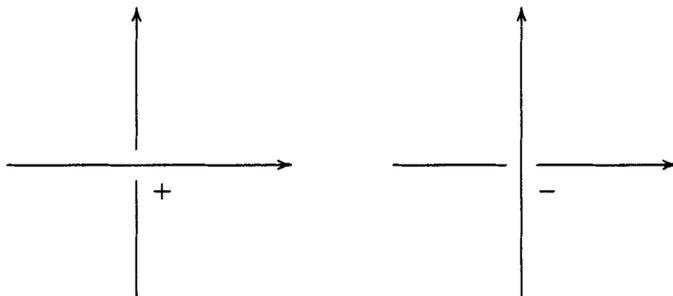


Figure 1: The sign-convention for projected arcs crossing each other in the plane.

In order to compute the writhe, this average value can be approximated in a stochastic fashion by randomly and uniformly choosing projection directions in which to sample the signed crossing number function. It is possible, however, to obtain exact calculations for curves constrained to live on spatially periodic lattices in R^3 . In order to do this, one converts the signed crossing number of a planar diagram to a linking number; the topological invariance of linking number supplies the needed flexibility to achieve an exact calculation. The signed crossing number of a knot diagram can be viewed as a linking number of a pair of oriented space curves determined by the knot diagram as follows: travel along the projected curve $\hat{\omega}$ and push off by a small amount

along a vector orthogonal to $\hat{\omega}$ in the direction of the unbounded region of the plane of the diagram. One obtains the planar pushoff $\hat{\omega}_p$, a companion planar curve which is parallel to $\hat{\omega}$. Orient $\hat{\omega}_p$ parallel to $\hat{\omega}$. Near each self-crossing of $\hat{\omega}$ (or $\hat{\omega}_p$) the overcrossing segment of the curve can be vertically pushed a small distance above the diagram plane, generating a pair of space curves (ω^*, ω_p^*) corresponding to the pair of planar curves $(\hat{\omega}, \hat{\omega}_p)$. Thought of as 3-dimensional objects in this way, the linking number $Lk(\omega^*, \omega_p^*)$ is the signed crossing number of the knot diagram $\hat{\omega}$. The pair (ω^*, ω_p^*) cobound a small annulus in 3-space. Think of the small annulus as a curtain, and rotate the curtain downward until it hangs vertically downward from ω^* toward the diagram plane. The planar pushoff has now moved to the position of a spatial pushoff $\hat{\omega}_p^*$ of ω^* in the projection direction, and $Lk(\omega^*, \omega_p^*) = Lk(\omega^*, \hat{\omega}_p^*)$. The pair $(\omega^*, \hat{\omega}_p^*)$ is ambient isotopic by pushing vertically to the pair (ω, ω_p) , where ω_p is a pushoff of ω vertically downward in the projection direction. Hence, the $Wr(\omega)$ is converted to the average value of linking numbers of ω with pushoffs in all projection directions.

If a simple closed space curve is constrained to lie on a regular space-filling lattice, the number of pushoffs needed for an exact computation of the writhe is finite. As will be argued below, this is because the "coordinate" planes of the lattice subtend finitely many compartments on a small 2-sphere about the origin, and the linking number of the curve and all pushoffs along directions in the interior of the same compartment have the same linking number. For simplicity, we argue⁴ the case for the simple cubic lattice Z^3 . Suppose that we wish to compute the writhe of the polygon ω in Z^3 . Let S^2 be the 2-sphere in R^3 of radius $\frac{1}{2}$; S^2 is the space of directions for push-offs. The three coordinate planes in R^3 separate S^2 into eight connected regions ("octants"), characterized by constancy of sign in each coordinate. If $\mu \in S^2$, then $\omega_\mu = \omega + \mu$ is the push-off of ω in the direction μ .

Claim 1 : *If μ lies in the interior of any octant on S^2 , then ω and ω_μ are disjoint space curves and $Lk(\omega, \omega_\mu)$ is defined.*

Proof: Suppose that $\mu = (\mu_1, \mu_2, \mu_3)$ lies in the interior of the first octant, where all three coordinates are positive; since μ lies on a 2-sphere of radius $\frac{1}{2}$, then $0 < \mu_i < \frac{1}{2}$ for $i = 1, 2, 3$. Now suppose that $(x, y, z) \in \omega$ and $(x + \mu_1, y + \mu_2, z + \mu_3) \in \omega \cap \omega_\mu$. Points in ω have the property that at least two of the coordinates are integers. By the pigeonhole principle, at least one of the following is true: both x and $x + \mu_1$ are integers; both y and $y + \mu_2$ are integers, or both z and $z + \mu_3$ are integers. This means that at least one of $\{\mu_1, \mu_2, \mu_3\}$ is an integer, which is impossible. A similar argument holds for

the other seven octants. \square

Claim 2 : *If μ and ν lie in the interior of the same octant, then $Lk(\omega, \omega_\mu) = Lk(\omega, \omega_\nu)$.*

Proof: Consider the shorter of the two great circle arcs connecting μ and ν on S^2 ; this arc lies in the interior of the same octant as do μ and ν . The points along this arc define a 1-parameter family of pushoff directions, starting with μ and ending with ν . Thus the pushoff ω_μ can be ambient isotoped to the pushoff ω_ν in the complement of ω , and the linking numbers are equal. Since all of the eight octants in S^2 have equal area, at this point we have shown that $Wr(\omega)$ is the average of eight linking numbers, one for each octant. We now use the symmetry of the simple cubic lattice to reduce the number of pushoffs needed to four. \square

Claim 3 : *If μ is not on a coordinate plane, then $Lk(\omega, \omega_\mu) = Lk(\omega, \omega_{-\mu})$.*

Proof: Let t be a parameter such that $-1 \leq t \leq 0$. As in Claim 1, for each value of t , ω_μ and $\omega_{t\mu}$ are disjoint. To see this, suppose that $(x + \mu_1, y + \mu_2, z + \mu_3) = (x^* + t\mu_1, y^* + t\mu_2, z^* + t\mu_3)$ for some $(x, y, z), (x^*, y^*, z^*) \in \omega$. Suppose also that both x and x^* are integers. It follows that $(x - x^*) = (1 - t)\mu_1$. But $(1 - t)\mu_1$ cannot be an integer, because $1 \leq (1 - t) \leq 2$, and $0 < |\mu_1| < \frac{1}{2}$. This means that we can ambient isotope ω to $\omega_{-\mu}$ in the complement of ω_μ , so $Lk(\omega_\mu, \omega) = Lk(\omega_\mu, \omega_{-\mu})$. Similarly, one can ambient isotope ω_μ to ω in the complement of $\omega_{-\mu}$, hence $Lk(\omega_{-\mu}, \omega_\mu) = Lk(\omega_{-\mu}, \omega)$. The claim now follows by symmetry of linking numbers in R^3 . \square

From the above three claims it is clear that $Wr(\omega)$ is the average of linking numbers of ω with pushoffs into four mutually non-antipodal octants. In particular, it follows that four times $Wr(\omega)$ is an integer for any polygon $\omega \in Z^3$. As we shall see, the fact that the writhe of a lattice polygon is the average of four linking numbers is useful in both calculational and theoretical (theorem proving) contexts.

3 Width of the writhe distribution

In this section we discuss how the width of the distribution of writhe of polygons in Z^3 depends on the number of edges in the polygon⁵. In particular, can we use the result in the previous section to examine the behaviour of the mean absolute writhe as a function of n ? We expect the mean absolute writhe to increase with n , and we shall prove that this increase is at least as fast as

\sqrt{n} .

We define two polygons in Z^3 to be identical if they can be superimposed by a translation. Let P_n be the set of polygons of length n in Z^3 , and let the number of polygons in P_n be p_n . If ω is a polygon in P_n , then let $Wr(\omega)$ be the writhe of ω . The expected value of the writhe for a polygon with n edges in Z^3 , $\langle Wr \rangle_n$, is defined by

$$\langle Wr \rangle_n = \frac{1}{p_n} \sum_{\omega \in P_n} Wr(\omega). \quad (1)$$

By symmetry, $\langle Wr \rangle_n = 0$. It will be more rewarding to study the expected absolute writhe, $\langle |Wr| \rangle_n$, or the expected square writhe, $\langle Wr^2 \rangle_n$, instead. We saw in the previous section that writhe of a given polygon can be computed by taking the average of linking numbers of the polygon with four pushoffs of itself into non-antipodal octants in Z^3 . We use this result now to study the expectation of the absolute writhe. We shall divide the proof into two parts. In the first part, we show that small curly-cues can be truncated from a given polygon, and that the writhe is additive with respect to the truncation. In the second part we adapt a coin-tossing argument to study the expected absolute writhe.

Define a *curly-cue* to be the self-avoiding walk defined by the sequence of canonical unit vectors $\{i, i, -k, -j, -i, j, j, k, -i\}$.^a The first and last vertices in a curly-cue can be joined by adding an extra edge in the j -direction giving a polygon \mathcal{P} (see figure 2).

In order to compute the writhe of \mathcal{P} , we construct its pushoffs and compute its linking numbers with its pushoffs. Let \mathcal{P}_1 be the pushoff of \mathcal{P} obtained by adding $(i + j + k)/2$ to each point of \mathcal{P} . Direct computation show that the linking number is $Lk(\mathcal{P}, \mathcal{P}_1) = 1$. Push-offs by $(i - j + k)/2$ and $(-i - j + k)/2$ also give linking numbers equal to 1, but the push-off by $(-i + j + k)/2$ gives a linking number of -1 . Hence, $Wr(\mathcal{P}) = \frac{1}{2}$. Similarly, if \mathcal{P}^* is the mirror image of \mathcal{P} , reflected through the xy -plane, then $Wr(\mathcal{P}^*) = -\frac{1}{2}$.

Let \mathcal{C} be the cube formed by the union of the 3-cells dual to any translation in Z^3 of the set of vertices with coordinates in the set $\{0, 1, 2\} \times \{0, 1, 2\} \times \{0, 1, 2\}$. Let A be a polygon which intersects \mathcal{C} in the curly-cue $\{i, i, -k, -j, -i, j, j, k, -i\}$ with first and last vertices P and Q . We can *truncate* this subwalk from A by removing it from A , and by joining P and Q with an edge in the j -direction in A . The subwalk can be turned into \mathcal{P} by joining

^a Any rotation of this sequence in Z^3 is also a curly-cue.

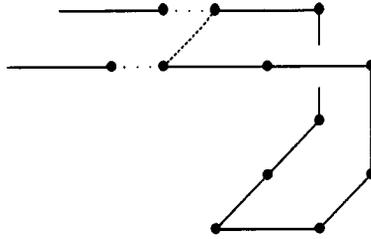


Figure 2: A curly-cue can be truncated from a polygon as shown here. The writhe of the polygon is additive with respect to this truncation.

its ends with an edge in the j -direction. Suppose we obtain a polygon A' if we truncate the curly-cue from A . Then

$$Wr(A) = Wr(A') + Wr(\mathcal{P}). \quad (2)$$

To see this, consider push-offs A_1 of A and A'_1 of A' in the $(i+j+k)/2$ direction, as illustrated in figure 3. Define $Lk(A, A_1)$ to be linking number of A and A_1 .

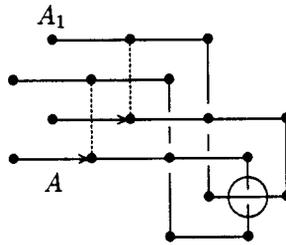


Figure 3: The projection in the xy -plane of a section of a polygon (and one of its push-offs) which contains a curly-cue. The curly-cue can be truncated at the dashed lines.

In figure 3 a projection of A and its push-off in the $(i+j+k)/2$ direction in the vicinity of the cube \mathcal{C} is shown. If the crossing which is circled is reversed, then A and A_1 can be deformed by an isotopy in the cube \mathcal{C} (for example, by performing Reidemeister moves on figure 3 after reversal of the crossing) to the pair A' and A'_1 (that is, with the curly-cue truncated at the dashed lines in figure 3). Thus, truncating the curly-cue is equivalent (in a topological sense) to reversing the single crossing encircled in figure 3. Let \mathcal{P} and \mathcal{P}_1 be

the polygons obtained by closing the curly-cues on truncation. Then

$$Lk(A, A_1) = Lk(A', A'_1) + 1 = Lk(A', A'_1) + Lk(\mathcal{P}, \mathcal{P}_1). \quad (3)$$

The remaining three push-offs can be similarly analyzed, and the result is equation (2). This is an important result for us, since we know exactly the change in the writhe of a polygon if we truncate a curly-cue. We can now show that the expected value of the absolute writhe increases at least as fast as $O(\sqrt{n})$.

Let $\langle |Wr| \rangle_n$ be the mean of the absolute value of the writhe of polygons of length n .

Theorem 1 *The probability $\text{Prob}(\langle |Wr| \rangle_n < f(n))$ approaches zero as $n \rightarrow \infty$ for any function $f(n) = o(\sqrt{n})$.*

Proof: Let $P = (C, B)$ be a ball-pair consisting of any translate of the cube C and the curly-cue $B = \{i, i, -k, -j, -i, j, j, k, -i\}$. We call P a pattern, and we define $P^* = (C, B^*)$ to be the mirror image of P (reflected through the xy -plane). Let $\epsilon > 0$ be given and let $\alpha_n(\epsilon)$ be the number of polygons of length n which have at least $\lfloor \epsilon n \rfloor$ occurrences of the patterns P or P^* . (In other words, these polygons intersect at least $\lfloor \epsilon n \rfloor$ translated copies of the cube C in either B or B^*). A pattern theorem by Kesten⁶ states that for sufficiently small ϵ and sufficiently large n , there exists a $\gamma > 0$ such that $(1 - e^{-\gamma n})p_n \geq \alpha_n(\epsilon)$, where p_n is the number of polygons of length n . Let a polygon counted by $\alpha_n(\epsilon)$ contain P or P^* exactly $D \geq \lfloor \epsilon n \rfloor$ times. Then P occurs exactly k times with probability

$$\binom{D}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{D-k}.$$

This is a maximum when $k = \lfloor D/2 \rfloor$, and we obtain the upper bound $1/\sqrt{\lfloor \epsilon n \rfloor}$ on this probability. The writhe of a polygon ω counted by $\alpha_n(\epsilon)$ has contributions from the $D \geq \lfloor \epsilon n \rfloor$ copies of P or P^* , as shown by the truncation in figure 3 and by equation (2). Let ω' be the polygon obtained by truncating all the curly-cues from ω . Then by equation (2), $Wr(\omega) = Wr(\omega') + Wr(\text{curly-cues})$. If $|Wr(\omega)| \leq f(n)$, then this implies that $Wr(\text{curly-cues})$ can assume at most $\lceil 2f(n) + 1 \rceil$ different values. Thence, the probability that the absolute writhe of ω is less than $f(n)$ is bounded from above by the probability that $Wr(\text{curly-cues})$ assume one of at most $\lceil 2f(n) + 1 \rceil$ values. But

$$\text{Prob}(|Wr(\omega)| \leq f(n)) \leq \frac{\lceil 2f(n) + 1 \rceil}{\sqrt{\lfloor \epsilon n \rfloor}}, \quad (4)$$

Thus

$$\text{Prob}(|Wr|_n \leq f(n)) \leq (1 - e^{-\gamma n}) \frac{[2f(n) + 1]}{\sqrt{[\epsilon n]}} + e^{-\gamma n} R, \quad (5)$$

where R is the contribution from polygons not counted by $\alpha_n(\epsilon)$. This approaches zero as $n \rightarrow \infty$ provided that $f(n) = o(\sqrt{n})$, as claimed above. \square

In theorem 1 we have focussed entirely on the contribution of small curly-cues to the writhe of a polygon. In particular, this restriction enables us to prove that the expected absolute writhe of a random polygon grows at least as fast as the square root of its length. A model of self-avoiding walks with a fugacity coupled to writhe has recently been studied by an identification to the $N \rightarrow 0$ limit (N is the number of scalar components) of a complex scalar Chern-Simon theory.⁷ In particular, they found that the conformation of the self-avoiding walk is decoupled from its writhe in the scaling limit. This can be interpreted as indicating that the total writhe (and therefore the total absolute writhe) of a self-avoiding walk is a consequence of contributions from local conformations in the walk, which are invisible from the long wavelength field theory (presumably, the local conformations are plectonemic wound conformations of size no larger than that of the persistence length of the walk). In our context, this suggests that the total writhe of a polygon is primarily determined by curly-cues. We therefore consider now the contribution to the absolute writhe of the polygon ω by the curly-cue pattern P and its mirror image P^* . Thus, if P occurs exactly k times, and there are a total of D occurrences of P or P^* , the contribution to the absolute writhe is $|2k - D|$. The expected value of this contribution for ω is given by

$$\begin{aligned} E_\omega(|2k - D|) &= \sum_k |2k - D| \binom{D}{k} 2^{-D} \\ &= 2^{-D+1} D \sum_{k < D/2} \binom{D}{k} - 2^{-D+2} \sum_{k < D/2} k \binom{D}{k}, \end{aligned} \quad (6)$$

after some algebra. This may be simplified to give

$$E_\omega(|2k - D|) = 2^{-D+1} D \binom{D-1}{(D-1)/2} \quad \text{if } D \text{ is odd,} \quad (7)$$

and

$$E_\omega(|2k - D|) = 2^{-D} D \binom{D}{D/2} \quad \text{if } D \text{ is even.} \quad (8)$$

But for any polygon counted by $\alpha_n(\epsilon)$, $[\epsilon n] \leq D \leq n$, and by noting that $\binom{D}{D/2} \sim 2^D D^{-1/2}$, we obtain the result that

$$E_\omega(|2k - D|) \sim n^{1/2}. \quad (9)$$

In other words, the contribution to the absolute writhe of the curly-cues grows as the square root of n . An important open question is to derive an upper bound on the n -dependence of the absolute writhe of the polygon, rather than focussing on the curly-cues. Numerical evidence⁵ suggests that the mean absolute writhe might grow like \sqrt{n} . Using techniques similar to the above, we can compute the expectation values $E_\omega(|2k - D|^M)$.

4 Numerical methods

In this section we discuss a Monte Carlo algorithm for the sampling of knotted polygons in Z^3 . In particular, we are interested in sampling polygons of a fixed (given) knot type. We shall show that the BFACF algorithm⁸ is suitable for our purposes.

The BFACF algorithm has two elementary moves, illustrated in figure 4. The first of these moves changes the length of the polygon, while the second is a length preserving move. By applying these moves, the algorithm samples along a realisation of a Markov chain defined on the state space of polygons. It has one free parameter β , which can be set to control the expected length of polygons in the sample.

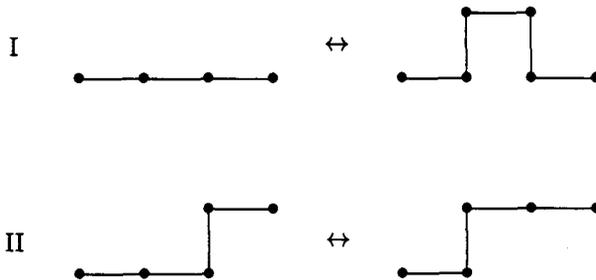


Figure 4: The elementary moves of the BFACF algorithm. Moves of type I can increase or decrease the length of the polygon by two edges. Moves of type II are length preserving.

The implementation of the algorithm is as follows: Let ω be the current state in the Markov chain. Select an edge uniformly in ω , and move the edge perpendicular to itself one lattice spacing in one of four possible lattice directions, while two new edges are inserted to keep the polygon connected. If this operation produces double edges then they are deleted. An enumeration of the possible outcomes shows that all possible results are one of the cases in figure 4. Let the new state be ν . If ν is not a polygon (i.e. if the vertices are not all of degree 2) then ν is rejected and ω is the next state of the Markov chain. If ν is a polygon then it is accepted or rejected as the next state by applying a Metropolis rule; if it is rejected, then ω will be taken as the next state in the chain. Let the number of edges in ω be $|\omega|$ and in ν be $|\nu|$. If $|\omega| \geq |\nu|$, then we accept ν as the next state in the chain; if $|\omega| = |\nu| - 2$, then we accept ν with probability β^2 as the next state. Since the edge is selected with probability $1/|\omega|$, and moved perpendicular to itself in one of four directions with probability $1/4$, the probability that ν is the next state in the Markov chains is given by

$$P(\omega \rightarrow \nu) = (\mathcal{I}(|\omega| \geq |\nu|) + \beta^2 \mathcal{I}(|\omega| < |\nu|)) / 4|\omega|, \quad (1)$$

where \mathcal{I} is an indicator function. On the other hand, by interchanging ω and ν in (1), we obtain the probability that ω is the next state in the Markov chain if ν is the current state. A comparison of these probabilities gives us the equation of detailed balance for this algorithm:

$$|\omega| \beta^{|\omega|} P(\omega \rightarrow \nu) = |\nu| \beta^{|\nu|} P(\nu \rightarrow \omega). \quad (2)$$

Notice that the algorithm is aperiodic (since there is a non-zero probability that a move will be rejected). The elementary moves in figure 4 are homeomorphisms on the polygon, and the knot type of the polygon along the Markov chain is therefore invariant. Therefore, the algorithm is reducible on the state space of all polygons, and we must find its *ergodicity classes*^b if we are to understand its behaviour in any detail. The knot type of the polygon is set by the first polygon sampled by the chain. Let $\mathcal{P}(K)$ be the set of all polygons of knot type K . Then any ergodicity class, defined by an initial polygon of knot type K , is a subset of $\mathcal{P}(K)$. Remarkably, the ergodicity classes of the polygon are exactly the sets $\mathcal{P}(K)$, since it is known⁹ that any two polygons of the same knot types can be connected by a finite sequence of the moves in figure 4. By summing the above over all possible states ν in a given ergodicity class, we see from the fundamental theorem of Markov chains that the invariant limiting

^bThese are the subsets of the state space on which the algorithm is irreducible.

distribution of the algorithm is given by

$$\Pi_\omega(K) = \frac{1}{Z_K(\beta)} |\omega| \beta^{|\omega|} \mathcal{I}(\omega \in \mathcal{P}(K)), \quad \text{where } Z_K(\beta) = \sum_{\omega \in \mathcal{P}(K)} |\omega| \beta^{|\omega|}. \quad (3)$$

We conclude that this algorithm is suitable for sampling polygons with a given knot type. The invariant limit distribution is given by (3), where polygons of the same length have the same weight. The expected length of polygons sampled by the algorithm is given by

$$\langle |\omega| \rangle_\beta = \sum_{\omega \in \mathcal{P}(K)} |\omega| \Pi_\omega(K). \quad (4)$$

This expected length is known to be infinite if β is larger than a critical value β_c . If $\beta < \beta_c$, then the expected length of polygons stays finite, and the algorithm can be used to sample finite polygons of a fixed knot type K from the distribution $\Pi_\omega(K)$ to compute their properties.

5 Writhe as a function of knot type

In 1980 Le Bret¹⁰ computed the writhe of closed Gaussian piece-wise linear curves in R^3 as a model of closed circular DNA. If the knot type of the curve is fixed, he noticed that the mean writhe is zero for the unknot and for the achiral knot 4_1 , but not for the chiral knots which he studied. These issues were revisited by Katritch *et al*¹ who published a similar study of “ideal knots” in \mathcal{R}^3 and of a simple model of circular DNA. These questions were also studied in the cubic lattice¹¹, and in this section we recount this study, and compare our findings with those of Le Bret and Katritch *et al*.

Any achiral knot K has a mirror image K^* of the same knot type. Since the writhe of K changes sign under reflection, but retains its magnitude, this observation proves that the mean writhe of any achiral knot is zero. Thus, if the mean writhe of a knot is not zero, then it is chiral.^c If writhe can be computed by a Monte Carlo algorithm, then we have a good method for determining, up to numerical error, whether a given knot is chiral. It is fortunate that the writhe of a lattice knot can be computed exactly by taking the average of the linking number of the knot with its pushoffs into four non-mutually antipodal octants. This is a tremendous improvement, since the writhe is defined as the mean of the sum of the signed crossings over all regular projections of the knot.

^cAlas, the converse of this statement is not known to be true.

We sampled along a Markov chain in the state space of lattice knots of a fixed knot type K by using the BFACF algorithm described in section 4. The writhe of the lattice knots along this chain was obtained by computing the linking numbers of the lattice knots with push-offs into four non-antipodal octants. The linking numbers were computed by performing heap sorting algorithms on the coordinates of the polygons. This allowed the calculation of a linking number in $O(n \log n)$ CPU-time (as opposed to the $O(n^2)$ CPU-time which would be required for a naive search for intersections in projections of the knot and its push-offs.) Several runs were performed for each knot type, with several different values of the parameter β selected each time in an attempt to sample adequately polygons with length up to 250 edges. For each fixed value of n we tabulated the distribution of writhe, and computed a mean writhe over the distribution. The distribution is plotted in figure 5 for the knot 5_1 . Notice that there are no data for $n < 34$; the best upper bound¹² on the number of edges necessary to tie the knot 5_1 is 34. The distribution widens as n increases, as we expect following the arguments in section 3, although this has not been made rigorous for the situation here. This increase in the width of the distribution was also observed by Le Bret¹⁰. The mean writhe of the knot seems to be independent of n , even though n increases from 34 up to 250. This observation seems to be true for every knot that we have examined. We can therefore estimate the mean writhe of the lattice knot by pooling all the data in figure 5 to arrive at a single estimate of the writhe.^d

Our best estimate for the writhe of the trefoil is 3.441 ± 0.025 , and estimating the writhe of 5_1 from the data in figure 5 gives 6.254 ± 0.011 . The mean writhe for other knots are tabulated in the second column of table 1. All the knots in this table are chiral, since we know that achiral knots have mean writhe zero. We also compare our results with estimates obtained by Le Bret¹⁰ and Katritch *et al*¹ (the estimates for the trefoil by Le Bret vary between 3.5 for the smallest polygons, and 3.15 for the largest polygons; we give the value obtained at the largest polygons in table 1). The data from the work of Katritch *et al*¹ are for an "ideal knot", and a model of circular DNA with 5400 base pairs and with 1800 base pairs respectively. The agreement in the data is remarkable. An interesting observation from table 1 is the appearance of linear relationships between the crossing number and the mean writhe in some knot families. (Katritch *et al*¹ also observed this in their model.) One such family is the family of $(2, 2k + 1)$ -torus knots with first members $3_1, 5_1, 7_1, 9_1$, and so on. A linear fit to our data gives $\langle Wr \rangle \approx 1.409C - 0.789$ where

^dWe achieved this by plotting the mean writhe as a function of n , and carrying out a linear least squares analysis. The y-intercept was taken as the mean writhe, and we observed that the slope of the fitted line was very close to zero in every case.

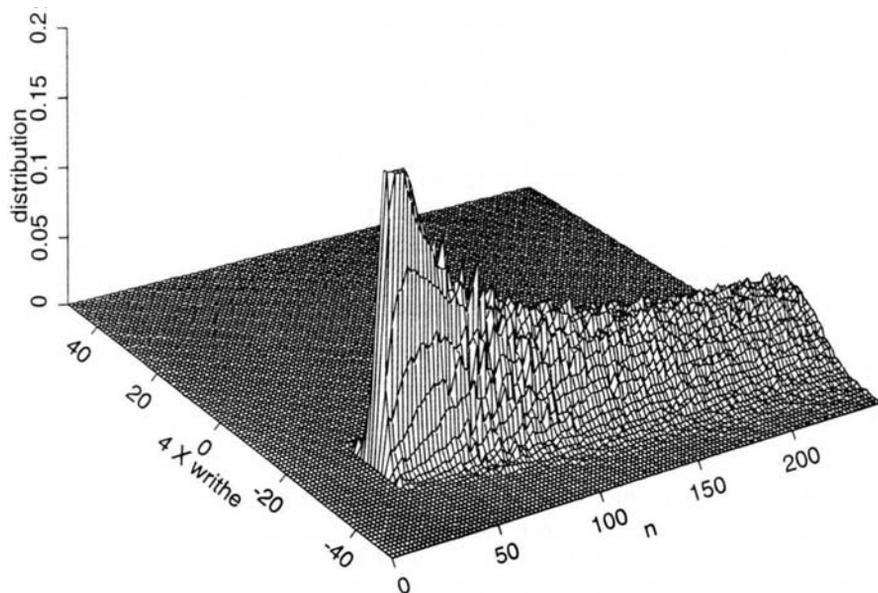


Figure 5: The distribution of writhe for a lattice knot of type 5_1 . The distribution for each value of n has been normalised to 1. Notice that the minimal number of edges to tie the knot in \mathbb{Z}^3 is 34, so that there are no data for polygons shorter than 34 edges.

C is the crossing number, and the accuracy of this fit is very good (the residual deviance is approximately 2×10^{-5}). Similar relationships can be obtained for other families of knots, such as the even crossing number twist knots ($4_1, 6_1, 8_1, \dots$) and odd crossing number twist knots ($3_1, 5_2, 7_2, \dots$). Linear fits to our data show that the rate of increase in mean writhe with crossing number is approximately 0.59 in both these families. These observations remain true even in the case of more “complicated” knot families. We have considered the families of knots with Conway symbols $(2k + 1, 1, 2k)$, $(2k - 1, 1, 2k)$ and $(1, 2k - 1, 1, 2k - 1, 2)$, and we observe similar linear behaviour.

We can concatenate prime knots in the lattice to form composite knots by the following construction: The *top edge* of a polygon is the edge with lexicographically largest midpoint, and the *bottom edge* is that edge with lexicographically least midpoint. Two polygons are concatenated by translating one polygon (and rotating it if necessary) such that the midpoint of its bottom edge has first coordinate which is one more than the first coordinate of the midpoint of the top edge of the second polygon. We concatenate the two polygons by first deleting the bottom edge of the first and the top edge of the second polygon, before we add two edges incident with the endpoints of the deleted top and

bottom edges to connect the two polygons into a single, larger polygon. If the first polygon had knot type K_1 and the second K_2 , then the new polygon has knot type $K_1 \# K_2$. The writhe of polygons is additive under concatenation.¹¹ This result suggests a much more general case: the possibility that writhe is additive under the connect sum of lattice knots: In other words, the mean writhe of a lattice knot of type $K_1 \# K_2$ is the sum of the mean writhe of lattice knots of types K_1 and K_2 . Numerical work supports this notion strongly; for example, we measured $\langle Wr(3_1^-) \rangle = -3.441 \pm 0.025$, $\langle Wr(5_1^-) \rangle = -6.254 \pm 0.011$, while $\langle Wr(5_1^- \# 3_1^-) \rangle = -9.657 \pm 0.58$ and $\langle Wr(5_1^- \# 3_1^+) \rangle = -2.790 \pm 0.048$. Additivity under the connect sum operation was also observed by Katritch *et al.*²

6 Writhe as a function of link type

One can ask similar questions about the dependence of mean writhe on link type, though less work has been carried out in this area. Gee and Whittington¹³ have recently used Monte Carlo methods to estimate the writhe of a polygon in a two component link of fixed type, as a function of the numbers of edges in the two polygons. We shall discuss these results and compare with work by Vologodskii and Cozzarelli³ on torus links in a model of circular DNA.

Two component links can be symmetric or asymmetric. Colour the two circles red and blue. An unoriented link is *symmetric* if there is an ambient isotopy which interchanges the colours. That is, if there is a smooth deformation whereby the two circles in the link can be interchanged in space. $(2, 2k)$ -torus links are symmetric. A link with one component knotted and the other unknotted (e.g. 7_8^2) is asymmetric. If the link is symmetric and the two polygons composing the link in Z^3 have the same number of edges, then each component of the link must have the same mean writhe, and this is the simplest case to consider. Suppose we consider $(2, 2k)$ -torus links, where each polygon has n edges. Gee and Whittington¹³ estimated the mean writhe for $k = 2, 3, 4, 5$ for $n \leq 140$. For each value of k the mean writhe seems to be more or less independent of n (except at very small values of n) and one can estimate the mean writhe on the assumption that it is really n -independent, and that the apparent small dependence is a mixture of statistical fluctuations and a slow convergence to a limiting value. It then seems that the mean writhe increases roughly linearly with increasing k . Vologodskii and Cozzarelli³ observed similar behaviour in their model of circular DNA. Although the mean writhe values for the two models are different (for the same torus link), the mean writhe increases linearly with k in both cases, and the slope of the line is very similar.

| knot type | Janse van Rensburg <i>et al</i> | | Le Bret | Katritch <i>et al</i> | | |
|------------|---------------------------------|---------|---------|-----------------------|--------|--------|
| | average | minimal | | ideal | 5400bp | 1800bp |
| 3_1 | 3.44 | 3.44 | 3.5 | 3.41 | 3.43 | 3.41 |
| 5_1 | 6.25 | 6.25 | 6.3 | 6.26 | 6.30 | 6.21 |
| 5_2 | 4.55 | 4.49 | 4.7 | 4.54 | 4.56 | 4.59 |
| 6_1 | 1.16 | 0.93 | | 1.23 | 1.20 | 1.18 |
| 6_2 | 2.83 | 2.69 | | 2.70 | 2.83 | 2.76 |
| 7_1 | 9.07 | 9.28 | | 9.15 | 9.10 | 8.95 |
| 7_2 | 5.74 | | | 5.82 | 5.79 | 5.77 |
| 8_1 | 2.35 | | | 2.33 | 2.47 | 2.43 |
| 8_{19} | 8.73 | | | 8.64 | | |
| 9_1 | 11.89 | | | 12.07 | 12.0 | |
| 9_2 | 6.95 | | | 6.84 | 6.94 | |
| 10_1 | 3.55 | | | 3.47 | 3.60 | |
| 10_{161} | 9.30 | | | 9.45 | | |
| $3_1\#3_1$ | 6.91 | | | 6.81 | | |

Table 1: Estimates of mean writhe for various chiral knots.

This establishes that linking induces writhe and that the mean writhe increases (roughly linearly with k) within the family of $(2, 2k)$ -torus links, when the two circles are the same length. If the circles have different lengths then the writhe can depend on the length of a circle and on the length (or relative length) of its partner. For instance, if a 20-gon and a 40-gon are linked to form a $(2, 4)$ -torus link, the mean writhe of the 20-gon is 0.53 ± 0.03 while the mean writhe of the 40-gon is 3.35 ± 0.13 (so that the longer polygon has higher writhe, as found by Vologodskii and Cozzarelli³). More generally, consider a polygon with n edges linked to a polygon with m edges. Gee and Whittington¹³ found that, for a fixed torus link type, the mean writhe of an n -gon decreases with increasing m at fixed n and increases with increasing n at fixed m . At fixed n the writhe of an n -gon seems to be approaching a limiting non-zero value as m increases, and this limiting value depends both on n and on the type of torus link, increasing both with n and with k .

7 Discussion

Lattice models have played an important role in the development of our understanding of the conformational properties of polymers and, since the recent

interest in topological and geometrical entanglement in polymers, lattice models have proved important in these areas. The first proofs of the inevitability of knotting in sufficiently long ring polymers were constructed for lattice models^{14,15} and we still know much more about the lattice version of this problem¹⁶ than about its continuum analogues.^{17,18,19}

The results which we described in Section 2 show that the computation of the writhe of a simple closed curve in Z^3 is much simpler than the corresponding calculation for curves in R^3 . With some modifications this argument can be applied to other lattices such as the body-centred and face-centred cubic lattices. The result described in that section plays a key role in the proof of the theorem which we described in Section 3 which gives a lower bound on the rate of increase of $\langle |Wr| \rangle_n$.

The remainder of the paper is concerned with the mean writhe of circles of a fixed knot type or which are members of a link of two circles with specified link type. The mean writhe isn't strongly dependent on n and our estimates of the mean writhe as a function of knot type are very similar to the values found by Katritch *et al*¹ for ideal knots in R^3 . In addition we observed that mean writhe is approximately additive under the connect sum operation (see also the results in Katritch *et al*²), and this might be an exact relation. We also observed that mean writhe increases linearly with crossing number in certain knot families, and that there is a similar behaviour in $(2, 2k)$ -torus links.

For the moment the results described in Sections 5 and 6 are observations based on numerical data, and it would be interesting to establish these results rigorously.

Acknowledgements

The authors would like to thank Carla Tesi and Enzo Orlandini for numerous helpful conversations. This work was financially supported by NSF and by NSERC.

1. V. Katritch, J. Bednar, D. Michoud, R. Scharein, J. Dubochet and A. Stasiak, *Geometry and physics of knots* Nature **384** 142-145 (1996).
2. V. Katritch, W.K. Olson, P. Pieranski, J. Dubochet and A. Stasiak, *Properties of ideal composite knots* Nature **388** 148-151 (1997).
3. A.V. Vologodskii and N.R. Cozzarelli, *Monte Carlo analysis of the conformation of DNA catenanes* J. Mol. Biol. **232** 1130-1140 (1993).

4. R.C. Lacher and D.W. Sumners, *Data Structures and Algorithms for the Computation of Topological Invariants of Entanglements: Link, Twist and Writhe*, in *Computer Simulations of Polymers*, R.J. Roe, ed., Prentice-Hall, 365-373 (1991).
5. E.J. Janse van Rensburg, E. Orlandini, D.W. Sumners, M.C. Tesi and S.G. Whittington, *The writhe of a self-avoiding polygon*, *J. Phys. A: Math. Gen.* **26** L981-986 (1993).
6. H. Kesten, *The number of self-avoiding walks*, *J. Math. Phys.* **4** 960 (1963).
7. J.D. Moroz and R.D. Kamien, *Self-Avoiding Walks with Writhe*, Preprint.
8. N. Madras and G. Slade, *The Self-Avoiding Walk*, Birkhäuser 1993, Chapter 9.
9. E.J. Janse van Rensburg and S.G. Whittington, *The BFACF algorithm and knotted polygons* *J. Phys. A: Math. Gen.* **24** 5553-5567 (1991).
10. M. Le Bret, *Monte Carlo computation of the supercoiling energy, the sedimentation constant, and the radius of gyration of unknotted and knotted circular DNA* *Biopolymers* **19** 619-637 (1980).
11. E.J. Janse van Rensburg, E. Orlandini, D.W. Sumners, M.C. Tesi and S.G. Whittington, *The writhe of knots in the cubic lattice* *J. Knot Theory and its Ramifications*, **6** 31-44 (1997).
12. E.J. Janse van Rensburg and S.D. Promislow, *Minimal Knots in the Cubic Lattice* *J. Knot Theory and its Ramifications*, **4** 115-130 (1995).
13. M. Gee and S.G. Whittington, *Induced writhe in linked polygons* *J. Phys. A: Math. Gen.* **30** L1-5 (1997).
14. D.W. Sumners and S.G. Whittington, *Knots in self-avoiding walks* *J. Phys. A: Math. Gen.* **21** 1689-1694 (1988).
15. N. Pippenger, *Knots in random walks* *Discrete Appl. Math.* **25** 273-278 (1989),
16. C.E. Soteros, D.W. Sumners and S.G. Whittington, *Entanglement complexity of graphs in Z^3* *Math. Proc. Camb. Phil. Soc.* **111** 75-91 (1992).
17. Y. Diao, *On Knotting of Randomly Embedded n -gons in \mathcal{R}^3* Ph.D Thesis, FSU 1990.
18. Y. Diao, *Unsplittability of Random Links* *J. Knot Theory and its Ramifications*, **3** 379-389 (1994).
19. Y. Diao, N. Pippenger and D.W. Sumners, *On Random Knots* *J. Knot Theory and its Ramifications*, **3** 419-429 (1994).

CHAPTER 6

MINIMAL LATTICE KNOTS

E.J. JANSE VAN RENSBURG

*Department of Mathematics and Statistics, York University
4700 Keele street, North York, Ontario, M3J 1P3, Canada.*

Abstract

How many unit edges and right angles are needed to construct a knot of a given type in the cubic lattice. The unknot can be made using 4 edges and 4 right angles, since the girth of the cubic lattice is 4, and the curvature of a lattice polygon must be at least 2π (define curvature of a piecewise linear curve as the sum over the excluded angles between successive line segments). It is also known that the trefoil can be constructed in the cubic lattice using 24 edges (and not with any fewer), and that it will have at least 12 right angles. In this chapter I review the properties of minimal lattice knots, concentrating of the minimum length and minimum curvature for given knot types.

1. Introduction

How many edges are necessary and sufficient to tie a knot of a given type in the cubic lattice \mathcal{Z}^3 ? A trefoil will need 24 edges [1], and it is conjectured that a figure eight knot needs 30 edges [2]. In this chapter I will consider aspects of this question; in particular, I consider the behaviour of the minimal edge number and the minimal lattice curvature, and estimate these numerically. Related questions were considered in references [1,2,3,4,5,8]. Similar questions can be asked about knotted polygons in other lattices. It is thought that 16 edges are necessary and sufficient to tie the knot¹ 3_1 in the face-centered cubic lattice [7].

¹ Knots will be indicated by Alexander-Briggs notation [6]; for example, 3_1 is the trefoil and 4_1 is the figure eight knot.

A *polygon* in \mathcal{Z}^3 is an unlabeled vertex-avoiding closed path of edges. I define two polygons to be equivalent if they can be superimposed by a translation and/or a rotation in \mathcal{Z}^3 . The resulting equivalence classes are also called *polygons*. One may consider a polygon to be a tame embedding of the circle in \mathcal{R}^3 , and therefore its knot type is defined, and may be determined by computing knot invariants. There is a natural one-to-one correspondence between the knot classes of polygons and tame embeddings of the circle in \mathcal{R}^3 ; by subdivision, any tame knot can be arbitrarily closely approximated by a polygon with the same knot type [9].

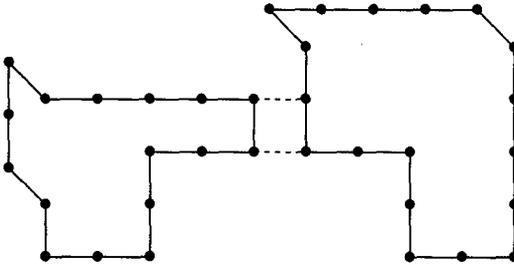


Figure 1: Concatenation of two polygons.

Concatenation is an operation which constructs a new polygon from two smaller polygons. A lexicographic ordering of the edges of a polygon P by the coordinates of their midpoints defines a *bottom edge*, as the lexicographic least edge, and a *top edge*, as the lexicographic most edge. It is simple to check that the bottom and the top edges of P are always perpendicular to the first direction.² A polygon P_1 can be concatenated with a polygon P_2 by translating P_1 , and rotating it about the $\hat{\mathbf{i}}$ -axis, if necessary, until the midpoint

² I define $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ to be the canonical unit vectors. The vector $\hat{\mathbf{i}}$ is the first direction, and the lexicographic ordering is done first in the $\hat{\mathbf{i}}$ direction, then the $\hat{\mathbf{j}}$ direction, and then the $\hat{\mathbf{k}}$ direction. The bottom and top edges are always in either the $\hat{\mathbf{j}}$ or $\hat{\mathbf{k}}$ directions.

of the bottom edge of P_2 has first coordinate exactly one more than the first coordinate of the midpoint of the top edge of P_1 (the other two coordinates are the same). By inserting two new edges between the end-vertices of the top edge of P_1 and the bottom edge of P_2 , and by deleting the top edge of P_1 and the bottom edge of P_2 , a new polygon P_3 is constructed. The list of coordinates of the vertices of P_3 is the union of the lists for P_1 and P_2 , and all the edges in P_1 and P_2 are in P_3 ; with the exception of the deleted top edge of P_1 and the deleted bottom edge of P_2 . The two new edges connect P_1 and P_2 into P_3 , concatenating the knots³; I illustrate this in figure 1.

In this chapter I will consider the minimal edge number of lattice knots, and the minimal curvature of a lattice knot. I will first consider some theoretical results before I resort to computer simulations to estimate the minimal edge number and the minimal curvature of lattice knots. In section 2 I explore the basic theoretical ideas before I consider numerical results in section 3. The paper is concluded in section 4.

2. Knot Complexity and the Minimal Number.

Let $\mathcal{M}(K)$ be the minimal edge number of the knot K , and let $\mathcal{C}(K)$ be the minimal lattice curvature of K . Note that contributions to $\mathcal{C}(K)$ is in units of $\pi/2$ (for each right angle in K). I first show that $\mathcal{M}(K)$ is subadditive with respect to concatenation (and therefore with respect to the connected sum of knots), and I explore the extent to which $\mathcal{M}(K)$ and $\mathcal{C}(K)$ are measures of knot complexity (see reference [9] for a comparison). Presumably, the more complicated a knot, the larger its minimal edge number.

Lemma 1: If K_1 and K_2 are knots, then

$$\mathcal{M}(K_1 \# K_2) \leq \mathcal{M}(K_1) + \mathcal{M}(K_2).$$

Thus, the minimal edge number is subadditive with respect to the connected sum of K_1 and K_2 .

³ Concatenation is usually applied to establish a sub-additive relation for the numbers of the objects involved. It will be slightly differently applied here: I will use it to show a sub-additive relation for the minimum size of the involved polygons.

Proof: Let P_1 and P_2 be minimal polygon representatives of K_1 and K_2 with $\mathcal{M}(K_1)$ and $\mathcal{M}(K_2)$ edges respectively. Concatenate P_1 and P_2 to get a polygon representation of $K_1\#K_2$ with $\mathcal{M}(K_1) + \mathcal{M}(K_2)$ edges. Hence, $\mathcal{M}(K_1\#K_2) \leq \mathcal{M}(K_1) + \mathcal{M}(K_2)$.

□

Application of lemma 1 gives an upper bound on the minimal edge number of compound knots in terms of the minimal edge number of their factors. The following lemma finds a lower bound by using the curvature of the knots. This also has some implications for the lattice curvature.

Lemma 2: If K_1 and K_2 are knots, then there exist constants $\alpha_K \geq 0$ for any knot K such that

$$\mathcal{M}(K_1\#K_2) \geq \alpha_{K_1} + \alpha_{K_2},$$

and

$$\mathcal{M}(K^n) \geq n\alpha_K,$$

where K^n is the connected sum of n knots, all of type K . If K is not the unknot, then $\alpha_K > 0$.

Proof: The *bridge number* of a projection of a knot K can be found by traversing the knot and recording a sequence of letters corresponding to over- and underpasses; one may suppose that an O indicates an overpass and a U indicates an underpass. Consider this sequence to be cyclic; the last letter is followed again by the first. The number of runs of O 's in the sequence is the bridge number of the projection. The *bridge index* of a knot K , $b(K)$, is the minimum bridge number over all projections and conformations of the knot [10,11]. It is known that the bridge index minus one is additive with respect to the connected sum of knots: if K_1 and K_2 are two tame knots, then $b(K_1) + b(K_2) - 1 = b(K_1\#K_2)$. Put $\alpha_K = 4(b(K) - 1)$. Then α_K is additive under the connected sum of knots. For every tame knot K in \mathcal{R}^3 , $b(K) = c(K)/2\pi$, where $c(K)$ is the infimum of the total curvature of K , as K varies within its knot type [11]. Let $C(K)$ be the curvature of a given knot K (then $c(K) \leq C(K)$). Then $C(K) \geq 2\pi b(K)$ for any given knot K . If a polygon in \mathcal{Z}^3 is considered, then contributions to the curvature occurs in units of $\pi/2$, since successive edges in the polygon can make angles of only 0 or $\pi/2$ with each other. Hence, if a polygon P has curvature $C(P)$, then P

has at least $2C(P)/\pi$ edges; one edge for every $\pi/2$ contribution to $C(P)$. We get $\mathcal{M}(K) \geq 2C(K)/\pi \geq 2c(K)/\pi = 4b(K) > \alpha_K$. From the additivity of α_K , $\mathcal{M}(K_1 \# K_2) \geq \alpha_{K_1 \# K_2} = \alpha_{K_1} + \alpha_{K_2}$. The second inequality is obtained similarly. □

In lemma 2, the constants α_K depend only on the knot type K , and there is it is enough, but not necessary, to take $\alpha_K = 4(b(K) - 1)$ in the above. The following theorem assigns a *minimal edge index* to every knot type.

Theorem 3: Let K be an arbitrary knot. Then the following limit exists, and it is the *minimal edge index* of K :

$$\lim_{n \rightarrow \infty} \frac{\mathcal{M}(K^n)}{n} = \mu_K.$$

In addition, $\mathcal{M}(K^n) \geq n\mu_K$. Moreover, $\mu_K > 0$ if K is not the unknot.

Proof: $\mathcal{M}(K^n)$ is subadditive with respect to the connected sum: by lemma 1 $\mathcal{M}(K^n \# K^m) \leq \mathcal{M}(K^n) + \mathcal{M}(K^m)$. By lemma 2, $\mathcal{M}(K^n)$ grows at least linearly in n if K is not the unknot; thus, applying a result from the theory of subadditive functions: $\mu_K = \lim_{n \rightarrow \infty} \mathcal{M}(K^n)/n \geq \alpha_K$ exists. Moreover, for every n , $\mathcal{M}(K^n) \geq n\mu_K$ [12]. If $K \neq \emptyset$, then $\alpha_K \neq 0$, and so $\mu_K > 0$. □

The minimal lattice curvature can be studied in the same way as $\mathcal{M}(K)$. Let K_1 and K_2 be two lattice knots with minimal lattice curvature $\mathcal{C}(K_1)$ and $\mathcal{C}(K_2)$. If K_1 and K_2 are concatenated, then at most 4 right angles can be removed when the bottom and top edges are deleted, and at most 4 right angles may be created when new edges are added to connect K_1 to K_2 . Therefore, the total number of right angles in K_1 and K_2 may change by a total of 8 (this is an overestimate). Thus $\mathcal{C}(K_1 \# K_2) \leq \mathcal{C}(K_1) + \mathcal{C}(K_2) + 4\pi$. In other words, $\mathcal{C}(K) + 4\pi$ is subadditive. On the other hand, it is known that $\mathcal{C}(K) \geq 6\pi$ [13]. By the proof of lemma 2 observe that $\mathcal{C}(K^n) \geq c(K^n) = 2\pi b(K^n) = 2\pi n b(K) - 2\pi(n - 1) \geq 2\pi n$ if K is not the unknot. The same arguments in the proof of theorem 3 gives:

Corollary 4: Let K be an arbitrary knot. Then the following limit exists, and it is the *minimal curvature index* of K :

$$\lim_{n \rightarrow \infty} \frac{\mathcal{C}(K^n)}{n} = \nu_K.$$

In addition, $\mathcal{C}(K^n) \geq n\nu_K$, and $\nu_K \geq 2\pi$ if K is not the unknot.

□

To what extent do $\mathcal{M}(K)$, $\mathcal{C}(K)$, μ_K and ν_K measure the “complexity” of a knot? A good measure of knot complexity [9] is any function \mathcal{F} defined on the equivalence classes of knots with the following properties: $\mathcal{F}(\emptyset) = 0$ (\emptyset is the unknot), and there exists a knot K such that for any knot L , $\mathcal{F}(K^n \# L) \geq n\mathcal{F}(K) > 0$ (n is an arbitrary integer). This last condition rules out any functions which are subadditive with respect to the connected sum of knots, including $\mathcal{M}(K)$ and $\mathcal{C}(K)$. Instead, define a *weak measure* of knot complexity [2] as a function \mathcal{F}_w defined on the equivalence classes of knots such that $\mathcal{F}_w(\emptyset) = 0$ and $\mathcal{F}_w(K^n \# L) \geq nA_K + A_L$, where $A_K \geq 0$ and $A_K = 0$ if and only if $K = \emptyset$. This definition suggests that $\mathcal{F}_w(K) = \mathcal{M}(K) - 4$ and $\mathcal{F}_w = \mathcal{C}(K) - 2\pi$ should qualify as weak measures of knot complexity. I state the properties of $\mathcal{M}(K)$ in the next theorem:

Theorem 5: $\mathcal{M}(K)$ has the following properties:

- (i) $\mathcal{M}(\emptyset) = 4$,
- (ii) $\mathcal{M}(K) > 4$ if K is not the unknot,
- (iii) $\mathcal{M}(K_1 \# K_2) \leq \mathcal{M}(K_1) + \mathcal{M}(K_2)$,
- (iv) and there exists constants A_K such that $\mathcal{M}(K_1^n \# K_2) > nA_{K_1} + A_{K_2}$ for any knot $K_1 \neq \emptyset$, K_2 an arbitrary knot, and any integer $n \geq 0$. Moreover, $A_K = 0 \Leftrightarrow K = \emptyset$.

Proof: (i) This follows from the fact that the unknotted polygon has shortest length 4. (ii) A theorem by Diao [1,3] states: $\mathcal{M}(K) = 24$ if and only if $K = 3_1$. Moreover, if $\mathcal{M}(K) < 24$, then K is the unknot. Hence, $\mathcal{M}(K) > 4$ if K is not the unknot. (iii) This was shown in lemma 1. (iv) $\mathcal{M}(K) \geq \alpha_K$ as shown

in the proof of lemma 2. Therefore, $\mathcal{M}(K_1^n \# K_2) \geq n\alpha K_1 + \alpha_{K_2}$. So take $A_K = \alpha_K$. Since $\alpha_K \geq 4$ if K is not the unknot, and $\alpha_K = 0$ if K is the unknot, the claimed bounds on A_K follows. □

Theorem 5 implies that $\mathcal{M}(K) - 4$ is a weak measure of knot complexity. A weak measure can distinguish the unknot, and it grows at least proportionally to the number of factors in a compound knot. $\mathcal{M}(K) - 4$ does not qualify as a good measure of knot complexity due to theorem 5(iii): it is subadditive, and it cannot satisfy the condition $\mathcal{F}(K^n \# L) \geq n\mathcal{F}(K) > 0$ for a knot K and any knot L .

Theorem 6: Let K be any piecewise linear knot. Then the minimal edge index of K has the following properties:

- (i) $\mu_K = 0 \Leftrightarrow K = \emptyset$,
- (ii) $\mu_{K^n} = n\mu_K$,
- (iii) $\mu_{K_1 \# K_2} \leq \mu_{K_1} + \mu_{K_2}$, where K_1 and K_2 are non-trivial piecewise linear knots.

(iv) There exists numbers \mathcal{A}_K such that if K_1 is a non-trivial knot and K_2 is an arbitrary knot, then $\mu_{K_1^n \# K_2} \geq n\mathcal{A}_{K_1} + \mathcal{A}_{K_2}$. In fact, $\mathcal{A}_K = 0 \Leftrightarrow K = \emptyset$ and $\mathcal{A}_K > 4$ if K is not the unknot.

Proof: (i) $\mathcal{M}(\emptyset^n) = 4$ thus $\mu_\emptyset = 0$. If $K \neq \emptyset$, then by lemma 2, $\mathcal{M}(K^n) \geq n\alpha_K$, where $\alpha_K > 0$. Divide by n and take the limit $n \rightarrow \infty$ to find $\mu_K > 0$. (ii) $\mu_{K^n} = \lim_{m \rightarrow \infty} \mathcal{M}((K^n)^m)/m = n \lim_{nm \rightarrow \infty} \mathcal{M}(K^{nm})/nm = n\mu_K$ by theorem 3. (iii) $\mu_{K_1 \# K_2} = \lim_{n \rightarrow \infty} \mathcal{M}((K_1 \# K_2)^n)/n \leq \lim_{n \rightarrow \infty} (\mathcal{M}(K_1^n) + \mathcal{M}(K_2^n))/n = \mu_{K_1} + \mu_{K_2}$ by lemma 1. (iv) By the methods of lemma 2 and theorem 4, $\mu_{K_1^n \# K_2} \geq \lim_{m \rightarrow \infty} (4nm c(K_1) + 4m c(K_2) - 2\pi(4nm + 4m))/(2m\pi)$ where $c(K)$ is the minimum curvature of knot K . But for any non-trivial knot, $c(K) > 4\pi$ [14], thus $\mu_{K_1^n \# K_2} \geq \lim_{m \rightarrow \infty} (8m(n+1) - 4m(n+1))/m = 4n + 4$. If $K_1 = K_2 = \emptyset$, then observe that $\mathcal{A}_{K_1} = 0$. Otherwise, $K_1 \neq \emptyset$, and $\mu_{K_1^n} \geq 4n$ so that $\mathcal{A}_{K_1} \geq 4 > 0$. □

Theorem 6(i) and (iv) implies that the minimal edge index is a weak measure of knot complexity. It is not a good measure of knot complexity in the sense of reference [9] because of theorem 6(iii). Theorem 6(ii) indicates that the minimal edge index is regular when applied to knots of the type K^n ; in fact, it measures the number of components in the compound knot. Moreover, this property suggests that theorem 6(iii) should be an equality (it is when $K_1 = K_2$), so the following conjecture seems natural:

Conjecture 7: Let K_1 and K_2 be arbitrary knots. Then $\mu_{K_1 \# K_2} = \mu_{K_1} + \mu_{K_2}$.

□

If conjecture 7 holds, then the minimal edge index should be a very powerful measure of knot complexity. To prove this seems hard: some detailed geometric information on exactly how the minimal knots change when they are concatenated are needed.

Similar results are found when $\mathcal{C}(K)$ and ν_K are considered instead. The arguments are very similar to those in the proofs of theorems 5 and 6, and I will only state the results; the interested reader can fill in the detail.

Theorem 8: $\mathcal{C}(K)$ has the following properties:

- (i) $\mathcal{C}(\emptyset) = 2\pi$,
- (ii) $\mathcal{C}(K) > 2\pi$ if K is not the unknot,
- (iii) $\mathcal{C}(K_1 \# K_2) \leq \mathcal{C}(K_1) + \mathcal{C}(K_2)$,
- (iv) and there exists constants A'_K such that $\mathcal{C}(K_1^n \# K_2) > nA'_{K_1} + A'_{K_2}$ for any knot $K_1 \neq \emptyset$, K_2 an arbitrary knot, and any integer $n \geq 0$. Moreover, $A'_K = 0 \Leftrightarrow K = \emptyset$.

□

And similarly:

Theorem 9: Let K be any piecewise linear knot. Then the minimal curvature index of K has the following properties:

- (i) $\nu_K = 0 \Leftrightarrow K = \emptyset$,
- (ii) $\nu_{K^n} = n\nu_K$,
- (iii) $\nu_{K_1 \# K_2} \leq \nu_{K_1} + \nu_{K_2}$, where K_1 and K_2 are non-trivial piecewise linear knots.
- (iv) There exists invariants $\mathcal{A}'_K > 0$ such that if K_1 is a non-trivial knot and K_2 is an arbitrary knot, then $\nu_{K_1 \# K_2} \geq n\mathcal{A}'_{K_1} + \mathcal{A}'_{K_2}$. In fact, $\mathcal{A}'_K = 0 \Leftrightarrow K = \emptyset$ and $\mathcal{A}'_K > 2\pi$ if K is not the unknot.

□

Thus, $\mathcal{C}(K) - 2\pi$ and ν_K are both weak measures of knot complexity, and **Conjecture 10:** Let K_1 and K_2 be arbitrary knots. Then $\nu_{K_1 \# K_2} = \nu_{K_1} + \nu_{K_2}$.

□

In other words, I conjecture that the lattice curvature index is additive with respect to the connected sum of lattice knots.

3. Numerical Results

In this section the BFACF algorithm [15] is implemented with simulated annealing to find upper bounds (and also to estimate) on the minimal edge numbers and minimal curvature of various lattice knots. The two elementary moves in the algorithm act locally on the polygons; the first is a length preserving move (figure 2(a)), and the second move may change the length of the polygon (figure 2(b)).



Figure 2: The two elementary moves of the BFACF algorithm

A simple Metropolis-style implementation [16] of the algorithm is as follows: Let ω be an arbitrary polygon with vertices denoted by $\omega_0, \omega_1, \omega_2, \dots, \omega_n$ if ω has length n . Pick an edge $\omega_{i-1}\omega_i$ in ω with uniform probability, and pick a unit vector \mathbf{u} perpendicular to the edge. Move the chosen edge parallel to itself in the direction of \mathbf{u} , inserting two new edges to keep the polygon connected. Erase any double edges which may be formed. It can be checked that this gives one of the two possible transitions in figure 2. Let the new polygon be ω' . If ω' is not self-avoiding, then this attempt is rejected, and ω is counted (again) as the next state. If ω' is shorter or of equal length than ω , then accept ω' as the next state. If ω' is longer than ω , then accept it as the next state with probability β^2 ; if it is not accepted, then ω is (again) counted as the the next state. β is a free parameter of the algorithm, increasing its value leads to the sampling of longer polygons, while reducing it bias the sampling in favour of shorter polygons. Simulated annealing [18] is implemented by slowly reducing the value of β to 0 during a simulation; while the shortest polygons and polygons with least curvature found are recorded. If the “cooling down” is done carefully enough, then good estimates of the least curvature and shortest polygons are found. Is it known that the algorithm is irreducible on the classes of lattice polygons with the same knot type this means that the polygons of a given knot type with least curvature and shortest length is always accessible to the algorithm [17].

3.1 Estimating the Minimal Number

Estimates of the minimal edge numbers for prime knots up to eight crossings in the standard knot tables, as well as a selected few knots with nine and ten crossings, are listed in table I. There seems to be some correspondence between $\mathcal{M}(K)$ and the minimal crossing number of knots: $\mathcal{M}(K)$ increases in generally with crossing number. Any polygon with n edges will have at most $n(n-3)/2$ crossings in any regular projection, thus a lower bound is $\mathcal{M}(K) \geq (3 + \sqrt{9 + 8C})/2$, if K has C crossings in a minimal projection. The knots with non-alternating minimal diagrams seem to be anomalously low in the table; they take fewer edges to construct than alternating knots with the same crossing number.

How many different knot types may have the same minimal edge number? Let $q(n)$ be the number of knot types with minimal edge number n , and let $Q(n) = \sum_{m \leq n} q(m)$. Then $Q(n)$ grows at most exponentially:

TABLE I: Estimates of $\mathcal{M}(K)$

| $\mathcal{M}(K)$ | K |
|------------------|---|
| 4 | 0_1 |
| 24 | 3_1 |
| 30 | 4_1 |
| 34 | 5_1 |
| 36 | 5_2 |
| 40 | 6_1 6_2 6_3 |
| 42 | 8_{19} |
| 44 | 7_1 7_3 7_4 7_7 8_{20} |
| 46 | 7_2 7_5 7_6 8_{21} |
| 48 | 8_3 8_7 9_{42} |
| 50 | 8_1 8_2 8_4 8_5 8_6 8_9 8_{13} 8_{16} 9_{47} 10_{161} |
| 52 | 8_8 8_{10} 8_{11} 8_{12} 8_{14} 8_{15} 8_{17} 8_{18} |
| 54 | 9_1 |
| 56 | 9_2 |
| 60 | 10_1 |

Theorem 10: There exists a constant \mathcal{Q} such that

$$\lim_{n \rightarrow \infty} [\log Q(n)]/n = \log \mathcal{Q}.$$

Moreover, $Q(n) \leq \mathcal{Q}^n$.

Proof: Let n be even, and let $q(n)$ and $Q(n)$ be defined as above. Since the number of polygons rises exponentially with n [19], $q(n)$ rises at most exponentially with n . Also, $Q(n)Q(m) \leq Q(n+m)$, since every knot of type K with $\mathcal{M}(K) = n$ can be concatenated with a knot of type K' with $\mathcal{M}(K') = m$ to give a compound knot with $\mathcal{M}(K \# K') \leq (n+m)$. But this does not count prime knots K with $\max\{n, m\} < \mathcal{M}(K) \leq (n+m)$. Consequently, $Q(n)$ is a supermultiplicative function. Since $q(n)$ is bounded exponentially, so is $Q(n)$. Thus, we conclude that $\lim_{n \rightarrow \infty} [\log Q(n)]/n = \log \mathcal{Q}$ exists. Moreover, $Q(n) \leq \mathcal{Q}^n$ [12].

□

Since $q(n) = Q(n) - Q(n-1)$, it follows from theorem 10 that

$$\lim_{n \rightarrow \infty} [\log q(n)]/n = \log \mathcal{Q}.$$

Lower bounds on \mathcal{Q} can easily be read from table I. If chiral knots are counted as the same, and if compound knots (see table III) are ignored, then $\mathcal{Q} \geq 1.047$. An immediate consequence of theorem 10 is:

Corollary 11: The number of distinct knot types with minimal edge number equal to n rises exponentially with n .

□

There are other interesting patterns in table I. Consider the sequence of knots 3_1 , 5_1 , 7_1 and 9_1 which is a sequence of torus knots of the type $T_{2,k}$ for $k = 3, 5, 7, 9$. The minimal edge numbers of these knots increase as 24, 34, 44 and 54, exactly in steps of 10 edges. However, this pattern likely will not persist indefinitely. If strands in the knot twist about each other as in figure 3, then it appears as if 10 each are needed to complete each twist, except that 2 extra edges will be required every eighth twist, for an average of 10.25 edges per twist. This pattern could explain the increments in steps of 10 in this sequence of torus knots, but I think that it is unlikely to persist.

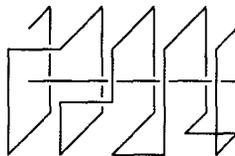


Figure 3: Twisting two strands.

Other interesting sequences in table I involve the twist knots, where two separate sequences can be identified: 4_1 , 6_1 , 8_1 , 10_1 , and 5_2 , 7_2 and 9_2 . In both sequences the increases are again in units of 10 edges, and a similar situation to the one above may be imagined, using the basic unit in figure 3 to add extra twists between two strands. 3_1 should be the first member in the sequence starting with 5_2 , but it appears that it does not fit the pattern. It is again hard to make definitive statements: to do so requires detailed knowledge of the geometric properties of the minimal knots in the lattice.

TABLE II: Estimates of $\mathcal{C}(K)$ (Units of $\pi/2$)

| $\mathcal{M}(K)$ | K |
|------------------|---|
| 4 | 0_1 |
| 12 | 3_1 |
| 14 | 4_1 |
| 17 | 5_1 5_2 |
| 18 | 6_2 6_3 9_{47} |
| 19 | 6_1 7_7 8_{20} |
| 20 | 7_4 7_6 8_{18} 8_{19} 8_{21} |
| 21 | 7_2 8_8 8_{13} 8_{17} 9_{42} |
| 22 | 7_1 7_3 8_2 8_5^* 8_7^* 8_9 8_{11} 10_{161}^* |
| 23 | 7_5 8_4 8_6 8_{10} 8_{23} 8_{16} |
| 24 | 8_1^* 8_3 8_{12}^* 8_{15} |
| 26 | 9_2 |
| 27 | 9_1 |
| 29 | 10_1 |

3.2 Minimal Curvature

The minimal curvature of lattice knots were estimated by simulated annealing. The results are displayed in table II. In most cases the minimal curvature was found on knots of minimal length; I indicate the exceptional cases with a *. As with the minimal edge number, the minimal curvature increases generally with increasing crossing numbers; there seems to be little, or no connection to the infimum of the curvature for these knots when realised in \mathcal{R}^3 .

In table II the torus knots $T_{2,2k+1}$ increases in minimal curvature in steps of 5; this pattern can again be explained by figure 3. The sequences of twist knots $4_1, 6_1, 8_1, 10_1$ similarly increases in steps of 5. On the other hand the sequence $3_1, 5_2, 7_2, 9_2$ increases in units of 5, except for the increment from 5_2 to 7_2 . The non-alternating knots appear early in the table, having relatively low curvature. The knot 9_{42} has very low curvature compared to other knots with minimal crossing number 9. Will knots with arbitrary high crossing number appear with low minimal lattice curvature? Since exactly $2\mathcal{C}(K)/\pi$ line segments appear in a lattice knot with curvature $\mathcal{C}(K)$, there can be at most $\mathcal{C}(K)(2\mathcal{C}(K)/\pi - 3)/\pi$ crossings in any regular projection of the lattice knot. Thus, if a knot has minimal crossing number C , then its minimal lattice curvature is at least $(3 + \sqrt{9 + 8C})\pi/4$.

3.3 Bounds on the minimal edge index and the minimal curvature index

Any vertex in the cubic lattice is at the barycentre of a dual 3-cube. The *lattice neighbourhood* of a (self-avoiding) walk is the union of all 3-cubes dual to the vertices of the walk. Let ω be a walk with first and last vertices the lexicographic least and most vertices respectively. Let $N(\omega)$ be the lattice neighbourhood of ω . Let ω' be the walk derived from ω by deleting its first and last steps. If $N(\omega')$ is a 3-ball, and if the pair $(N(\omega'), N(\omega') \cap \omega)$ is knotted, then ω is a knotted ball-pair or a *tight knot*. A tight knot ω has the *Kesten property* if three copies of $N(\omega') \cap \omega$ can be concatenated endpoint-to-endpoint to produce a tight knot which contains the original tight knot exactly three times [20]. The knot type of a tight knot is determined by taking rays from its endpoints (on the boundary of the 3-ball) to infinity, giving a knot in S^3 . The following lemma was proven in [9]:

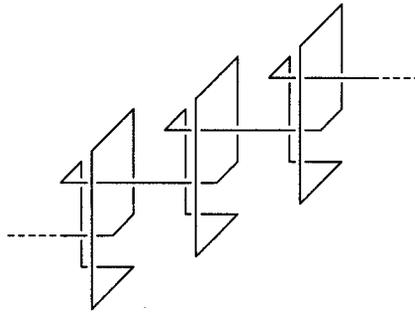


Figure 4: Tight trefoils in a string.

Lemma 12: Let K be any knot, then there exists a tight knot with the Kesten property which has knot type K .

□

Theorem 13: Let ω be a tight knot of type K with $|\omega|$ edges. If ω has the Kesten property, then $\mu_K \leq |\omega| - 1$. Moreover, if ω has C right angles, then $\nu_K \leq \pi C/2$.

Proof: Let ω' be defined as above. If ω has the Kesten property, then we may concatenate three copies of $N(\omega') \cap \omega$ in a row, as illustrated in figure 4. Adding a fourth copy to the third gives a new sequence of three, starting from the second copy, and this process can be continued indefinitely to find a sequence of tight knots in the lattice. Continue this process until N copies of $N(\omega') \cap \omega$ are included. Take four copies of this string, arranged in a square, where (say) m edges are added to join them into a polygon. The resulting polygon has $4(|\omega| - 1)N + m$ edges, since the length of each copy of $N(\omega') \cap \omega$ is $|\omega| - 1$. Divide by $4N$ and let $N \rightarrow \infty$ to find the upper bound. Observe that there are $4N$ copies of ω in this polygon, with non-overlapping corners, for a total of $4CN$ right angles. Thus, by taking $N \rightarrow \infty$, $\nu_K \leq \pi C/2$.

□

In figure 4 a string of tight trefoils is shown; each copy has 17 edges, representing 3_1 as a tight knot. Since there are 9 corners in each component, $\mu_{3_1} \leq 17$ and $\nu_{3_1} \leq 9\pi/2$. In order to test these predictions, I estimated $\mathcal{M}(3_1^n)$ and $\mathcal{C}(3_1^n)$ for n taking values 1 through n . The results are displayed in table III where $\mathcal{C}(3_1^n)$ is stated in units of $\pi/2$.

Theorem 3 and Corollary 4 indicates that the data in Table III may be used to compute strict upper bounds on μ_{3_1} and ν_{3_1} . The best bound on μ_{3_1} is obtained for $n = 8$, which gives $\mu_{3_1} \leq 17$; this is as good as the bound obtained from theorem 13. Similarly $\nu_{3_1} \leq 7.36\pi/2$, also obtained when $n = 8$. The increase in $\mathcal{M}(K)$ with n is in steps of 16 until $n = 8$; this bound is much better than the 9 obtained from figure 4. I illustrate the minimal 3_1^8 in figure 5; note the occurrences of “knots-of-knots”; these make effective use of right angles in the knot, producing a better lower bound. By corollary 4, $\nu_K \geq 2\pi$; is it true that $\nu_{3_1} > 2\pi$?

There are other interesting patterns visible in table III, for example, there are increases of 48 in the minimal edge number in the compounded left-handed trefoils 3_1^n for $n = 2, 5, 8$; it was stated in reference [2] that this spacing is 50, but more extensive simulations have improved on those results. For $n = 3, 5, 7$ the spacing is 32 and for $n = 1, 4, 7$ the spacing is 48 again.

4. Discussion

The results in this chapter is closely linked to work on the *stick number*

TABLE III: $\mathcal{M}(3_1^n)$ and $\mathcal{C}(3_1^n)$ $[\pi/2]$

| n | $\mathcal{M}(3_1^n)$ | $\mathcal{C}(3_1^n)$ |
|-----|----------------------|----------------------|
| 1 | 24 | 12 |
| 2 | 40 | 20 |
| 3 | 56 | 28 |
| 4 | 72 | 36 |
| 5 | 88 | 41 |
| 6 | 104 | 49 |
| 7 | 120 | 57 |
| 8 | 136 | 59 |
| 9 | 154 | 72 |
| 10 | 176 | 81 |

of a knot [8,21]: How many unit length line segments are necessary to realise a piecewise linear knot of type K (by joining the segments end-to-end into a polygon)? It is known that the stick number of a trefoil is 6 [8], and that the stick number of a figure eight knot is at least 7 [21].

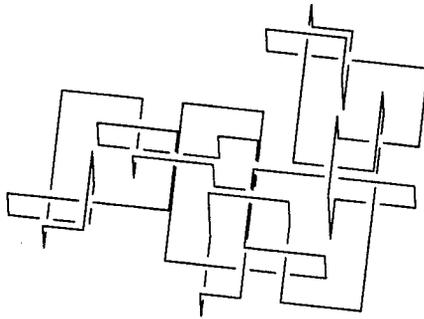


Figure 5: The shortest version of 3_1^8 found by simulated annealing.

From the perspective of the cubic lattice, it seems natural to define ideal conformations of lattice knots as those with minimal curvature. This minimizes the number of corners in a realisation of the knot as a lattice polygon (but does not necessarily minimize its edge number). It is not obvious that there is

always a conformation of a minimal lattice curvature which is also of minimal edge length, for any given knot K . The results in table II suggest that in some cases, lattice knots with minimal curvature cannot have minimal edge number; I suspect that for complicated knots this will be the rule, rather than the exception. On the other hand, one may want to minimise the curvature of lattice knots with minimal edge number to define those as “ideal lattice knots”.

The results in tables I, II and III classifies knots according to a measure of complexity defined by minimal edge numbers or curvature; more complicated knots appear lower in those tables. This observation may be compared to experimental work on knots in DNA [22], where it is found that no knots beyond a certain complexity will appear in a piece of DNA of given length. DNA has a natural stiffness, and it is energetically unfavourable to tie knots with (for example) high curvature in it (for simulation results, see reference [7]).

Acknowledgements:

This work was supported by a grant from NSERC of Canada.

References

- [1] Y. Diao, *Minimal Knotted Polygons on the Cubic Lattice*, J. of Knot Theory and its Ramifications, **2** (1993) 413-425.
- [2] E.J. Janse van Rensburg and S.D. Promislow, *Minimal Knots in the Cubic Lattice*, J. of Knot Theory and its Ramifications, **4** 115-130 (1995).
- [3] Y. Diao, *The Number of Smallest Knots on the Cubic Lattice*, J. Stat. Phys., **74** (1994) 1247-1254.
- [4] E.J. Janse van Rensburg, *Lattice Invariants for Knots*, in *Mathematical Approaches to Biomolecular Structure and Dynamics*, (eds. J.P. Mesirov, K. Schulten and D.W. Sumners), Proc. IMA Summer Prog. in Mol. Biol., 11-20 (Springer-Verlag, New York, 1995).
- [5] R. Uberti, E.J. Janse van Rensburg, E. Orlandini, M.C. Tesi and S.G. Whittington, *Minimal Links in the Cubic Lattice*, in *Topology and Geometry in Polymer Science*, (eds. S.G. Whittington, D.W. Sumners and T. Lodge), Proc. IMA 1995-96 Prog. on Math. Meth. in Material Science, Workshop 8 (1996).

- [6] J. W. Alexander and G. B. Briggs, *On Types of Knotted Curves*, Ann. of Math. (2), **28** (1927) 562-586.
- [7] E. J. Janse van Rensburg and S. G. Whittington, *The Knot Probability in Lattice Polygons*, J. Phys. A: Math. and Gen., **23** (1990) 3573-3590.
- [8] R. Randell, *A Molecular Conformation Space*, Proc. of the Int. Course and Conf. on Interfaces between Math., Chem. and Comp. Sci., **54** (1987) 125-140. R. Randell, *Conformation Spaces of Molecular Rings*, Proc. of the Int. Course and Conf. on Interfaces between Math., Chem. and Comp. Sci., **54** (1987) 141-156.
- [9] C. E. Soteros, D. W. Sumners and S. G. Whittington, *Entanglement Complexity of Graphs in \mathcal{Z}^3* , Math. Proc. Camb. Phil. Soc., **111** (1992) 75-91.
- [10] H. Schubert, *Über eine Numerische Knoteninvariante*, Math. Z. **61** (1954) 245-288.
- [11] D. Rolfsen, *Knots and Links*, Math. Lect. Ser., **7** (1990) (Publish or Perish, Inc.: Houston, Texas).
- [12] E. Hille, *Functional Analysis and Semi-Groups*, Am. Math. Soc. Colloq. Publ., **31** (1948) (American Mathematical Society: New York).
- [13] E.J. Janse van Rensburg and S.D. Promislow, *The Curvature of Lattice Knots*, Preprint.
- [14] J. Milnor, *On the total curvature of closed space curves*, Math. Scand., **1** (1953) 289-296.
- [15] B. Berg and D. Foester, *Random Paths and Random Surfaces on a Digital Computer*, Phys. Lett., **106B** (1981) 323-326. C. Aragao de Carvalho, S. Caracciolo and J. Frölich, *Polymers and $g|\phi|^4$ Theory in Four Dimensions*, Nucl. Phys. B [FS7], **215** (1983) 209-248.
- [16] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, *Equation of State Calculations by Fast Computing Machines*, J. Chem. Phys., **21** (1953) 1087-1092.
- [17] E.J. Janse van Rensburg and S.G. Whittington, *The BFACF Algorithm and Knotted Polygons*, J. Phys. A: Math. Gen., **24** (1991) 5553-5567.
- [18] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, *Optimization by Simulated An-*

nealing, Science, **220** no. 4598 (1983) 671-680.

- [19] J.M. Hammersley, *The Number of Polygons on a Lattice*, Proc. Camb. Phil. Soc., **57** (1961) 516-523.
- [20] D.W. Sumners and S.G. Whittington, *Knots in Self-Avoiding Walks*, J. Phys. A: Math. Gen. **21** (1988) 1689-1694.
- [21] K.C. Millett, *Random knotting of regular polygons*, Talk given at the Int. Joint Meetings of the AMS-CMS-MAA (August 1993), Vancouver. R. Randell, *Polygonal Knots*, Talk given at the Int. Joint Meetings of the AMS-CMS-MAA (August 1993), Vancouver.
- [22] S.Y. Shaw and J.C. Wang, *Knotting of a DNA chain during ring closure*, Science **260** (1993) 533-535.

CHAPTER 7

MINIMAL EDGE PIECEWISE LINEAR KNOTS

J. A. CALVO

*Department of Mathematics, University of California,
Santa Barbara, CA 93106, USA*

K. C. MILLETT

*Department of Mathematics, University of California,
Santa Barbara, CA 93106, USA*

The space of n -sided polygons embedded in three-space consists of a smooth manifold in which points correspond to piecewise linear or “geometric” knots, while paths correspond to isotopies which preserve the geometric structure of these knots. Two cases are considered: (i) the space of polygons with varying edge length, and (ii) the space of equilateral polygons with unit-length edges. In each case, the spaces are explored via a Monte Carlo search to estimate the distinct knot types represented. Preliminary results of these searches are presented. Additionally, this data is analyzed to determine the smallest number of edges necessary to realize each knot type with nine or fewer crossings as a polygon, *i.e.* its “minimal stick number.”

1. Introduction, vocabulary, and history of geometric knots.

The topological and geometric knotting of circles occurs in many contexts in the natural sciences.^{17,25} By geometric knotting we mean the imposition of geometric constraints on allowed configurations and their transformations. These constraints can arise by taking into consideration local “stiffness” of molecular structures such as DNA or other polymers. An attractive structure providing a useful model is that of the spatial polygon. These polygonal configurations are determined by a list of n points in three-space, which we call the *vertices* of the polygon. Straight line segments, or *edges*, connect each successive pairs of vertices, including the first and last one, producing a closed loop. When this configuration is embedded, so that there are no intersections of edges except at common vertices, one has a polygonal knot. The entire collection of such knots determines an open subset of euclidean space whose dimension is three times the number of vertices. Requiring that all vertices lie within the unit cube, that each edge have unit length, or that the angle between adjacent edges be constrained will determine other knot spaces and knot theories of interest.

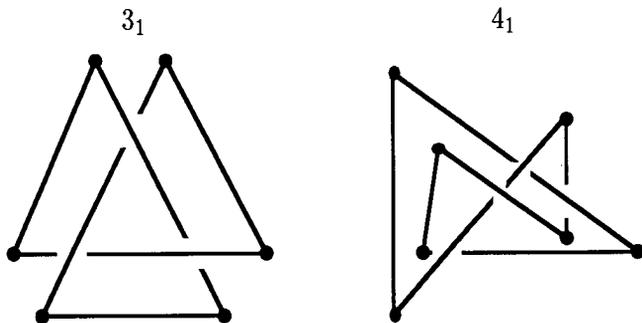


Figure 1: A hexagonal trefoil knot and a heptagonal figure eight knot.

1.1. Minimal stick number

An n -sided *spatial polygon* P in \mathbb{R}^3 is a closed, piecewise linear loop with no self-intersections consisting of n points of \mathbb{R}^3 , called *vertices*, joined by n straight line segments, called *edges*. We think of an n -gon as the result of glueing n sticks end to end to end. Define the *minimal stick number* $s(K)$ of a topological knot type K as the smallest number of edges required to realize K as a knotted polygon.^{1,23} It takes at least six sticks to construct a knotted polygon. A trefoil can be built with six sticks, while at least seven are required to build a figure-eight knot.^a Thus $s(3_1) = 6$ and $s(4_1) = 7$. Figure 1 shows projections of a hexagonal right-handed trefoil and a heptagonal figure-eight knot. In addition, every five and six crossing prime knot $(5_1, 5_2, 6_1, 6_2, 6_3)$, the square and granny knots $(3_1 \pm 3_1)$, the $(3, 4)$ -torus knot (8_{19}) , and the knot 8_{20} can all be built using eight sticks. Figure 2 shows octagonal realizations of these knots. Since only the trefoil and the figure eight can be constructed with fewer edges, all of these knots have stick number $s(\cdot) = 8$. However, it remains an open question whether this is a complete list of the eight-stick knots. Towards this goal, Calvo⁵ rules out every possibility other than 8_{18} , which has a minimal stick number of either 8 or 9. In addition to those knots in Figures 1 and 2, all of the seven crossing prime knots $(7_1, \dots, 7_7)$, as well as knots 8_{16} , 8_{17} , 8_{21} , 9_{40} , and 9_{41} are known to have nine-stick realizations, showing that these knots have minimal stick number $s(\cdot) = 9$.^b

Formulae for stick number are known only for a couple of families of knots.

^a For instance, see Proposition 1.3 in Millett.¹⁸ In pp.266–268, this proposition is followed not only by pictures, but also by coordinates of the vertices of equilateral realizations of these knots.

^b The construction of these nine-stick realizations are due to Monica Meissen^{15,16} and Robert Scharein²⁴.

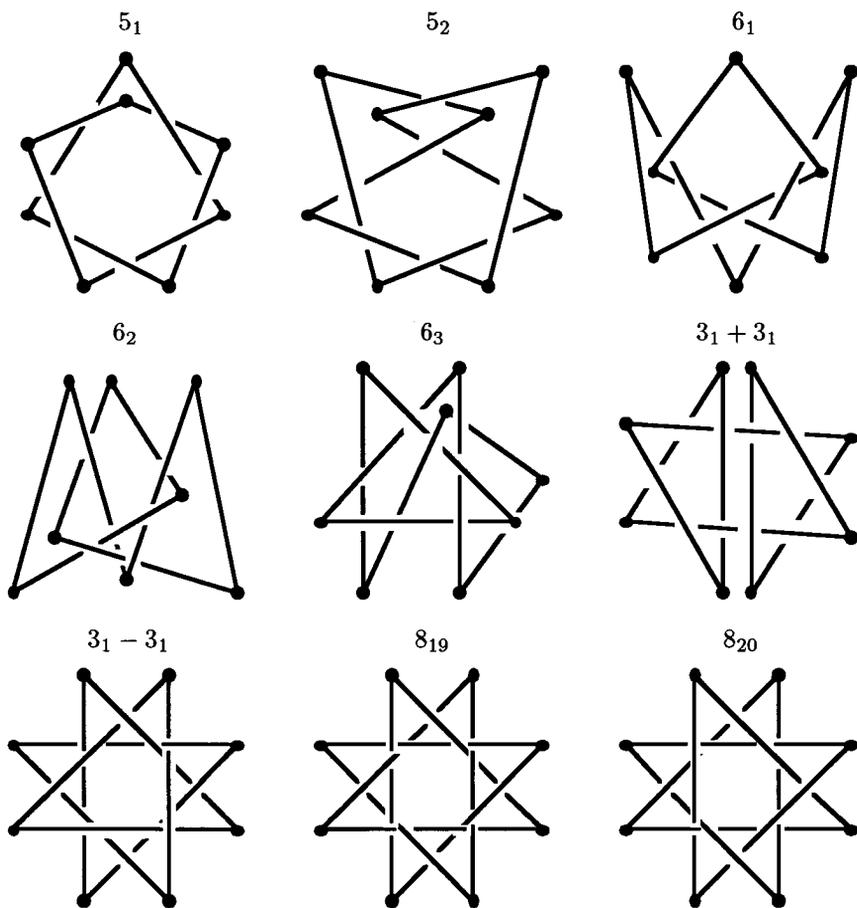


Figure 2: Octagonal knots.

Firstly, if p and q are coprime integers with $2 \leq p < q < 2p$, the stick number of the (p, q) -torus knot $T_{p,q}$ is

$$s(T_{p,q}) = 2q \quad (1)$$

(Theorem 7 in Jin¹¹). Note that this shows that $s(3_1) = s(T_{2,3}) = 6$ and that $s(8_{19}) = s(T_{3,4}) = 8$. Secondly, the connected sum of any combination of n right- and left-handed trefoils has stick number

$$s(3_1 \pm 3_1 \pm \cdots \pm 3_1) = 2n + 4 \quad (2)$$

(Theorem 7.1 in Adams *et al*²). Thus the square and granny knots have $s(3_1 + 3_1) = s(3_1 - 3_1) = 8$. This is an improvement on the general case of a connected sum, in which

$$s(K_1 + K_2) \leq s(K_1) + s(K_2) - 3 \quad (3)$$

(Theorem 3.1 in Adams *et al*²).

Relatively little more is known about stick number. Negami¹⁹ shows that given a nontrivial knot K with crossing number $c(K)$,

$$\frac{5 + \sqrt{9 + 8c(K)}}{2} \leq s(K) \leq 2c(K). \quad (4)$$

Here, the upper bound is obtained using results in graph theory, while the lower bound is found by projecting an n -sided polygon onto a plane perpendicular to one of the edges. The result is an $(n - 1)$ -sided polygonal knot diagram having at most $c = \frac{1}{2}n(n - 3)$ crossings. Completing the square and solving for n then gives the inequality in (4). Note that the trefoil knot is the only known example for which the upper bound is tight. In fact, Furstenberg *et al*⁹ show that if K is a knot with a one-, two-, or three-integer Conway notation and $c(K) > 5$, then this bound can be improved to

$$\frac{5 + \sqrt{9 + 8c(K)}}{2} \leq s(K) \leq c(K) + 2. \quad (5)$$

On the other side of the spectrum, Jin¹¹ uses Kuiper's superbridge index $sb(K)$ to obtain the lower bound

$$2sb(K) \leq s(K). \quad (6)$$

The superbridge index $sb(K)$ is the minimum over all embeddings of K of the largest number of local maxima obtained when projecting the knot in any direction in \mathbb{R}^3 (see Kuiper¹²). Furstenberg *et al*⁹ point out that no bound on stick number $s(K)$ gotten from the superbridge index $sb(K)$ can ever be very

efficient. A case in point is the family of two-bridge knots which, by (4), have arbitrarily large stick number but whose superbridge index is bounded above by seven: $sb(K) \leq 7$. Nonetheless, (6) can lead to some interesting bounds. For instance, if $2 \leq p < q$ then the (p, q) -torus knot has $sb(T_{p,q}) = \min\{2p, q\}$ (Theorem B in Kuiper¹²). A systematic construction of polygonal realizations of torus knots then shows that, if $2 \leq p < q$, then

$$2 \min\{2p, q\} \leq s(T_{p,q}) \leq p \left\lceil \frac{2q}{p} \right\rceil, \quad (7)$$

where the “ceiling brackets” denote rounding up, so $\lceil x \rceil = \min\{n \in \mathbb{Z} : n \geq x\}$ (Corollary 5 and Theorem 8 in Jin and Kim¹⁰). Notice that in the special case when $p = 2r + 1$ and $q = 3r + 1$, we have $2 \min\{2(2r + 1), 3r + 1\} = 6r + 2$ while $(2r + 1) \lceil \frac{2(3r+1)}{2r+1} \rceil = 6r + 3$. Therefore,

$$6r + 2 \leq s(T_{2r+1,3r+1}) \leq 6r + 3 \quad (8)$$

for any positive integer r (Corollary 9 in Jin and Kim¹⁰).

Although here $s(\cdot)$ is defined in the general setting of polygons with arbitrary edge lengths, similar notions of minimal stick number exist for more special sorts of “geometric knots.” For instance, one might restrict attention to polygons with unit-length edges, with vertices on the integral lattice \mathbb{Z}^3 , with restricted vertex angles, or with vertices on the unit-radius sphere about the origin.^c In this way, stick number might well depend upon the specific type of geometric knot under consideration. For example, Diaó⁷ has shown that the trefoil knot requires 24 edges for its vertices to lie on the lattice \mathbb{Z}^3 and its edges to have unit length. Later, we shall give special attention to equilateral polygons and consider the minimal *equilateral* stick number $s'(K)$ of a topological knot type K . At this time, however, there are no knot types known to have equilateral stick numbers which are different from their standard stick number.

Question 1. How many sticks are required to construct a knot K ? In particular, what are the stick numbers for all knots with, say, nine crossings or fewer? Does this depend on whether we use unit-length edges or not?

^c See Adams *et al.*²

1.2. The space of geometric knots

The general framework for the space of geometric knots was introduced by Randell.^{21–22} Consider an n -sided polygon P in \mathbb{R}^3 , together with a distinguished vertex, or *root*, v_1 and a choice of orientation. We can view P as a point of \mathbb{R}^{3n} by listing the triple of coordinates for each of its n vertices, starting with v_1 and proceeding in sequence as determined by the orientation.

In the spirit of Vassiliev,^{3,26} define the *discriminant* $\Sigma^{(n)}$ to be the collection of points in \mathbb{R}^{3n} which correspond in this way to non-embedded polygons. A polygon fails to be embedded in \mathbb{R}^3 when two or more of its edges intersect, so $\Sigma^{(n)}$ is the union of the closure of $\frac{1}{2}n(n-3)$ real semi-algebraic cubic varieties, each consisting of polygons with a given pair of intersecting edges.^d For example, the collection of polygons $\langle v_1, v_2, \dots, v_{n-1}, v_n \rangle$ for which v_1v_2 intersects v_3v_4 is the closure of the locus of the system

$$\begin{aligned}(v_2 - v_1) \times (v_3 - v_1) \cdot (v_4 - v_1) &= 0 \\ (v_2 - v_1) \times (v_3 - v_1) \cdot (v_2 - v_1) \times (v_4 - v_1) &< 0 \\ (v_4 - v_3) \times (v_1 - v_3) \cdot (v_4 - v_3) \times (v_2 - v_3) &< 0.\end{aligned}$$

In particular, the closure of each of these semi-algebraic varieties forms a codimension-1 submanifold (with boundary) of \mathbb{R}^{3n} . Hence the subspace

$$\mathfrak{Geo}^{(n)} = \mathbb{R}^{3n} - \Sigma^{(n)}$$

corresponding to embedded polygons is an open $3n$ -manifold which we will call the *embedding space* of rooted oriented n -sided geometric knots.

A path $h : [0, 1] \rightarrow \mathfrak{Geo}^{(n)}$ corresponds to an isotopy of polygonal simple closed curves, so each path-component of $\mathfrak{Geo}^{(n)}$ contains polygons of the same topological knot type. If two polygons lie in the same path-component of $\mathfrak{Geo}^{(n)}$, we will say they are *geometrically equivalent*. Also, a polygon is *geometrically unknotted* if it is geometrically equivalent to a standard planar polygon; since all planar n -sided polygons are geometrically equivalent, the component of geometric unknots is well-defined.

The geometric equivalence of two knots implies their topological equivalence. However, not much is known about the converse. For instance, it is unknown whether there exist topological unknots which are geometrically knotted. In fact, until recently there were no known examples of any topological knot type corresponding to two distinct geometric knot types.^e

^d If $n = 3$, $\Sigma^{(3)}$ is the collection of triangles $\langle v_1, v_2, v_3 \rangle$ for which $(v_2 - v_1) \times (v_3 - v_1) = 0$.

^e See Millett¹⁸ p.265, and compare with Theorem 1 in ⁴

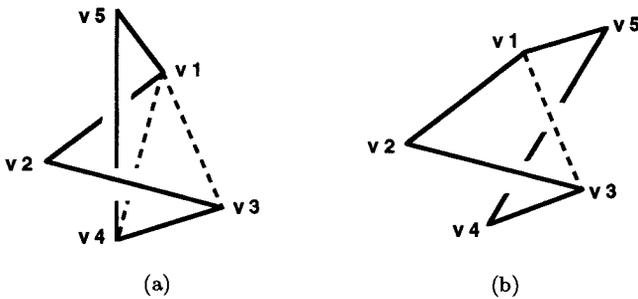


Figure 3: All pentagons are geometric unknots.

Question 2. How many distinct geometric (or topological) knot types are there in $\mathcal{G}\mathcal{E}\mathcal{O}^{(n)}$ as a function of n ?

Question 3. What can be said about the topology of the components of these knot spaces?

A classical theorem of Whitney²⁷ guarantees that, for each n , there are only finitely many geometric knot types. Furthermore, it is a “folk theorem” that $\mathcal{G}\mathcal{E}\mathcal{O}^{(n)}$ consists of a single path-component when $n \leq 5$. Since triangles are planar, the embedding space $\mathcal{G}\mathcal{E}\mathcal{O}^{(3)}$ of rooted oriented triangles is connected. A quadrilateral (tetragon) consists of two triangles hinged along a common edge; since we can change the dihedral angle at the hinge to flatten the quadrilateral out, we find that $\mathcal{G}\mathcal{E}\mathcal{O}^{(4)}$ is also connected. Finally, suppose that $P = \langle v_1, v_2, v_3, v_4, v_5 \rangle$ is some pentagon. If the edge v_4v_5 intersects the triangular disc determined by vertices v_1, v_2 , and v_3 , then P can be deformed by an isotopy of the linkage $v_4v_5v_6$ across the disc determined by v_4, v_5 , and v_6 until it coincides with the quadrilateral $\langle v_1, v_2, v_3, v_4 \rangle$ (see Figure 3a). On the other hand, if the edge does not intersect that triangle, then P can be deformed by an isotopy of $v_1v_2v_3$ across the triangular disc determined by v_1, v_2 , and v_3 until it coincides with the quadrilateral $\langle v_1, v_3, v_4, v_5 \rangle$ (see Figure 3b). In either case, P can then be pushed into a plane just like a quadrilateral. Therefore the space $\mathcal{G}\mathcal{E}\mathcal{O}^{(5)}$ of pentagons is connected, as well.

The situation when $n = 6$ is described in Calvo.^{4–5} In this case, we have to contend with the hexagonal realizations of the trefoil knot. Recall that trefoils are chiral, *i.e.* topologically different than their mirror image. This means that every hexagonal trefoil will lie in a different component of $\mathcal{G}\mathcal{E}\mathcal{O}^{(6)}$ than its mirror image. Therefore, there must be at least three distinct path-components in $\mathcal{G}\mathcal{E}\mathcal{O}^{(6)}$, corresponding to the unknot, the right-handed trefoil,

and the left-handed trefoil. The embedding space $\mathfrak{Geo}^{(6)}$ contains, in fact, five path-components. These consist of a single component of unknots, two components of right-handed trefoil knots, and two components of left-handed trefoil knots. Thus there are two distinct geometric realizations of each type of topological trefoil. In particular, hexagonal trefoil knots are not reversible: In contrast with trefoils in the topological setting, reversing the orientation on a hexagonal trefoil yields a different geometric knot. Hence geometric knottedness is actually stronger than topological knottedness.

It turns out that the distinction between the two geometric types of right-handed trefoils is a consequence of our original choice of root and orientation. If we eliminate this choice by taking the quotient of $\mathfrak{Geo}^{(6)}$ modulo the action of the dihedral group of order 12, we find that the spaces of non-rooted oriented hexagonal knots and of non-rooted non-oriented hexagonal knots each consist of three components (Corollary 7 in Calvo⁴). Randell has reported another approach, using a spectral sequence analysis, confirming these results.^f Nevertheless, the fact that $\mathfrak{Geo}^{(6)}$ consists of five components may prove to be relevant to questions about the topology of DNA, in which there are intrinsic base points and orientations due to the sequences of base pairs.

The classification of hexagonal knots is completed by means of the *joint chirality-curl* \mathcal{J} , a combinatorial invariant which distinguishes between all five components of $\mathfrak{Geo}^{(6)}$. In particular, \mathcal{J} takes values as follows:

$$\mathcal{J}(H) = \begin{cases} (0, 0) & \text{iff } H \text{ is an unknot,} \\ (+1, \pm 1) & \text{iff } H \text{ is a right-handed trefoil,} \\ (-1, \pm 1) & \text{iff } H \text{ is a left-handed trefoil.} \end{cases}$$

Reversing orientation on a hexagon will change the sign of the second coordinate of \mathcal{J} , while taking mirror its image will change the sign of both coordinates.

The minimal stick number for the figure-eight knot is $s(4_1) = 7$. Thus, the space $\mathfrak{Geo}^{(7)}$ contains at least four path-components containing the unknot, the right- and left-handed trefoil knots, and the figure eight knot. Topologically, these are the only four knots that can occur with seven edges. Geometrically, there are five heptagonal knot types, two of which correspond to the figure-eight knot (Theorem 4.1 in Calvo⁵). In fact, heptagonal trefoil knots are achiral but not reversible. This is another example demonstrating the difference between topological and geometric knottedness.

Unlike the hexagonal trefoils, though, the irreversibility of heptagonal figure-eights does not depend on our choice of root. In fact, whereas the

^f Personal communication, AMS meeting, Iowa (March 1996).

space of non-rooted non-oriented embedded heptagons consists of four path-components, the space quotient of non-rooted oriented embedded heptagons consists of five path-components (Corollary 4.7 in Calvo⁵).

Of the nine knot types known to have stick number $s(\cdot) = 8$ (see Figure 2), only two are achiral. Together with the four topological knot types which already occur in $\mathfrak{Geo}^{(7)}$, this gives at least 20 path-components in $\mathfrak{Geo}^{(8)}$. The exact number of geometric – or, for that matter, topological – knot types that can occur when $n = 8$ remains unknown.

1.3. Equilateral polygons

A path in the embedding space $\mathfrak{Geo}^{(n)}$ corresponds to a deformation which can stretch or shrink the edges of a polygon. This type of deformation might be unrealistic when one uses geometric knot theory to model phenomena like DNA molecules. In such cases, we may want a stronger notion of “geometric knot theory” in which the length of the edges remain invariant under deformation. Depending on the relative size of the edges, this new notion of knottedness may actually be different than the more general geometric knottedness described in Section 1.2 above. For instance, Cantarella and Johnston⁶ show that for certain choices of edge length, there are “stuck” hexagonal unknots, *i.e.* polygons which are topologically unknotted but cannot be made planar via geometric deformations that preserve edge lengths.

Let us restrict our attention to the class of equilateral polygons and to deformations which preserve the lengths of their edges. Define the embedding space $\mathfrak{Equ}^{(n)}$ of n -sided *equilateral knots* as the collection of polygons $\langle v_1, v_2, \dots, v_{n-1}, v_n \rangle$ in $\mathfrak{Geo}^{(n)}$ with unit-length edges. Therefore $\mathfrak{Equ}^{(n)}$ is a codimension- n quadric subvariety of $\mathfrak{Geo}^{(n)}$ defined by the equations

$$\|v_1 - v_2\| = \|v_2 - v_3\| = \dots = \|v_{n-1} - v_n\| = \|v_n - v_1\| = 1.$$

Consider the map $f : \mathfrak{Geo}^{(n)} \rightarrow \mathbb{R}^n$ given by the n -tuple

$$f(\langle v_1, v_2, \dots, v_{n-1}, v_n \rangle) = (\|v_1 - v_2\|, \|v_2 - v_3\|, \dots, \|v_{n-1} - v_n\|, \|v_n - v_1\|).$$

The point $p = (1, 1, \dots, 1) \in \mathbb{R}^n$ is a regular value for f (Corollary 1 in Randell²²), so that $\mathfrak{Equ}^{(n)} = f^{-1}(p)$ is a $2n$ -dimensional smooth submanifold which intersects a number of the components of $\mathfrak{Geo}^{(n)}$, some perhaps more than once.

Another helpful way in which to think of the space $\mathfrak{Equ}^{(n)}$ is to use a vector description. An n -sided polygon can be entirely described by its root vertex v_1 and a list of n displacement vectors from one vertex to the next:

$$\vec{V}_1 = v_2 - v_1, \quad \vec{V}_2 = v_3 - v_2, \quad \dots \quad \vec{V}_{n-1} = v_n - v_{n-1}, \quad \vec{V}_n = v_1 - v_n.$$

Each of these is a unit vector and is, therefore, enumerated by a point of the unit-radius 2-sphere \mathbb{S}^2 in \mathbb{R}^3 . A list of n such vectors is subject to the requirement that their sum is the zero vector in order to ensure a closed polygon. This shows that the collection of n -sided equilateral polygons can be considered to be the codimension-3 subset \mathcal{S} of the product $\mathbb{R}^3 \times \mathbb{S}^2 \times \cdots \times \mathbb{S}^2$ determined by the condition that the sum of the n vectors is zero. Note that \mathcal{S} is a real algebraic variety of dimension $2n$. Then the space $\mathfrak{E}qu^{(n)}$ of equilateral knots is the open subset of \mathcal{S} consisting of the points $(v_1, \vec{V}_1, \vec{V}_2, \dots, \vec{V}_{n-1}, \vec{V}_n)$ which correspond to embedded polygons.

We will say two polygons are *equilaterally equivalent* if they lie in the same component of $\mathfrak{E}qu^{(n)}$, and that a polygon is an *equilateral unknot* if it is equilaterally equivalent to a standard planar polygon. Millett has shown that all planar polygons are equilaterally equivalent, so the component of equilateral unknots is well-defined.⁹

As in the geometric case, the space $\mathfrak{E}qu^{(3)}$ of triangles is connected; in fact, $\mathfrak{E}qu^{(3)}$ consists of rotations and translations of a rigid equilateral triangle and is thus homeomorphic to $\mathbb{R}^3 \times SO(3)$. All rhombi (equilateral quadrilaterals) are equilaterally unknotted, since we can view a rhombus as the sum of two isosceles triangles “hinged” along one of the rhombus’s diagonals. For instance, a rhombus $Q = \langle v_1, v_2, v_3, v_4 \rangle$ corresponds to the sum of the triangles $\langle v_1, v_2, v_3 \rangle$ and $\langle v_3, v_4, v_1 \rangle$. We can move v_2 and keep $\|v_1 - v_2\| = \|v_2 - v_3\| = 1$ by rotating the triangular linkage $v_1v_2v_3$ about the axis through v_1 and v_3 until Q lies completely in a plane. Hence $\mathfrak{E}qu^{(4)}$ is connected.

Randell²² showed that any equilateral pentagon can be deformed to a planar one without changing the length of any of its edges. For suppose $P = \langle v_1, v_2, v_3, v_4, v_5 \rangle$ is an equilateral pentagon. Let \mathcal{P} be the plane determined by vertices v_1, v_2 , and v_3 . If \mathcal{P} separates v_4 from v_5 , then either

- (i) both v_3 and v_4 lie on one side of the plane containing v_1, v_2 , and v_5 ,
- or
- (ii) both v_1 and v_5 lie on one side of the plane containing v_2, v_3 , and v_4 .

Thus, after relabeling, we can assume that both v_4 and v_5 lie to one side of \mathcal{P} . In this case, rotate the triangular linkage $v_1v_2v_3$ about the axis through v_1 and v_3 until it lies coplanar with v_4 . We can then deform the quadrilateral linkage $v_1v_2v_3v_4$ in its plane until it misses the line through v_1 and v_4 ; this is easy to achieve since the set of quadrilateral linkages $v_1v_2v_3v_4$ embedded in the plane forms a connected one-parameter family described entirely by the angle $\angle v_4v_1v_2$. We can then rotate the linkage $v_4v_5v_1$ about the axis through

⁹ For example, see steps 2 and 3 in the proof of Proposition 2.1 in Millett.¹⁸

v_1 and v_4 until the entire pentagon lies in a single plane. Since any equilateral pentagon can be flattened out, $\mathfrak{Equ}^{(5)}$ must also be connected.

Consider the case when $n = 6$. We have equilateral examples of each of the five types of hexagons in $\mathfrak{Geo}^{(6)}$. For example, the regular hexagon

$$H_0 = \langle (1, 0, 0), (.5, .866025, 0), (-.5, .866025, 0), \\ (-1, 0, 0), (-.5, -.866025, 0), (.5, -.866025, 0) \rangle$$

is an equilateral unknot, while the hexagon

$$H_1 = \langle (0, 0, 0), (.886375, .276357, .371441), \\ (.125043, -.363873, .473812), (.549367, .461959, .845227), \\ (.818041, 0, 0), (.4090205, -.343939, .845227) \rangle$$

is an equilateral trefoil with $\mathcal{J}(H_1) = (+1, +1)$. Let ρH and rH denote the mirror image (or obverse) and the reverse of a hexagon H ; then ρ , r , and ρr are involutions of $\mathfrak{Geo}^{(6)}$ taking H_1 to equilateral trefoils of the other three types. Therefore $\mathfrak{Equ}^{(6)}$ intersects each of the five components of $\mathfrak{Geo}^{(6)}$ at least once.

The first in-depth analysis of $\mathfrak{Equ}^{(6)}$ was done by Millett and Rosa Orelana. They show that any topologically unknotted equilateral hexagon can be deformed to a planar one without changing the length of any of its edges.^h Thus $\mathfrak{Equ}^{(6)}$ contains a single component of unknots. Calvo⁴⁻⁵ completes the study of equilateral hexagons, showing that any two equilateral hexagons are equilaterally equivalent exactly when they are geometrically equivalent. Therefore, $\mathfrak{Equ}^{(6)}$ contains exactly five path-components, consisting of a single component of unknots, two components of right-handed trefoil knots, and two components of left-handed trefoil knots (Theorem 2 in Calvo⁴). As with $\mathfrak{Geo}^{(6)}$, the joint chirality-curl \mathcal{J} distinguishes among these components. Nevertheless, each component of trefoils in $\mathfrak{Equ}^{(6)}$ contains essential loops which are null-homotopic in $\mathfrak{Geo}^{(6)}$, so that the inclusion $i : \mathfrak{Equ}^{(6)} \hookrightarrow \mathfrak{Geo}^{(6)}$ has a nontrivial kernel at the level of fundamental group. Thus, the trefoil components of $\mathfrak{Equ}^{(6)}$ are not homotopy equivalent to those in $\mathfrak{Geo}^{(6)}$ (Theorem 15 in Calvo⁴). In particular, this shows that, despite the fact that equilateral and geometric knot types coincide in the case $n = 6$, the two notions of knottedness are quite different in nature.

Question 4. Are there values of n for which the number of path-components in $\mathfrak{Geo}^{(n)}$ and \mathfrak{Equ}^n differ? This can occur if there exist either topological knot

^h This result is mentioned, for example, in Proposition 1.2 of Millett.¹⁸ An alternate proof is presented in Calvo.⁴⁻⁵

types which are realizable only by “scalene” n -sided polygons, or equilateral isotopes of the same geometric knot type.

2. Monte Carlo search methods.

2.1. Random knot generation

The complexity of the knot spaces $\mathfrak{Geo}^{(n)}$ and $\mathfrak{Equ}^{(n)}$ has proven to be very difficult to penetrate analytically. Instead, we shall explore these spaces probabilistically, by selecting a large number of “random” configurations in these spaces. As the size of the sample of this Monte Carlo search increases, one obtains a better understanding of the spaces, the topological knot types realized, and the minimal stick numbers of those knots.

Consider generating a random geometric knot in $\mathfrak{Geo}^{(n)}$. By a homothety, any knot type which occurs in $\mathfrak{Geo}^{(n)}$ will be realizable by a polygon in the cube $[0, 1]^3$. This allows us to restrict our attention to the subspace $\mathfrak{Geo}^{(n)} \cap [0, 1]^{3n}$. In this case, one Monte Carlo approach is quite straight forward. With respect to the uniform distribution on the interval $[0, 1]$, one selects a list of $3n$ numbers to represent the coordinates of the n vertices of the polygon. Geometric knots are obtained by connecting these vertices by linear segments cyclically.

A helpful tool in the study of the spaces $\mathfrak{Equ}^{(n)}$ of equilateral knots is the *pivot transformation*. A pivot is determined by a pair of non-adjacent vertices of a polygon, which separate the regular n -gon into two pieces, together with a *pivot angle* $\phi \in [-\pi, \pi]$. The pivot transforms the polygon by holding the image on one of these pieces fixed and rotating the image of the other piece about the axis through the two designated vertices by the given angle. Up to a rotation of euclidean 3-space about the same axis, the result of the pivot will be equivalent to the one given by reversing the roles of the two pieces. Furthermore, if ϕ is sufficiently small, the new polygon will have the same equilateral knot type as the one with which we started. The following two theorems assist in the study of equilateral knot spaces.

Theorem 1. (Proposition 2.1 in Millett¹⁸) For any two equilateral polygons in $\mathfrak{Equ}^{(n)}$, there is a finite sequence of translations, rotations, and pivots taking one polygon to the other.

This theorem is helpful in the study of Question 2, as it provides a method to construct any possible knot type. If the knots are based at $(0,0,0)$ and the second vertex is $(1,0,0)$, then translations and rotations are not required. Since all knot types have a representative of this sort, pivots are the only transformations required. Similar methods show that any path connecting two

equilateral polygons in $\mathfrak{E}qu^{(n)}$ can be approximated as closely as desired by a sequence of pivots. This implies that the pivots generate all possible paths, a fact of relevance in the study of a geometric knot type.

2.2. Recognition of knot type

One important tool in the study of knot spaces has been the calculation of knot invariants that can be used to identify knot types in the standard classification of knots. Historically, the Alexander polynomial has been the principal example, and it continues to be a popular one due to the relative ease of its calculation. Since 1984, however, this has changed with the creation of the Jones polynomial and its successors, the HOMFLY polynomial, the Kauffman polynomial and, more recently, the “quantum” and Vassiliev finite-type invariants. While, in theory, these are impractical due to the complexity of their computation, they actually are remarkably effective in practice. In addition, certain simplifications have proved to be helpful.²⁰ Based upon the observed computational complexity of, for example, the HOMFLY polynomial, one might make the conjecture: “The invariant of the generic knot is easy to compute.”

The HOMFLY polynomial is a finite Laurent polynomial in two variables, l and m , with integer coefficients associated with each topological knot type.^{8,13–14} For the 2977 prime knots represented with fewer than 13 crossings there are only 76 cases that have the same first term as the trivial knot. By considering the entire invariant, these are easily eliminated. Furthermore, most – though not all – chiral knots are distinguished by their HOMFLY polynomial.ⁱ Thus, although there are small families of knots having the same invariant, the HOMFLY polynomial is a good assay for determining topological knot type when dealing with small crossing and stick numbers. For a first estimate, we use distinct HOMFLY polynomials as a surrogate for distinct topological knot type; note that we *do not* identify chiral presentations.

3. Results of search.

In this section, we will describe a number of numerical calculations. Many of these are in a rather preliminary and incomplete state. In particular, we address Questions 1 and 2:

- How many distinct knot types are there in $\mathfrak{G}eo^{(n)}$ or $\mathfrak{E}qu^{(n)}$ as a function of the number of vertices?

ⁱ Knot 9₄₂ is an example of a chiral knot that has the same HOMFLY polynomial as its mirror image.

- What are the stick numbers $s(K)$ and $s'(K)$ of all knots K with nine or fewer crossings?

For $\mathfrak{Geo}^{(n)}$, partial results of research in progress are shown in Figure 4 and Table 1. In this computation, n points are selected randomly within the unit cube $[0, 1]^3$ and connected cyclically by linear segments. The HOMFLY polynomial of the resulting knot is then calculated. The polynomials are counted to estimate the number of distinct knot types.

Figure 4 shows a plot of the number of distinct HOMFLY polynomials observed in $\mathfrak{Geo}^{(n)}$ as a function of n . Note that growth in the number of polynomials gives an estimate of the number of knot types represented. This data clearly indicates the exponential growth in the topological knot types as a function on n . Since, asymptotically, knots are chiral one should divide the number of HOMFLY polynomials by two to estimate the number of topological knot types up to mirror image.

Table 1 displays the observed stick number $s(\cdot)$ for all knots with nine or fewer crossings. Where possible, exact results, such as those discussed in Section 1.1, are given; these are marked by stars (\star). Otherwise, the table indicates the smallest n observed in the Monte Carlo search for which a realization of a given knot type exists.

For $\mathfrak{Equ}^{(n)}$, partial results of research in progress are shown in Figure 5 and Table 2. In this case, the pivot transformation is successively applied, beginning at the regular polygon, and the HOMFLY polynomial is calculated. Then the number of distinct polynomials is counted. Figure 5 shows a plot of the number of distinct polynomials obtained in the search of each space, providing a rough estimate for the total number of distinct knot types in $\mathfrak{Equ}^{(n)}$ as a function of n . In addition to exploring $\mathfrak{Equ}^{(n)}$ for the relatively small values of n shown in Figure 5, the Monte Carlo search was also performed for $\mathfrak{Equ}^{(50)}$, in which case realizations of every knot listed in Table 4 were found. However, the exploration of $\mathfrak{Equ}^{(n)}$ for the smaller values of n is still at an early stage. Table 2 displays the observed equilateral stick number $s'(\cdot)$ for all knots with nine or fewer crossings found thus far by these Monte Carlo searches. As in Table 1, minimal stick numbers are marked by stars (\star).

For larger n , the computational power that is required to obtain an accurate approximation to the total number of knot types realizable in $\mathfrak{Geo}^{(n)}$ or $\mathfrak{Equ}^{(n)}$ can be overwhelming. In practice, one observes only a fraction of the total number of knot types, even after a large number of observations have been made. For example, the case of fifty edges is shown in Figure 6a, which plots the growth in the number of distinct knot types \mathcal{K} observed as a function of the number t of samples taken. After a total of 13,750,000 Monte Carlo observations in $\mathfrak{Equ}^{(50)}$, a total of 2935 distinct HOMFLY polynomials have

Table 1: Observed geometric stick numbers $s(\cdot)$ for knots with nine or fewer crossings. Stars (*) indicate cases for which the minimal stick number is actually known.

| K | $s(K)$ | K | $s(K)$ | K | $s(K)$ |
|---------------------------------|--------|---------------------------------|--------|--|--------|
| 0 | 3* | 8 ₁₃ | 11 | 9 ₂₂ | 14 |
| 3 ₁ | 6* | 8 ₁₄ | 11 | 9 ₂₃ | 14 |
| 4 ₁ | 7* | 8 ₁₅ | 12 | 9 ₂₄ | 12 |
| 5 ₁ | 8* | 8 ₁₆ | 9* | 9 ₂₅ | 15 |
| 5 ₂ | 8* | 8 ₁₇ | 9* | 9 ₂₆ | 12 |
| 6 ₁ | 8* | 8 ₁₈ | 9 | 9 ₂₇ | 12 |
| 6 ₂ | 8* | 8 ₁₉ | 8* | 9 ₂₈ | 12 |
| 6 ₃ | 8* | 8 ₂₀ | 8* | 9 ₂₉ | 15 |
| 3 ₁ + 3 ₁ | 8* | 8 ₂₁ | 9* | 9 ₃₀ | 13 |
| 3 ₁ - 3 ₁ | 8* | 3 ₁ + 5 ₁ | 12 | 9 ₃₁ | 13 |
| 7 ₁ | 9* | 3 ₁ - 5 ₁ | 11 | 9 ₃₂ | 12 |
| 7 ₂ | 9* | 3 ₁ + 5 ₂ | 12 | 9 ₃₃ | 12 |
| 7 ₃ | 9* | 3 ₁ - 5 ₂ | 12 | 9 ₃₄ | 12 |
| 7 ₄ | 9* | 4 ₁ + 4 ₁ | 11 | 9 ₃₅ | 13 |
| 7 ₅ | 9* | 9 ₁ | 13 | 9 ₃₆ | 14 |
| 7 ₆ | 9* | 9 ₂ | 14 | 9 ₃₇ | 14 |
| 7 ₇ | 9* | 9 ₃ | 12 | 9 ₃₈ | 15 |
| 3 ₁ + 4 ₁ | 10 | 9 ₄ | 14 | 9 ₃₉ | 13 |
| 8 ₁ | 10 | 9 ₅ | 13 | 9 ₄₀ | 9* |
| 8 ₂ | 11 | 9 ₆ | 13 | 9 ₄₁ | 9* |
| 8 ₃ | 12 | 9 ₇ | 12 | 9 ₄₂ | 9* |
| 8 ₄ | 10 | 9 ₈ | 13 | 9 ₄₃ | 10 |
| 8 ₅ | 12 | 9 ₉ | 13 | 9 ₄₄ | 10 |
| 8 ₆ | 12 | 9 ₁₀ | 13 | 9 ₄₅ | 10 |
| 8 ₇ | 12 | 9 ₁₁ | 13 | 9 ₄₆ | 9* |
| 8 ₈ | 11 | 9 ₁₂ | 12 | 9 ₄₇ | 12 |
| 8 ₉ | 12 | 9 ₁₃ | 13 | 9 ₄₈ | 12 |
| 8 ₁₀ | 12 | 9 ₁₄ | 14 | 9 ₄₉ | 11 |
| 8 ₁₁ | 10 | 9 ₁₅ | 11 | 3 ₁ + 6 ₁ | 13 |
| 8 ₁₂ | 12 | 9 ₁₆ | 14 | 3 ₁ - 6 ₁ | 13 |
| | | 9 ₁₇ | 14 | 3 ₁ + 6 ₂ | 14 |
| | | 9 ₁₈ | 13 | 3 ₁ - 6 ₂ | 14 |
| | | 9 ₁₉ | 13 | 3 ₁ + 6 ₃ | 13 |
| | | 9 ₂₀ | 13 | 4 ₁ + 5 ₁ | 14 |
| | | 9 ₂₁ | 14 | 4 ₁ + 5 ₂ | 15 |
| | | | | 3 ₁ + 3 ₁ ± 3 ₁ | 10* |

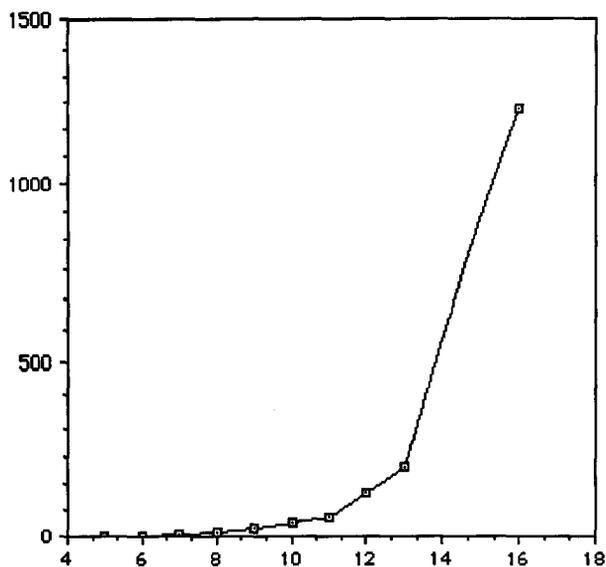


Figure 4: Growth in the number of distinct HOMFLY polynomials observed in $\mathcal{C}_{eo}^{(n)}$, plotted as a function of n .

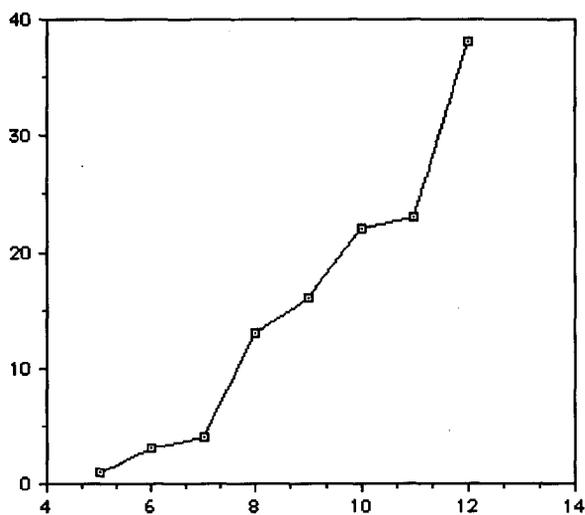


Figure 5: Growth in the number of distinct HOMFLY polynomials observed in $\mathcal{C}_{qu}^{(n)}$, plotted as a function of n .

Table 2: Observed equilateral stick numbers $s'(\cdot)$ for knots with nine or fewer crossings. Stars (\star) indicate cases for which the minimal equilateral stick number is actually known.

| K | $s'(K)$ | K | $s'(K)$ |
|-------------|---------|-------------|---------|
| 0 | 3^* | 7_3 | 12 |
| | | 7_4 | 12 |
| 3_1 | 6^* | 7_5 | 11 |
| | | 7_6 | 12 |
| 4_1 | 7^* | 7_7 | 12 |
| | | $3_1 + 4_1$ | 11 |
| 5_1 | 8^* | 8_2 | 12 |
| 5_2 | 8^* | 8_{13} | 12 |
| | | 8_{14} | 12 |
| 6_1 | 8^* | 8_{19} | 9 |
| 6_2 | 8^* | 8_{20} | 10 |
| 6_3 | 8^* | 8_{21} | 10 |
| $3_1 + 3_1$ | 9 | | |
| $3_1 - 3_1$ | 10 | | |
| | | 9_{42} | 9^* |
| 7_1 | 12 | 9_{45} | 14 |
| 7_2 | 12 | 9_{46} | 14 |

been found. However the shape of the graph indicates that this number is likely to increase significantly after more sampling.

The situation can be likened to a “fish problem,” where one wishes to determine the number of species of fish inhabiting a lake via random sampling of the population. The problem, in the case of knots, is complicated by the fact that the relative proportions of species are far from being uniform. One solution is to approximate the observed values of \mathcal{K} with a function which can then be used to estimate the total number of knot types possible. The data collected during the Monte Carlo searches suggests a function of the form

$$\mathcal{K}(t) = N(1 - Ae^{-kt}). \quad (9)$$

Here the parameter N represents the total number of knot types realizable with n sticks. We consider the sequence of total knot types observed after, say, every 250,000 samples and find the best fitting curve of the form in (9) for these data points. This is done by taking some small integer $j \geq 1$ and considering successive differences of the form

$$\mathcal{K}(t+j) - \mathcal{K}(t) = NA(1 - e^{-kj})e^{-kt}.$$

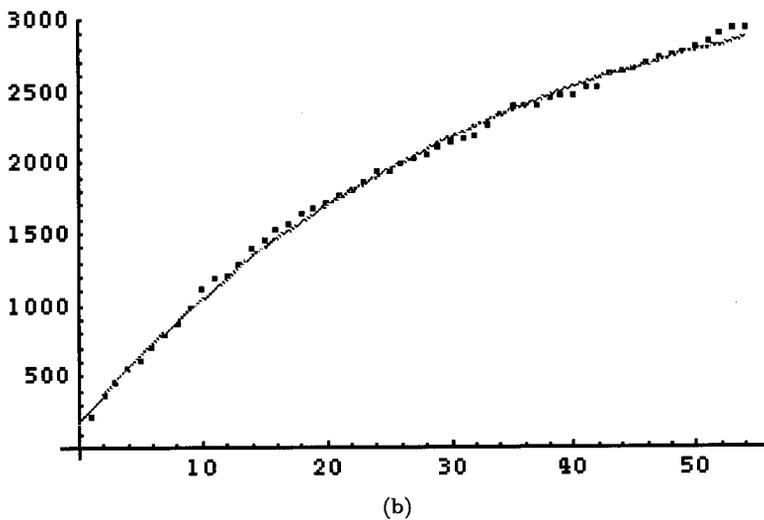
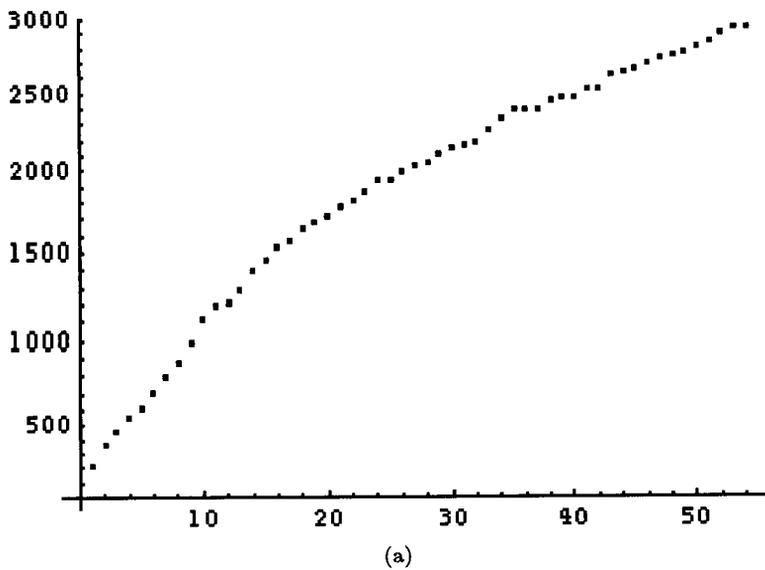


Figure 6: Growth in the number \mathcal{K} of distinct knot types observed in $\mathcal{E}qu^{(50)}$ as a function of the number t of samples. Each unit in t corresponds to 250,000 samples.

Table 3: Approximate number of knot types N in $\mathfrak{E}qu^{(50)}$.

| j | N | R^2 |
|-----|--------|----------|
| 1 | 3471.6 | 0.997536 |
| 2 | 3701.0 | 0.997046 |
| 3 | 3828.0 | 0.996360 |
| 4 | 3826.4 | 0.996369 |

A least-squares regression fitting a line to $\ln(\mathcal{K}(t+j) - \mathcal{K}(t))$ will then yield a “good” value for the parameter k . The desired parameter N is then found by a second least-squares regression fitting a curve of the form in (9) to the data points of \mathcal{K} .

Table 3 shows the approximate total number of knot types N obtained by using $j = 1, 2, 3, 4$ for the Monte Carlo searches for $\mathfrak{E}qu^{(50)}$. The coefficient of determination R^2 for the corresponding curve fit is also indicated in each case.^{*j*} Of these, the better approximation seems to come from $j = 1$, in which case

$$\mathcal{K}(t) = 3471.61 - 3305.46e^{-0.03138t}. \quad (10)$$

Figure 6b shows a plot of this function. This predicts a total of about 3,472 distinct HOMFLY polynomials, providing a conjectured lower bound for chiral knot types in $\mathfrak{E}qu^{(50)}$.

4. Conclusions

The Monte Carlo approach described in Section 2 seems an effective means for producing rough estimates, both in the general case of polygons with arbitrary edge length and in the case of equilateral polygons, of the total number of knot types in each knot space and of the minimal stick number for each topological knot type. In particular, Tables 1 and 2 in Section 3 are the only available compilation of stick number information for all knots with nine or fewer crossings of which we are aware.

^{*j*} The coefficient of determination R^2 is a classical statistical tool measuring how well a curve fits a data set. The better the fit, the closer R^2 is to 1. Given a data set $\{p_i\}_{i \in \{1, \dots, m\}}$ with mean \bar{p} and a curve $q(t)$ approximating it, the coefficient of determination is defined as the ratio

$$R^2 = \frac{\sum_{i=1}^m (p_i - \bar{p})^2 - \sum_{i=1}^m (p_i - q(i))^2}{\sum_{i=1}^m (p_i - \bar{p})^2}.$$

Although the results in Section 3 are only preliminary, one lesson is clear: The amount of sampling required to give an accurate estimates for these quantities will be staggering. For example, consider the data for the Monte Carlo search in $\mathbb{C}u$ ⁽⁵⁰⁾. After 13,750,000 observations the search has revealed 2935, or about 85%, out of a conjectured 3,472 distinct HOMFLY polynomials. According to the prediction curve (10), it should take an *additional* 20 million observations before finding 98% of the conjectured total.

Of course, deeper statistical analysis of the data, especially in regards to the non-uniform distribution in the population of knot types in these spaces, is likely to yield better information in both these areas.

Acknowledgments

The first author was funded in part by a National Science Foundation Graduate Research Fellowship (1995–1998). We would also like to thank Paolo Gardinali for his statistical pointers.

References

1. Adams, C. C., *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*, W. H. Freeman and Co., New York (1994).
2. Adams, C. C., B. M. Brennan, D. L. Greilsheimer, and A. K. Woo, *Stick numbers and composition of knots and links*, *Journal of Knot Theory and its Ramifications* **6** no.2 (1997) 149–161.
3. Birman, J. S. and X. S. Lin, *Knot polynomials and Vassiliev's invariants*, *Inventiones Mathematicae* **111** (1993) 225–270.
4. Calvo, J. A., *The embedding space of hexagonal knots*, preprint (1998).
5. ———, *Geometric knot theory*, Ph.D. thesis, University of California, Santa Barbara (1998).
6. Cantarella, J. and H. Johnston, *Nontrivial embeddings of polygonal intervals and unknots in 3-space*, preprint (1998).
7. Diao, Y., *Minimal knotted polygons on the cubic lattice*, *Journal of Knot Theory and its Ramifications* **2** no.4 (1993) 413–425.
8. Freyd, P., D. Yetter, J. Hoste, W. B. R. Lickorish, K. C. Millett, and A. Ocneanu, *A new polynomial invariant of knots and links*, *Bulletin of the American Mathematical Society (New Series)* **12** no.2 (1985) 239–246.
9. Furstenberg, E., J. Lie, and J. Schneider, *Stick knots*, preprint (1997).

10. Jin, G. T., and H. S. Kim, *Polygonal knots*, Journal of the Korean Mathematical Society **30** no.2 (1993), 371–383.
11. Jin, G. T., *Polygon indices and the superbridge indices of torus knots and links*, Journal of Knot Theory and its Ramifications **6** no.2 (1997), 281–289.
12. Kuiper, N. H., *A new knot invariant*, Mathematische Annalen **278** (1987) 193–209.
13. Lickorish, W. B. R. and K. C. Millett, *A polynomial invariant of oriented links*, Topology **26** no.1 (1987) 107–141.
14. ———, *The new polynomial invariants of knots and links*, Mathematics Magazine **61** no.1 (1988) 3–23.
15. Meissen, M., *Edge number results for piecewise-linear knots*, Knot Theory, Banach Center Publications Vol.42, Polish Academy of Sciences, Institute of Mathematics, Warsaw (1998).
16. ———, *Lower and upper bounds on edge numbers and crossing numbers of knots*, Ph.D. thesis, University of Iowa (1997).
17. Millett, K. C. and D. W. Sumners (eds.), *Random Knotting and Linking*, Series on Knots and Everything Vol.7, World Scientific, Singapore (1994).
18. Millett, K. C., *Knotting of regular polygons in 3-space*, Journal of Knot Theory and its Ramifications **3** no.3 (1994) 263–278; also in *Random Knotting and Linking*, K. C. Millett and D. W. Sumners (eds.), Series on Knots and Everything Vol.7, World Scientific, Singapore (1994) 31–46.
19. Negami, S., *Ramsey theorems for knots, links, and spatial graphs*, Transactions of the American Mathematical Society **324** (1991) 527–541.
20. Przytycka, T. M., and J. H. Przytycki, *Subexponentially computable truncations of Jones-type polynomials*, Graph Structure Theory, Contemporary Mathematics Vol.147, American Mathematical Society, Providence, Rhode Island (1993) 63–108.
21. Randell, R., *A molecular conformation space*, MATH/CHEM/COMP 1987, R. C. Lacher (ed.), Studies in Physical and Theoretical Chemistry **54** (1988) 125–140.
22. ———, *Conformation spaces of molecular rings*, MATH/CHEM/COMP 1987, R. C. Lacher (ed.), Studies in Physical and Theoretical Chemistry **54** (1988) 141–156.
23. ———, *An elementary invariant of knots*, Journal of Knot Theory and its Ramifications **3** no.3 (1994) 279–286; also in *Random Knotting and Linking*, K. C. Millett and D. W. Sumners (eds.), Series on Knots

- and Everything Vol.7, World Scientific, Singapore (1994) 47–54.
24. Scharein, R. G., *Interactive topological drawing*, Ph.D. thesis, University of British Columbia (1998).
 25. Summers, D. W. (ed.), *New Scientific Applications of Geometry and Topology*, Proceedings of Symposia in Applied Mathematics Vol.45, American Mathematical Society, Providence, Rhode Island (1992).
 26. Vassiliev, V. A., *Complements of Discriminants of Smooth Maps: Topology and Applications*, Translations of Mathematical Monographs Vol.98, American Mathematical Society, Providence, Rhode Island (1992).
 27. Whitney, H., *Elementary structure of real algebraic varieties*, Annals of Mathematics (Second Series) **66** (1957) 545–556.

CHAPTER 8

ENTROPY OF A KNOT: SIMPLE ARGUMENTS ABOUT DIFFICULT PROBLEM

A.Yu. GROSBERG

*Department of Physics and Center for Materials Science and Engineering,
Massachusetts Institute of Technology,
Cambridge, MA 02139, USA*

and

*Institute of Biochemical Physics, Russian Academy of Sciences,
Moscow 117977, Russia*

The idea of maximally inflated tube representation of a knot is employed to examine the question of knot entropy, which essentially reduces to how many “conformations” are there corresponding to a given knot topological type. Simple scaling arguments are given, bringing the entire seemingly intractable problem of knot entropy into the realm of simple conventional methods.

1 Introductory remarks

For this book, I don't have to argue that knots are beautiful and important. They are indeed. Knots appear in very different contexts throughout physics, from astrophysics to string theory. They are also challenging as regards their mathematical and/or physical understanding. To be short, they are interesting.

Most of what is known about knots mathematically has to do with their classification and the theory of topological invariants. This has been the main stream of knot theory since the time of P.G.Tait, W.Thomson, and J.C.Maxwell, and at the present day it still remains in the focus of attention.

However important, the theory of invariants does not solve another major aspect, which remains at best on the periphery of attention. I have in mind here what can be called the general problem of “knot entropy.” To explain, I shall concentrate on the special case of knots made up by a long molecular chain of a polymer, such as DNA.

In general, polymers provide a unique application for knot theory. Indeed, while in many other arenas of physics knots appear as a result of an abstraction, knots in DNA as well as in other polymers are very real ones. They can be directly seen in experiments^{1,2}. They are of very obvious meaning, as polymer molecule is a classical object very much similar to a tiny rope.

In this paper, I shall be using approach and terminology which may be very unusual and even disturbing for the knots community. I do that because the problem I am trying to address is so difficult that does not seem to allow fruitful mathematical sophistication. It remains on the level where one can

only exercise physical intuition in an attempt to extract testable (by a real and/or computer experiment) conclusions. This is why I do not attempt doing any mathematics and resort to physical arguments. Nevertheless, I make every effort possible to keep mathematically or topologically oriented reader as long as possible. If the reader can forgive me using sometimes “no knots” instead of “trivial knot” - we can continue. Having said that, I now directly switch to my subject.

2 Basic definitions

Imagine that we are given some polymer chain (for instance, DNA) which is closed circular and, therefore, makes a certain knot. Even if we knew the knot type, and were able to fully characterize it, this itself does not allow to understand physical properties of interest, such as, for instance, averaged size of the molecule in solution, its fluctuations, etc. All those quantities are affected by the topology, they depend on the knot type, and this is relevant both physically and biologically, not the topology itself. As long as thermodynamic equilibrium is concerned, all one needs to compute is the partition function, which can be written in the form

$$Z_K = \sum_C^{(K)} e^{[-E(C)/T]} . \quad (1)$$

Here I use the following informal notations: K denotes the topological type of a knot; C stays for a closed curve in three-dimensional space; $E(C)$ is energy of the polymer when its backbone conformation follows C ; T is an absolute temperature in energetic units; most importantly, $\sum_C^{(K)} \dots$ means summation over “all” curves C whose topology is that of K .

The topology-oriented reader may be wondering as to what is the energy $E(C)$. Physically, it can be either attractive or repulsive interaction between parts of the polymer that come close to each other in space, although they may be arbitrarily far apart along the chain contour. with $0 \leq s \leq 1$ fraction of total length. distances. I shall not explain more about the energy; I mention that physicists believe to have a pretty good understanding of that part, and this is why I shall concentrate on the (mathematically) simplest case when $E(C) = 0$.

Other details in equation (1) may also vary: C may be a path on a lattice for lattice models of polymers; C may be a broken line for the freely-jointed models of polymers; many other models are also possible. I leave aside here the (difficult) problem as to how to understand this sum for the continuous

cases, whether Wiener measure or something else should be employed, etc. Whatever its precise mathematical meaning, the partition function (1) is the physical quantity of interest (not the only one, though; other quantities may be of interest for kinetics of knotted polymers). $F_K = -T \ln Z_K = E_K - TS_K$ is the relevant free energy. As I assume the energetic part E_K to vanish, the problem is about S_K , which I call entropy of a knot. For example, in the notations of the equation (1), equilibrium dimension of a polymer can be written as

$$\begin{aligned}
 R^2 \equiv \langle R_C^2 \rangle_K &= \sum_C^{(K)} R_C^2 \frac{e^{-E(C)/T}}{\sum_{C'}^{(K)} e^{-E(C')/T}} = \\
 &= \sum_C^{(K)} R_C^2 e^{[F_K - E(C)/T]} = \\
 &= \sum_C^{(K)} R_C^2 e^{-S_K}, \tag{2}
 \end{aligned}$$

where the $E = 0$ condition is taken in the last transformation, and

$$R_C^2 = \frac{1}{2} \int ds \int ds' |\vec{r}(s) - \vec{r}(s')|^2 \tag{3}$$

gives the gyration radius of the polymer while in microconfiguration C ; $\vec{r}(s)$ represents the curve C , with $0 \leq s \leq 1$ being the arc length along the curve, expressed as a fraction of total length. Of course, the average value R^2 depends on topology K .

Relatively few computational studies, let alone analytical theories, have been attempted so far to examine the problem. In the work³, we employed the idea of knot inflation, which is the central subject of this book, and suggested a kind of Flory theory to describe equilibrium dimensions of a knotted polymer, depending on both knot topology and a variety of factors controlling interactions, such as temperature, solvent conditions, etc. In this article I present a generalized and simplified version of that theory, which neglects and does not treat the energy part (which is physically interesting, because it can be controlled by solvent and temperature conditions).

The work is organized as follows. In the next section 3, the critical exponent μ is defined which determines the size of an ideal ring polymer with the topology of a trivial knot. This incorporates also fair amount of introductory information about knots, useful for a polymer-oriented reader, as well as some scaling concepts which may appear new for a topology-oriented reader. Then,

in the section 4 we discuss the main results of the work, which is how to apply the idea of maximally inflated knot representation to scaling analysis of knot entropy.

3 Critical exponent for a polymer ring with the topology of a trivial knot

I keep considering the simplest case of no interaction energy, $E(C) = 0$. In polymer physics language, this is called “an ideal polymer,” because it is in some ways similar to an ideal gas. Physically, if there is really no interaction between parts of a polymer, then one would assume these parts to freely pass through each other, with the consequence of relaxed topological constraints: such a polymer would freely go from one knot topology to the other. Thus, what I really want to speak about is a delicate extreme: I assume that $E(C) = 0$, but there is still infinitely high, albeit infinitely narrow, barrier which prohibits passing of the polymer segments through each other. If that is the case, then the only remaining part of free energy is entropy

$$S_K = \ln \sum_C^{(K)} 1 . \quad (4)$$

which is precisely the entropy of a knot K . It has also clear probabilistic meaning, as

$$\mathcal{P}_K = \frac{\sum_C^{(K)} 1}{\sum_C 1} = \frac{e^{S_K}}{\sum_K e^{S_K}} . \quad (5)$$

is the probability to find the knot K upon random choice of a curve C (from a set of interest).

3.1 A reminder: probability of a trivial knot for an N -link ring decays exponentially with N .

Probabilities \mathcal{P}_K have been extensively studied numerically. Some results are summarized in the Figure 1. Many interesting observations can be done using this wealth of information. What becomes clear almost immediately is the fact that the data for a trivial knot probability fit remarkably well to the equation

$$\mathcal{P}_0(N) = \exp[-N/N_0] , \quad N_0 \approx 335 . \quad (6)$$

The very fact of exponential behavior is by no means unexpected: it was predicted by a theorem proven as early as in 1988^{6,7}. What is unexpected,

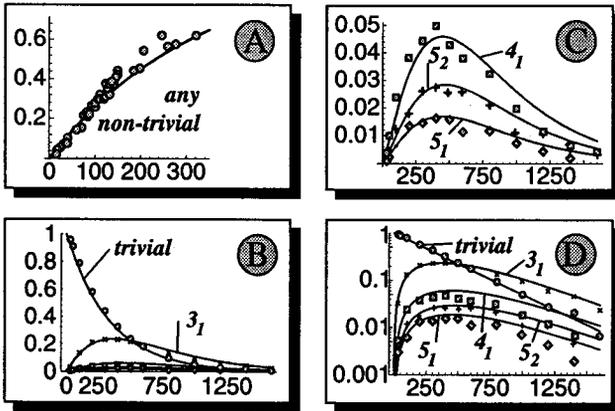


Figure 1: Simulation data for probabilities to find certain knots upon random closure of an N -segment broken line, as a function of N . (A) Early data from the review article⁴. At this time, people used to think of conformations with knots being something unusual, which may or may not be seen at all. This is why what is plotted here is probability of any non-trivial knot $\sum_{K \neq 0} \mathcal{P}_K = 1 - \mathcal{P}_0$. (B-D) Data from the work⁵. The data for trivial knot probability fit remarkably well to the exponential equation 6.

however, is the value of N_0 . Apart from a highly controversial approach⁸, I am not aware of any attempt to compute this quantity theoretically. It is also even qualitatively unclear, why this number is so large. One could relate it to the entanglement number N_e , which is known in the reptation theory¹⁰ and which is also of order of tens or hundreds. It remains a challenge to understand the relation of these two numbers.

Exponential formula (6) can be “understood” in terms of blobs: our polymer consists of N straight segments and can be viewed as a connection of N/g blobs, each consisting of g segments. In order for the

entire polymer to be unknotted, it is necessary that each of the blobs is unknotted, which yields $\mathcal{P}_0(N) \leq [\mathcal{P}_0(g)]^{N/g}$. The exponential formula results from here by replacing inequality with equality. This means that the most “dangerous” knots, which dominate the probability, are those occurring at relatively small scales, as opposed to the scale of a polymer as a whole. This fact itself validates the wording employed above (“each blob is unknotted” etc).

Thus, in $N \rightarrow \infty$ extreme, vast majority of all contours will be of non-trivial topology.

3.2 A conjecture about the size of an ideal trivial-knotted ring

A naive intuition often misses the difference between trivial knot, which is “no knots” at all, and no topological constraints, which is the mixture of all possible knots. For the later problem, computation of the size, or average gyration radius (eq. (2) at $E(\mathcal{C}) = 0$), is straightforward. The result is of the form

$$R^{(0)} \simeq aN^{1/2}, \quad (7)$$

where a is either a length of one straight segment if \mathcal{C} are broken lines; or it is a lattice constant if \mathcal{C} are paths on the lattice; or a is a typical curvature radius (so-called persistence length) if some more sophisticated model for \mathcal{C} is employed (such as, say, wormlike polymer). In what follows, let us choose units of measurement such that $a = 1$. Square root formula (7) is due to the fact that the average is dominated by the majority of contours, which is of the random-walk type.

Now imagine, that we want to compute R for a trivial knot. As trivial knotted rings represent very tiny subset of all contours, there are no grounds to assume that average over the subset will yield the unchanged answer (7). On the other hand, simple self-similarity argument suggests that R has to remain a power law function of N , and thus we can write

$$R|_{\text{trivial}} \simeq N^\mu; \quad (a = 1). \quad (8)$$

Here μ is some critical exponent, the value of μ we do not yet know. Simple arguments show that

$$1/2 \leq \mu < 1, \quad (9)$$

where the left-hand-side estimate follows from the fact that more compact contours are more likely to be non-trivially knotted, and thus the unknotted subset is enriched with more extended forms. I stress, however, that I am not suggesting any proof, and it is still possible that in reality $\mu = 1/2$. In general, the problem to find μ seems to be interesting and challenging. The first attempt to approach it by means of Monte-Carlo experiment has been reported in⁹, but I am not aware of any theoretical works in this direction.

In what follows, I shall simply assume that there exists some exponent μ , and shall express all other quantities of interest in terms of μ . What seems to be attractive in such an approach is how many various quantities appear to be related to μ .

3.3 Applying scaling arguments: confinement of an ideal trivial knot.

The power law expression (8) reflects self-similar character of polymer conformations and can, therefore, serve as the basis for far-reaching scaling consid-

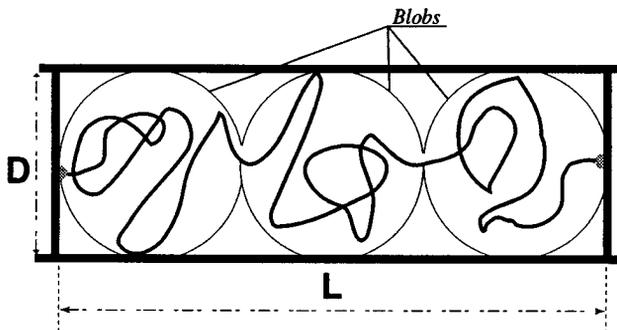


Figure 2: Schematic representation of a trivial-knot polymer confined in a tube.

erations, well known in polymer physics¹⁰. As an (important) example, let us consider the following problem. Consider a cylindrical tube of the length L and diameter D , with axes ratio $p = L/D$ (Figure 2). Imagine now an ideal polymer and suppose that its ends are attached to the two opposing ends of the tube. Imagine further, that the polymer is confined within the tube. In this situation, although polymer is not a ring, its topology is perfectly defined (because ends can be formally connected outside of the tube). Let us assume, that this topology is the one of a trivial knot.

The question one may ask is this: how does the confinement affect the entropy of a knot? Clearly, if the tube is very long, then most of the conformations are more or less close to the axis line of the tube, this restricts the number of conformations and thus reduces entropy. On the other hand, if the tube is too short, restrictions from the sides come into play and reduce entropy as well. It turns out that these changes in knot entropy can be expressed in terms of μ .

Before approaching this problem, however, it is useful to consider a simpler one. Imagine that one of the ends of the tube is designed as a piston and can move freely along the tube, such as the occupied length along the tube, L_0 , is defined by the thermodynamic equilibrium of the system. What will be an equilibrium length? In other words, what is “the most comfortable” length L_0 for an unknotted polymer in the tube of a given diameter D ? Mathematically, this length corresponds to the average, and we assume that the average is the same (to scaling accuracy) as the most probable (or entropy maximizing) value; this assumption is a kind of mean field approximation, of course. To find L_0 , or $p_0 = L_0/D$, we can follow the lines of the scaling consideration of a similar situation for a polymer with excluded volume¹⁰. Namely, let us introduce

blobs (Figure 2) such that their sizes are controlled by the tube diameter, D . Each blob is unknotted; indeed, its “ends” are attached to the neighboring blobs and, therefore, its topology is defined - just as well as the topology of the entire polymer. Therefore, the length of the polymer, g , within each blob is given by $g^\mu \simeq D$ (compare at eq. 8)), or $g \simeq D^{1/\mu}$. As the number of blobs is about N/g , and each of them occupies the length about D along the tube, we finally arrive at

$$L_0 \simeq (N/g)D \simeq ND^{1-1/\mu} ; \quad p_0 = L_0/D \simeq N/g \simeq ND^{-1/\mu} . \quad (10)$$

Clearly, the entropy change due to the restriction in the tube is negative, because the restriction reduces the number of possibilities, and its absolute value of order unity per each blob. Therefore, entropy change due to the confinement is of order

$$\Delta S_0 \simeq -N/g \simeq -p_0 \simeq -ND^{-1/\mu} . \quad (11)$$

As one would expect, both L_0 and p_0 decrease as D increases. Entropy of confinement, $|\Delta S_0|$, also decreases simultaneously, because increase of D means weakening of the restrictions.

In this argument, one assumption is actually hidden: it is assumed that consecutive blobs prefer not to overlap. This is justified by the fact that lion's share of conformations of two overlapping coils correspond to non-trivial links (for the same reason as most of rings are non-trivial knots), while our blobs must avoid being entangled to each other, as chain as a whole is a trivial knot.

Thus, the general line of arguments is very similar to that for the “real” polymers with excluded volume¹⁰. I shall, therefore, bring further scaling arguments without going into their details.

Let us now return to the original question: suppose that L is fixed, and let us find the entropy change due to confinement. Two cases should be considered separately: $L < L_0$ and $L > L_0$, when polymer is compressed or extended, respectively. If $L > L_0$, the polymer is extended and, therefore, dominated by the conformations which (almost) never touch the walls of the tube; vast majority of conformations are restricted within the sausage-shaped area whose diameter can be estimated using Pincus argument¹⁰, yielding entropy estimate of the form $\Delta S \simeq -(L/N^\mu)^{1/(1-\mu)}$, which is conveniently presented in the form $\Delta S \simeq \Delta S_0 (p/p_0)^{1/(1-\mu)}$. In the opposite extreme, when $L < L_0$, the polymer is compressed, lion's share of conformations are such that they fill more or less evenly all the available volume. In this case, blobs must be based on concentration considerations¹⁰, and this yields $\Delta S \simeq -(LD^2/N^{3\mu})^{1/(1-3\mu)}$, which can be also rewritten as $\Delta S \simeq \Delta S_0 (p/p_0)^{1/(1-3\mu)}$.

The above results can be conveniently summarized in the form of the following interpolation expression:

$$\Delta S \simeq - \left(\frac{L}{N\mu} \right)^{1/(1-\mu)} - \left(\frac{N^{3\mu}}{LD^2} \right)^{1/(3\mu-1)} \quad (12)$$

$$\simeq -p_0 \left[\left(\frac{p}{p_0} \right)^{1/(1-\mu)} + \left(\frac{p_0}{p} \right)^{1/(3\mu-1)} \right]. \quad (13)$$

This reproduces both asymptotics of $L \gg L_0$ and $L \ll L_0$ cases, and has also maximum at $L = L_0$, reflecting the fact that L_0 is the most probable size for the polymer in the tube with open ends. As all these arguments do not pretend for more than just scaling accuracy, we don't have to worry about inaccuracies of the interpolation expression (13).

4 The central idea of this work

Consider now a polymer with the topology of an arbitrary knot. Let us assume that this polymer occupies the spatial region of overall size R . The question is how does the knot topology affects the conformational entropy?

To address this question, the idea of knot inflation (or, the same, ideal knot representation¹¹; see also¹²) was suggested³. My favorite way to describe this construction is the following. Imagine that the length of our polymer is fixed and equal L . Let us now draw the tube of some very small radius which contains our polymer as an axis and which has the same topology as the polymer. Let us now inflate this tube. We assume, that parts of the tube cannot penetrate each other, that the axis line remains of the length L , and that the cross-sections of the tube remain all of the same diameter. Then inflation will eventually stop at some maximal diameter D , and both maximally inflated shape of the tube and, in particular, axis ratio $p = L/D$ are topological invariants. This is described numerously throughout this book, albeit maybe in somewhat different words. Now comes the way (or, perhaps, one of the ways) to actually use this topological invariant.

Let us perform the affine transformation of the maximally-inflated tube (only the tube, not the polymer!) such that the tube fits into the sphere of the size R ; let us call this R -size tube. Length L_R and diameter D_R of this R -size tube are easily found from the conditions that $L_R/D_R = p$ (affine transformation) and $L_R D_R^2 \simeq R^3$ (fits into R):

$$L_R \simeq R p^{2/3} \quad D_R \simeq R p^{-1/3}. \quad (14)$$

Here comes the main point of this work: *let us assume that conformations of a ring polymer with the topology of a knot K are dominated by those of a trivial knot confined within the R -size tube of K .* To motivate this idea, let us note that the topology of an unknotted polymer confined within the properly knotted tube is bound to be that of K . On the other hand, confinement entropy does not depend on the way the tube is embedded in the surrounding Euclidian 3D space. Therefore, eq (13) directly applies, and, upon substituting the values of L_R and D_R , yields

$$\begin{aligned} S_K &= S_{\text{trivial}} + \Delta S \simeq \\ &\simeq S_{\text{trivial}} - \beta^{1/(1-\mu)} p^{2/(3(1-\mu))} - \beta^{-3/(3\mu-1)}. \end{aligned} \quad (15)$$

Here S_{trivial} is the entropy of a trivial knot (with the same value of N , of course), and

$$\beta \equiv R/N^\mu \quad (16)$$

characterizes the actual size of a polymer, R , as compared to the averaged size of a trivial knot, N^μ .

We can now maximize entropy with respect to R (or β). The value

$$\beta \simeq p^{(1-3\mu)/3}, \text{ or } R \simeq N^\mu p^{(1-3\mu)/3} \quad (17)$$

is the most probable one, and, to the approximation employed here, is also the average. Upon substituting it into eq (15), we get the following approximate expression for the entropy of an arbitrary knot:

$$S_K = S_{\text{trivial}} - \text{const} p, \quad (18)$$

where const is some number generally of the order unity which depends neither on polymer length N nor on the knot characteristic p . Finally, we can easily note that (i) S_{trivial} should be proportional to N and thus can be generally written as $S_{\text{trivial}} = sN$, and (ii) the number of possible trajectories should be substantially reduced such that $S_K \simeq 0$ for maximally tightened knot, for which $p \simeq N$. From those two conditions, we arrive at

$$S_K = s[N - p]. \quad (19)$$

This is the main result of the work.

5 Discussion

Equation (19) tells us that the entropy of a knot decreases with increasing knot complexity, p ; specifically, it appears linear in p . Although the scaling-based

approach suggested in ³ and further discussed here is unable to compute an absolute value of entropy, presented as the coefficient s , what it can do is to follow the change of entropy from one knot to the other.

To properly understand this result, one has to keep in mind the range of change of p . It reaches its smallest possible value of 2π for the trivial knot, and it grows up to about N for maximally tightened knots. To scaling accuracy, minimal value of p is negligible, because it should be compared to N . On the other extreme, the condition that $p \leq N$ means simultaneously, that the minimal length of a polymer which is able to create a given knot is given as $N = p$.

The result (19) yields very simple conclusion regarding the probabilities of knots \mathcal{P}_K , (5). To find it, one has to compute the normalization factor (or the partition function), $\sum_K \exp[S_K]$. To do that, it is natural to switch from summation over all types of knots, K , to integration over p . This brings about the corresponding Jacobian, or yet another entropy, which is the number of different topological types at the given value of p . This quantity, $\mathcal{K}(p)$, has been argued ³ to grow exponentially with p : $\mathcal{K}(p) \simeq e^{\lambda p}$. With that in mind, one can write

$$\begin{aligned} \mathcal{P}_K &= \frac{e^{S_K}}{\sum_K e^{S_K}} = \frac{e^{s[N-p]}}{\int_0^N dp' e^{\lambda p' + s[N-p']}} = \\ &= \frac{1}{\lambda - s} \exp[-(\lambda - s)N - sp] . \end{aligned} \quad (20)$$

Given that $N \geq p$, we finally arrive at

$$\mathcal{P}_K(N) \sim \begin{cases} 0 & \text{when } N \leq p \\ \exp[s(N-p) - \lambda N] & \text{when } N > p \end{cases} , \quad (21)$$

where constant factor is not written. Clearly, the plot of $\mathcal{P}_K(N)$, according to the result eq (21) and as shown in the Figure 3, looks somewhat similar to the numerical data, Figure 1, C: it is peaked at intermediate value of N and decays exponentially at large N . It follows that $(\lambda - s) = 1/N_0$.

Of course, the very primitive and schematic theory presented here cannot pretend to quantitatively reproduce numerical data. However, it is interesting to understand in which region this theory encounters difficulties: as is seen from comparison the figures, the problems are in the moderate N region, or, better to say, in the N of order p region. It becomes, therefore, clear that imperfections of the theory are in the range where knot is close to maximal tightness, and cannot be really inflated too much. In this case, the situation is obviously largely non-universal, it depends on the microscopic details, such as

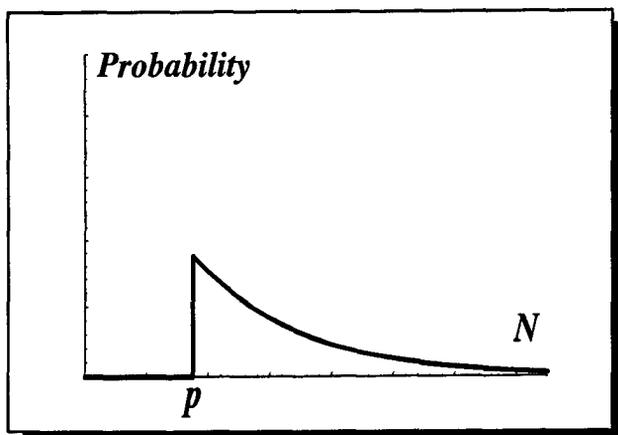


Figure 3: N -dependence of the probability that N -segment polymer ring will be found in the state whose topology is characterized by a given value of p . This dependence is qualitatively reminiscent of the one shown in Figure 1, C.

a particular mechanism of flexibility of the polymer (lattice walks as opposed to broken line type trajectories or to worm-like ones).

Another aspect of the results is the equation (17), which tells us how the most probable size of an (ideal) polymer depends on p , i.e. on the knot topology. Clearly, the qualitative trend is that the polymer is getting more compact when knot is getting more complex. This aspect was discussed in more details in the work³.

To summarize, it has been shown in this article that the idea of maximally inflated knot representation can be fruitfully used to approach the problem of knot entropy. It does yield qualitatively reasonable predictions for knotting probabilities. According to the work³, it can also serve as a reasonable basis to examine the response of a knotted polymer to a variety of physical influences, such as repulsion or attraction between polymer segments. It remains a challenge to examine the problem of “phase segregation” of knots, as it was mentioned in³.

6 Additional Note

When the present work has already been completed, the Editor of this book, Dr. A.Stasiak, paid my attention to the following interesting aspect of the results and their comparison to computational data.

In my original version of the Figure 1, there was a mistake (noticed by A.Stasiak) in labeling data for the knots 5_1 and 5_2 . Correctly labeled figure

indicates that the probability is higher for 5_2 than for 5_1 ($\mathcal{P}_{5_1} < \mathcal{P}_{5_2}$). On the other hand, equation (21) can be treated in terms of its p -dependence: it says, that for the polymers of a given length, N , the probability to obtain a given knot decreases exponentially with the value of p characteristic for that knot. Formally, this contradicts to the example of 5_1 and 5_2 : topological invariant p is higher for 5_2 than for 5_1 . However correct formally, this argument does not rule out the theory, but requires that both the scaling nature of the theory and its limitations be stressed once again. Specifically, one must keep in mind that the approximate scaling theory cannot and does not pretend to be numerically accurate; a "good" scaling theory can capture only overall tendencies and scaling dependencies. Of course, it remains to be seen if the scaling theory described above is "good" or not. But even if it is good to the best of my hope, it remains a challenge to improve it and incorporate additional properties of knots, such as writhe and others.

Acknowledgments

I would like to cordially acknowledge discussions with Y.Rabin and S.Nechaev. The comments of the Editor A.Stasiak helped to improve this paper. The support from National Science Foundation DMR 90-22933 and MCB-93-16-186 is gratefully acknowledged.

References

1. V.V.Rybenkov *et al*, *Proc. Nat. Ac. Sci.* **90**, 5307 (1993).
2. S.Y.Shaw and J.Wang, *Science* **260**, 533 (1993).
3. A.Yu.Grosberg *et al*, *Phys. Rev. E* **54**, 6618 (1996).
4. A.V.Vologodskii and M.D.Frank-Kamenetskii, *Sov. Phys. Uspekhi* **134**, 641 (1981); A.V.Vologodskii *Topology and Physics of Circular DNA* (CRC Press, Boca Raton, 1992).
5. T.Deguchi and K.Tsurusaki, *Knot Theory and its Ramifications* **3**, 321 (1994).
6. D.W.Summers and S.G.Whittington, *J. of Physics A: Math.&Gen.* **21**, 1689 (1988).
7. N.Pippenger, *Disc. Appl. Math.* **25**, 273 (1989).
8. A.L.Kholodenko *Trends in Chem. Physics* **3**, 63 (1994).
9. E.J.Janse van Rensburg and S.G.Whittington, *J. of Physics A: Math.&Gen.* **24**, 3935 (1991).
10. P.-G. de Gennes *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).

11. V.Katritch *et al*, *Nature* **384**, 142 (1996).
12. H.K. Moffat, *Nature* **347**, 367 (1990).

CHAPTER 9

Approximating the Thickness of a Knot

Eric J. Rawdon

*Department of Mathematics, Chatham College, Woodland Road,
Pittsburgh, PA 15232, USA*

The thickness of a unit length C^2 knot is the radius of the thickest "rope" one can place about a knot at the instant that the "rope" self-intersects. Thickness is difficult to compute for all but a few examples. To use computers, a polygonal version of thickness must be defined. The most natural definition does not correctly approximate thickness so a different polygonal version is necessary. This paper contains a definition of a continuous polygonal thickness which correctly approximates smooth thickness. Results on approximation and continuity are stated and examples given of thickness approximations.

1 Introduction

The thickness of a C^2 knot was first defined in ¹. While several subsequent definitions have been given in ²⁻⁵, we focus our attention on the definition in ¹ for the remainder of this paper. Computing the exact thickness of a knot is difficult task; the thickness of a planar circle and some torus knots can be computed, but few other exact calculations are possible. Thus, we need to define a polygonal version of thickness which can be implemented in a computer application to approximate the thickness of a smooth knot. For the computations to be stable, the polygonal thickness function must also be continuous on an appropriate space of polygonal knots. There is a natural definition of polygonal thickness used in ⁶⁻⁸ which is defined by analyzing the intersection of non-consecutive cylinders placed about each side of the polygon. However, it does not correctly approximate the thickness of a large class of knots.

In this paper, we define a polygonal thickness function and sketch the proof that it correctly approximates the thickness defined in ¹ and is continuous. We then use the polygonal version to approximate the smooth thickness of several curves. In section 2, we review the theory of smooth thickness presented in ¹. In section 3, we define a natural cylinder thickness function and show that it does not correctly approximate the thickness of a planar circle. Section 4 contains the definition of polygonal thickness and a sketch of the approximation proof. In section 5, we consider continuity. In section 6, we show examples of thickness approximation for knots with known and unknown thickness measures. This paper is a shortened version of ⁹ and is based on work from my Ph.D. thesis ¹⁰. The proofs of the theorems presented here appear in both ⁹ and ¹⁰.

2 Thickness Review

This section contains background material on smooth thickness from¹. Suppose K is a C^2 submanifold of \mathbb{R}^3 which is homeomorphic to S^1 . We call K a *smooth knot*. For each $x \in K$, let $N_r(x)$ be the disk centered at x which lies in the normal plane of K at x . For sufficiently small r , the union of $N_r(x)$ form a tube about K in which $N_r(x) \cap N_r(y) = \emptyset$ for $x \neq y$ (call $r > 0$ *acceptable* if it satisfies this condition). Define the injectivity radius, $R(K)$, as the supremum of acceptable radii. Thickness, τ , is then the ratio of injectivity radius to the length of the knot and ropelength, ρ , the ratio of knot length to injectivity radius. Note that τ and ρ are both scale invariant. Let $\min Rad(K)$ be the minimum radius of curvature over the points of the knot and $dcsd(K)$ (doubly-critical self-distance) be the minimum distance between distinct pairs which are critical points of the map $(x, y) \rightarrow \|x - y\|$ (called doubly critical pairs). It is shown in¹ that $R(K) = \min\{\min Rad(K), \frac{dcsd(K)}{2}\}$. Polygonal versions of $\min Rad$ and $dcsd$ are utilized in defining polygonal injectivity radius.

3 Cylinder Thickness

There is a natural definition of polygonal injectivity radius. Consider placing cylinders of a given radius about each side of a polygonal knot. Call a radius *acceptable* if no two non-consecutive cylinders intersect and define the cylinder injectivity radius to be the supremum of acceptable radii. The cylinder injectivity radius is continuous on a space of polygonal knots, however it fails to correctly predict the injectivity radius of a planar circle of radius r (which is r) using any sequence of inscribed n -gons.

More specifically, consider a circle of radius 1. After inscribing regular n -gons in the circle, elementary trigonometric calculations show that the cylinder injectivity radius is $\frac{\sin \frac{\pi}{2n}}{\sin \frac{\pi}{n}}$. As n tends to infinity, the injectivity radius actually decreases to $\frac{1}{2}$. Thus, a different definition of polygonal thickness is needed to satisfy both the approximation and continuity conditions.

4 Polygonal Thickness

Our goal is to prove that if $P_n \rightarrow K$, then $R(P_n) \rightarrow R(K)$. Thus, we must define both $P_n \rightarrow K$ and $R(P_n)$. Our strategy for $R(P_n)$ is to define polygonal versions of $\min Rad$ and $dcsd$ defined in section 2 and prove that these functions converge to their smooth counterparts component-wise. This strategy does not work. However, the minimum of the discrete versions does converge to the minimum of the smooth versions (see theorem 4.9). In the remainder of this

section, we define $P_n \rightarrow K$, define polygonal thickness, and sketch a proof of the approximation theorem.

We first present some terminology needed for the definitions. P_n denotes an n -sided oriented polygonal knot with vertices $\{v_0^n, v_1^n, \dots, v_{n-1}^n\}$ in \mathfrak{R}^3 and sides $\{S_0^n, S_1^n, \dots, S_{n-1}^n\}$, where v_i^n is an endpoint of S_{i-1}^n and S_i^n . For ease, the subscripts are always taken modulo n . Let $|S_i^n|$ denote the length of the stick S_i^n and $length(P_n) = \sum_{i=0}^{n-1} |S_i^n|$. The norm of the polygon is defined by $\|P_n\| = \max_{i \in \{0, \dots, n-1\}} |S_i^n|$ and the consecutive edge distortion by

$$E(P_n) = \max \left\{ \frac{|S_0^n|}{|S_1^n|}, \frac{|S_1^n|}{|S_0^n|}, \frac{|S_1^n|}{|S_2^n|}, \frac{|S_2^n|}{|S_1^n|}, \dots, \frac{|S_{n-1}^n|}{|S_n^n|}, \frac{|S_n^n|}{|S_{n-1}^n|} \right\}.$$

Definition 4.1 Suppose $\{P_n\}_{n=N_0}^\infty$ is a sequence of polygonal knots and K is a smooth knot. Then, $P_n \rightarrow K$ if

1. The vertices of P_n lie on the smooth knot K (i.e. P_n is inscribed in K).
2. The vertices of P_n lie in the “correct” order along the oriented curve K , (i.e. v_i^n lies between v_{i-1}^n and v_{i+1}^n on K).
3. $\lim_{n \rightarrow \infty} \|P_n\| = 0$.
4. $\lim_{n \rightarrow \infty} E(P_n) = 1$.
5. P_n has the same knot type as K .

Note that properties 1-4 imply property 5 for sufficiently large n . However, properties 1-3 are not sufficient (see ⁹). While properties 1-3 are standard, property 4 appears potentially restrictive on the sequence of inscribed polygons one can use for approximation. The following result shows that the natural choice for a sequence of inscribed polygons satisfies properties 1-5.

Theorem 4.2 Suppose $f : \mathfrak{R} \rightarrow \mathfrak{R}^3$ is an L -periodic C^2 parameterization (not necessarily by arclength) of a smooth knot K . Let P_n be the polygon created by connecting $f(0)$ to $f\left(\frac{L}{n}\right)$ to $f\left(\frac{2L}{n}\right)$ to \dots to $f\left(\frac{(n-1)L}{n}\right)$ to $f(L) = f(0)$. Then there exists an N_0 such that $P_n \rightarrow K$ for the sequence $\{P_n\}_{n=N_0}^\infty$.

To define the polygonal version of the function *minRad*, we must first define an analogue of radius of curvature for a vertex of a polygon. To motivate the radius of curvature definition, consider two equal-length line segments, R and S , which share a common endpoint v with the external angle at v measuring $\alpha(v)$. Let l_r and l_s be the perpendicular bisectors of R and S respectively. If $0 < \alpha(v) < \pi$, then l_r and l_s intersect at a point P which is

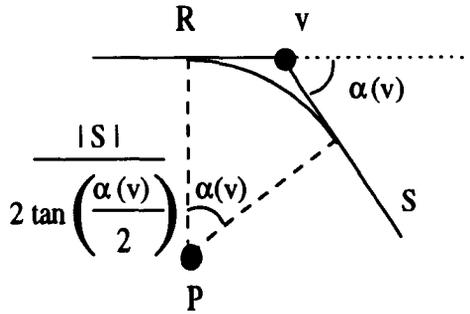


Figure 1: motivation for the definition of polygonal radius of curvature

distance $\frac{|S|}{2 \tan\left(\frac{\alpha(v)}{2}\right)}$ from the midpoints of R and S . Thus, the arc of a circle with center P , radius $\frac{|S|}{2 \tan\left(\frac{\alpha(v)}{2}\right)}$, and total curvature $\alpha(v)$ can be drawn which is tangent to R and S at their respective midpoints. See figure 1.

Definition 4.3 For a vertex v_i^n on P_n , let $\alpha(v_i^n)$ be the exterior angle created by S_{i-1}^n and S_i^n at v_i^n . Then $r_-(v_i^n) = \frac{|S_{i-1}^n|}{2 \tan\left(\frac{\alpha(v_i^n)}{2}\right)}$, $r_+(v_i^n) = \frac{|S_i^n|}{2 \tan\left(\frac{\alpha(v_i^n)}{2}\right)}$, and $\min Rad(P_n) = \min\{r_-(v_0^n), r_+(v_0^n), r_-(v_1^n), r_+(v_1^n), \dots, r_-(v_{n-1}^n), r_+(v_{n-1}^n)\}$.

Definition 4.4 For a point x on P_n , let $D_x : P_n \rightarrow \mathfrak{R}$ be defined by $D_x(y) = \|x - y\|$. We call y a *turning point* for x if y is a local maximum or local minimum of D_x . We call (x, y) a *doubly turning pair* if $x \neq y$, y is a turning point for x , and x is a turning point for y . Let $C(P_n)$ be the set of all doubly turning pairs of P_n . Then $dcsd(P_n) = \min_{(x,y) \in C(P_n)} \|x - y\|$.

Definition 4.5 Define the *polygonal injectivity radius* by

$$R(P_n) = \min \left\{ \min Rad(P_n), \frac{dcsd(P_n)}{2} \right\},$$

the *polygonal thickness* by

$$\tau(P_n) = \frac{R(P_n)}{\text{length}(P_n)},$$

and the *polygonal ropelength* by

$$\rho(P_n) = \frac{\text{length}(P_n)}{R(P_n)}.$$

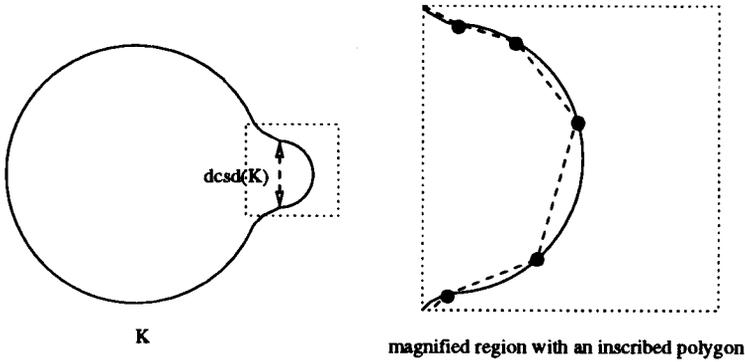


Figure 2: $dcsd$ need not converge for inscribed polygons

Note that τ and ρ are both scale invariant.

It is true that if $P_n \rightarrow K$, then $\min Rad(P_n) \rightarrow \min Rad(K)$ (see theorem 4.6). However if $P_n \rightarrow K$, $dcsd(P_n)$ need not converge to $dcsd(K)$. In figure 2, we see a knot K where $dcsd(P_n) \not\rightarrow dcsd(K)$ regardless of the sequence of inscribed polygons chosen. The $dcsd$ of K is realized at the antipodal pair of the small semi-circle. However, upon magnification of the semi-circle, we see that no inscribed polygon contributes a doubly turning pair in this section of curve. Consequently, $dcsd(P_n)$ converges to the diameter of the larger circle.

The proof of the approximation theorem is technical and relies on three major lemmas.

Theorem 4.6 If $P_n \rightarrow K$, then $\min Rad(P_n) \rightarrow \min Rad(K)$.

Lemma 4.7 If $P_n \rightarrow K$ and $\min Rad(K) \leq \frac{dcsd(K)}{2}$, then $\liminf \frac{dcsd(P_n)}{2} \geq \min Rad(K)$.

Lemma 4.8 If $P_n \rightarrow K$ and $\frac{dcsd(K)}{2} < \min Rad(K)$, then $dcsd(P_n) \rightarrow dcsd(K)$.

Combining these three results proves the desired approximation theorem.

Theorem 4.9 If $P_n \rightarrow K$, then $R(P_n) \rightarrow R(K)$, $\tau(P_n) \rightarrow \tau(K)$, and $\rho(P_n) \rightarrow \rho(K)$.

We can extend smooth thickness to a family of thickness functions by

defining $R_k(K) = \min \left\{ k \cdot \text{minRad}(K), \frac{\text{dcsd}(K)}{2} \right\}$ where $0 < k \leq 1$. In the definition of $R(K)$, we assume that the rope is infinitely compressible; that is, an unknot could be “tied” with 2π inches of one-inch radius rope. By changing k , we can simulate tying knots with rope that is less compressible. This was suggested first by Y. Diao¹¹ and later discussed in². Define τ_k and ρ_k analogously and by a similar argument, the following theorem holds.

Theorem 4.10 *If $P_n \rightarrow K$, then $R_k(P_n) \rightarrow R_k(K)$, $\tau_k(P_n) \rightarrow \tau_k(K)$, and $\rho_k(P_n) \rightarrow \rho_k(K)$ for $0 < k \leq 1$.*

Note that if $P_n \rightarrow K$, then the cylinder injectivity radius converges to $R_{\frac{1}{2}}(K)$. Now we know that polygonal thickness correctly approximates smooth thickness and shift our focus to showing that polygonal injectivity radius, thickness, and ropelength are continuous.

5 Continuity

The proof of continuity relies on an alternate characterization of polygonal thickness. The theorems are more easily stated if we provide the following definitions. For two fixed points x and y on P_n , let B and C be the two arcs of P_n which connect x and y . Then the total curvature between x and y is given by $tc(x, y) = \min \left\{ \sum_{v_i^n \in B} \alpha(v_i^n), \sum_{v_i^n \in C} \alpha(v_i^n) \right\}$. Note that if x happens to be a vertex, then $\alpha(x)$ is a summand in both terms (similarly for y). Let $A_{P_n} = \{(x, y) \in P_n \times P_n \mid tc(x, y) > \pi\}$.

Theorem 5.1 $R(P_n) = \min \left\{ \text{minRad}(P_n), \min_{(x,y) \in A_{P_n}} \frac{\|x-y\|}{2} \right\}$.

Definition 5.2 For a fixed n , let $\mathcal{P}_n = \{(v_0, v_1, \dots, v_{n-1}) \in \mathbb{R}^{3n} \mid \text{the resulting polygon is a non-singular knot}\}$.

Theorem 5.3 $R_k, \tau_k,$ and ρ_k are continuous on the \mathcal{P}_n for $0 < k \leq 1$.

Now we know that polygonal thickness satisfies the desired approximation and continuity properties. The final section uses the techniques stated here to approximate the injectivity radius of some smooth knots.

6 Approximation

In this section, we use the polygonal versions of R , τ , and ρ to approximate their respective smooth values on four different knots. The images shown in [Rawdon Plate 1], [Rawdon Plate 2], [Rawdon Plate 3], [Rawdon Plate 4] are created by TOROS¹³, a descendent of Ying-Qing Wu's Ming¹⁴. In the pictures, the radius of the tube surrounding each side of the polygon is the calculated polygonal injectivity radius.

In [Rawdon Plate 1], we approximate the injectivity radius of a planar circle of radius 1 (which is 1) using regular inscribed n -gons. The number below the picture is the computed polygonal injectivity radius.

In [Rawdon Plate 2], we approximate the injectivity radius of a smooth (2,1)-torus-knot using the algorithm in theorem 4.2. The knot lies on a torus where the larger radius is 4 and the smaller radius is 1. Each point on the knot has an antipodal point (on one of the cross-sectional circles of radius 1) with which it is a doubly critical pair. Thus, the injectivity radius of this curve is 1. The number below each picture represents the computed injectivity radius of the inscribed n -gon.

Harmonic knots¹⁵ are knots which have coordinate parameterizations given by trigonometric polynomials. In [Rawdon Plate 3] and [Rawdon Plate 4], we approximate the ropelength of two harmonic knot parameterizations found in ¹⁵ by applying the algorithm in theorem 4.2. The first is a figure-8 knot and the other an 8_{12} -knot. The number below each picture is the computed polygonal ropelength.

7 Conclusion

To approximate the thickness of a knot, a polygonal version of thickness is needed. To be useful, the polygonal thickness function must correctly approximate the thickness of smooth knots and be continuous. The most natural definition of thickness fails to correctly approximate the thickness of a large class of knots, most notably a planar circle. The polygonal thickness function presented here satisfies both criteria. In addition, an elementary algorithm yields appropriate thickness values.

Acknowledgments

This work is based on my Ph.D. thesis completed under the direction of Jonathan Simon in August 1997. I would like to thank Jonathan Simon for

the many helpful discussions which lead to its completion. Special thanks to Ying-Qing Wu for making the source code for Ming available.

References

1. R. A. Litherland, J. Simon, O. Durumeric, and E. Rawdon. Thickness of knots. *Topology and its Applications*, to appear.
2. R. Kusner and J. Sullivan. On distortion and thickness of knots. preprint, 1996.
3. Y. Diao, K. Ernst, and E.J.J. Van Rensburg. Knot energies by ropes. preprint, April 1996.
4. Y. Diao, K. Ernst, and E.J.J. VanRensburg. Thicknesses of knots. preprint, November 1997.
5. J. Cantarella, D. DeTurck, and H. Gluck. Upper bounds for the writhing of knots and the helicity of vector fields. preprint, December 1997.
6. V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet, and A. Stasiak. Geometry and physics of knots. *Nature*, 384:142–145, November 1996.
7. A. Stasiak, V. Katritch, J. Bednar, D. Michoud, and J. Dubochet. Electrophoretic mobility of DNA knots. *Nature*, 384:122, November 1996.
8. V. Katritch, W. K. Olson, P. Pieranski, J. Dubochet, and A. Stasiak. Properties of ideal composite knots. *Nature*, 388:148–151, July 1997.
9. E. Rawdon. Approximating smooth thickness. submitted, December 1997.
10. E. Rawdon. *The Thickness of Polygonal Knots*. PhD thesis, University of Iowa, 1997.
11. Y. Diao, June 1996. personal communication.
12. O. Durumeric, R. A. Litherland, E. Rawdon, and J. Simon. Thickness of knots 2. preprint.
13. E. Rawdon. *TOROS*. University of Iowa, <http://www.math.uiowa.edu/~rawdon>. Program for visualizing, manipulating, and thickness maximizing polygonal knots.
14. Y.-Q. Wu. *Ming*. University of Iowa, <http://www.math.uiowa.edu/~wu>. Program for visualizing, manipulating, and energy minimizing polygonal knots.
15. A. Trautwein. *Harmonic Knots*. PhD thesis, University of Iowa, 1995.

CHAPTER 10

Preface to Energy functions for knots: beginning to predict physical behavior

Jonathan Simon
Department of Mathematics
The University of Iowa
Iowa City, IA 52242-1419 USA
 simon@math.uiowa.edu

Ideal Knots

In November, 1991, I gave a talk on “energy functions for knots” in a symposium on Topology and Chemistry held in Bielefeld. Topologists (e.g. S. Fukuhara, J. O’Hara, M. Freedman et al., G. Buck and JS) had recently begun defining and analyzing several so-called energy functions that would separate different types of knots by infinitely high “potential walls” and produce minimum energy conformations that I rashly called *ideal*.

After the talk, Andrzej Stasiak suggested that the term “ideal” should better be used for something based on more classical considerations: think of knotted tubes and try to minimize the amount of surface area needed to enclose a given volume. We didn’t pursue the discussion much at that time, and so missed the obvious connection between his thoughts and the work (with R. Litherland) on “Thickness of knots” that I was to present at an AMS meeting the following week. (Those results subsequently were improved^{2,9}.)

A smooth knot in 3-space has a tubular neighborhood whose cross-sections are disks of some radius centered at each point of the knot, and perpendicular to the knot. For each curve, we can think of making the tube-radius smaller or larger; if we make the radius too large, then the disks will overlap. The supremum of all radii for which the disks are pairwise disjoint is a measure of how wrinkled or crumpled is the curve. However, if we expand the knot by stretching equally in all directions (e.g. triple the $[x, y, z]$ coordinates of all points) then this number will change; we want the tube-thickness to measure how crumpled the knot is, as a matter of shape, not absolute size. We can accomplish this by studying the ratio of tube-radius to the length of the curve; the supremum of these ratios is the *thickness* of the knot.

What about surface area vs. volume? Theorems of Pappus (Alexandria, 340 c.e.) on figures of revolution apply equally well to solids generated by disks (or surfaces generated by the boundaries of the disks) moving perpendicularly

to a central curve: if the disks have radius r (small enough that the disks do not overlap) and the central curve has length ℓ , then the volume of the solid torus generated is $\pi r^2 \ell$ and the surface area is $2\pi r \ell$. So the ratio of surface area to volume is $\frac{2}{r}$. The first conclusion is that unless we constrain the problem, e.g. by insisting that the length of the curve is fixed, the ratio of surface area to volume is just a matter of scale, not of shape: without changing the shape of the knot, we can enlarge or shrink the scale to make the ratio arbitrarily close to zero or arbitrarily large. If we fix the scale by insisting, for example, that the knot have arclength $\ell = 1$, then minimizing the ratio of surface area to volume is the same as maximizing $r = \frac{r}{\ell}$, the *thickness*.

For a given smooth curve, embedded tubes exist, and we can contemplate the supremum of admissible radii. A deeper question arises when we allow the curve to move (while preserving the knot-type) in hopes of discovering “better” conformations. Does there exist a curve within each knot type for which the thickness is the maximum among all curves of that knot-type? Such a curve exists^{8,5}, but the theory does not guarantee that it is very smooth (in the sense of differentiable of class C^1 vs. class C^2). Except for a round circle, it may be that an “ideal” knot is not so ideal! This is a problem we mathematicians and scientists create when we use value-laden words to describe technical phenomena; but still we are drawn to the question of geometric ideals.

We might note that while the intuitive motivations for “energy functions” for knots and “thickness” are different (in one case, the knot is made of stuff that somehow repels itself; in the other case the knot is made of stuff that occupies space and excludes itself), there is a lot of mathematical similarity. It has been observed by many people that the reciprocal of thickness, which we may call the “rope length energy” of a knot, has the same kinds of properties as other energy functions. We have obtained^{2,3} relations between rope-length and other energies, leading to the result that (up to a constant of proportionality) the crossing number of a knot is bounded by the $\frac{4}{3}$ power of the rope-length.

In a color plate elsewhere in this volume, we illustrate two “ideal” trefoil knots; the curves are drawn as thick tubes to make the geometry more visible. In one, we see a trefoil knot that is “ideal” in the sense of having low *minimum distance energy* (see e.g. reference 11 reprinted below). The other “ideal” trefoil knot has large *thickness* as described above. Which is the more ideal?

In another color plate, we illustrate the “evolution” or “flow” of a tangled knot to a more recognizable form, where the evolution is controlled by the minimum distance energy mentioned above. It is difficult to call the final conformation “ideal” because it does not have some evident quality of great symmetry or unusual beauty. But perhaps this is as close as this particular knot type can come to having an “ideal” realization.

Polygonal and smooth knots

It is clear in the abovementioned illustrations that the knots being shown are polygons. Sometimes we use computer enhancement to create the illusion that the knots we are drawing are smooth curves or smooth tubes. But of course the computer can only encode and communicate discrete data.

In his work on thickness of polygonal knots, E. Rawdon¹⁰ developed a “thickness” for polygons that approximates, and converges to, the thickness of a smooth curves that is being approximated by the polygons. The most straightforward definition of “polygonal thickness” in fact does not work, so one needs a more complicated version.

Several years ago, we explored an alternative approach (which has occurred to others as well—see e.g. work of L. Kauffman⁶). We can use a computer to represent a smooth curve by a finite list of coefficients of trigonometric polynomials $x(t)$, $y(t)$, $z(t)$ of some degree (i.e. finite Fourier approximations). A. Trautwein¹³ found bounds on the “harmonic degree” and the crossing number of the knot in terms of each other.

Gel velocity

In the following paper¹¹ we discuss the phenomenon of gel electrophoresis of knotted DNA loops and, in particular, the question of whether one can compute geometric measures of complexity for knot types that will predict which kinds of knots move faster in the gel than which others.

The geometric property that was first observed to influence velocity was the basic fact of being knotted: nontrivial knots move in the gel more quickly than unknotted loops of the same intrinsic length.

The next observation was that the minimum crossing numbers of the knot types usually predict (qualitatively) the relative velocity. This was a case of measuring some quantity evident in “ideal” conformations (in this case, minimum crossing number projections).

Energy functions for knots can provide even more accurate predictions than crossing number. Knots (represented in a computer as polygons) are allowed to flow to apparent minimum energy conformations, and those minimum energy numbers correctly predict¹¹ the ordering of velocity for certain knots with the same crossing number (as well as the previously known examples with different crossing numbers).

More recently, A. Stasiak and colleagues have shown¹² that the maximum thickness of knots (see above—again, a number taken from an “ideal” conformation of a polygonal knot) even gives a *quantitative* prediction of relative

velocity: there is (at least for relatively simple knots, and especially within homologous families of knots) a good linear correlation between rope-length and gel velocity.

In order to fully test the extent to which purely geometric properties of knot types, in particular, numbers taken from special conformations of the knots, can predict gel velocity, we would need more experimental data on velocities of fairly complicated knots. An alternative route might be careful measurement of velocities (of even simple knots) under varying experimental conditions. One essential lesson⁷ is that the experimental conditions cannot be ignored: a particular knot type can be faster than another one under certain conditions, and the order reversed under different conditions.

The paper¹¹ that is reprinted following this Preface is reprinted here with permission of the publisher. As the present volume attests, the subjects of thickness (i.e. rope-length), energies, and other topics in *physical knot theory* have attracted considerable attention, and even some lively debate^{1,4} in the last few years.

References

1. G. BUCK, *Four-thirds power law for knots and links*, Nature **392**, No. 6673 (1998), 238–239.
2. G. BUCK AND J. SIMON, *Energy and length of knots*, Lectures at Knots 96 (S. Suzuki, ed.), World Scientific Publ., 1997, pp. 219–234.
3. G. BUCK AND J. SIMON, *Thickness and crossing number of knots*, Topology and its Applications to appear.
4. J. CANTARELLA, R. KUSNER, AND J. SULLIVAN, *Tight knot values deviate from linear relations*, Nature **392**, No. 6673 (1998), 237–238.
5. O. DURUMERIC, work in progress.
6. L. KAUFFMAN, *Fourier knots*, in this volume.
7. S.D. LEVENE AND H. TSEN, *Analysis of DNA knots and catenanes by agarose-gel electrophoresis*, Protocols in DNA Topology and Topoisomerases, Vol. I (M. Bjornsti and N. Osheroff, eds.), Humana Press, 1996.
8. R. LITHERLAND*, J. SIMON, AND O. DURUMERIC, *Thickness of knots*, talk in AMS Special Session on Physical Knot Theory, Iowa City, 3/96.
9. R. LITHERLAND, J. SIMON, O. DURUMERIC, AND E. RAWDON, *Thickness of knots*, Topology and its Applications (1998 to appear).
10. E. RAWDON, *Thickness of polygonal knots*, in this volume.
11. J. SIMON, *Energy functions for knots: beginning to predict physical*

behavior, *Mathematical Approach to Biomolecular Structure and Dynamics* (J.P. Mesirov, K. Schulten, and D.W. Sumners, eds.), Proc. of 1994 IMA Summer Program on Molecular Biology, IMA Volumes No. 82, Springer-Verlag, 1996, pp. 39–58.

12. A. STASIAK, V. KATRITCH, J. BEDNAR, D. MICHOU, AND J. DUBOCHET, *Electrophoretic mobility of DNA knots*, *Nature* **384**, Nov. 1996, 122.
13. A. TRAUTWEIN, *Harmonic knots*, in this volume.

Energy functions for knots: beginning to predict physical behavior

Jonathan Simon
Department of Mathematics
The University of Iowa
Iowa City, IA 52242-1419 USA
simon@math.uiowa.edu

Several definitions have been proposed for the “energy” of a knot. The intuitive goal is to define a number $u(K)$ that somehow measures how “tangled” or “crumpled” a knot K is. Typically, one starts with the idea that a small piece of the knot somehow repels other pieces, and then adds up the contributions from all the pieces. From a purely mathematical standpoint, one may hope to define new knot-type invariants, e.g by considering the minimum of $u(K)$ as K ranges over all the knots of a given knot-type. We also are motivated by the desire to understand and predict how knot-type affects the behavior of physically real knots, in particular DNA loops in gel electrophoresis or random knotting experiments. Despite the physical naïveté of recently studied knot energies, there now is enough laboratory data on relative gel velocity, along with computer calculations of idealized knot energies, to justify the assertion that knot energies can predict relative knot behavior in physical systems. The relationships between random knot frequencies and either gel velocities or knot energies is less clear at this time.

“Suppose you have a knotted loop of string, and you spread an electric charge along the string and then let go; what will happen?” This question has been a common “cocktail party” topic among knot theorists for many years, but recently questions like this have been the objects of serious study. The motivation is twofold: We want to develop new knot-type invariants that describe, in intuitively satisfying ways, how complicated/tangled/interwound one knot is relative to another, and also we want to build a mathematical environment in which one can hope to model, understand, and predict how different types of knots behave in physical situations. If one imagines different knots tied from the same length of string, then it is at least plausible that more complicated knots would be spatially more compact, so a quantitative measure of complexity might be used as a quantitative measure of compaction.

This paper provides an expository introduction to knot energy functions, preliminary announcement of our computational results on energy minima for various knot types, and an analysis of how these energy numbers relate to other measures of knot complexity, in particular frequency of different knots in random knotting experiments and, most of all, relative velocity of differently knotted DNA loops in agarose gel electrophoresis.

To provide the most direct route to the results, we present first the main discussion, and then provide the background on knots and energy functions in the subsequent sections.

1 Gel velocity, random knot frequency, and topological ground state energy of a knot

There is clear laboratory evidence that when closed DNA loops are knotted, different types of knots behave differently in gel electrophoresis^{10,12,13,18,49}. The duplex loops are long enough and nicked, so they are conformationally flexible; the differences in gel velocity are detecting different *configurations* (or time-averaged conformations, as opposed to individual rigid conformations), what the topologists would call different *knot types*. This differential mobility rapidly evolved from being a discovery to being an analytical tool^{12,39}, so that in reference 39, the knot types are determined entirely by gel mobility, without recourse to electron microscope confirmation that was typical in earlier papers. The intuition is that if a “complicated” and a “simple” knot are made from string of the same length, then the complicated knot will be, in time averaged conformation, more tightly crumpled, hence move faster in agarose gel.

There is similarly convincing evidence that different knot types have different probabilities of being created in random DNA knotting experiments^{39,40,41}. Here the intuition might be that, other factors being equal, a particular “complicated” knot is kinetically or entropically less likely to occur than a particular “simpler” one. However, this seemingly reasonable belief clashes with the equally appealing intuition (and, in fact, a theorem^{33,45}) that a *very* long “random loop” is not likely to be unknotted, and in fact is likely to be a complicated knot.

One possible way to reconcile the intuitive belief that ‘more complicated’ implies ‘less likely’ with the fact that loops of long strings ‘must’ be complicated knots would be if, as the number of segments in a random polygonal loop increases, while the probability of finding one particular simple knot type decreases towards zero, perhaps that probability still is greater than the probability of finding *one particular* complicated knot type. However, in the computer simulations of reference 11, there is data saying that for certain numbers of segments, a knot of type A is more likely than a knot of type B, while for more segments, type B is more likely than type A. It might be significant that this phenomenon is observed for composite knots (see section 2 below for knot terminology), but not (yet) observed for prime knots. In reference 11, and in the computer simulations in reference 39, the ordering of prime knot types by their probabilities among random polygons of different lengths is independent of segment number.

There is a perfect (qualitative) correlation between relative gel velocity and relative frequency for the first few knot types, i.e. the unknot, trefoil, and figure-eight (see section 2 below for definitions). Knots which move faster in

the gel are less likely to be produced by random knotting than knots which migrate more slowly. However, when we get to knots with five crossings (so-called knot types 5_1 and 5_2) the knot 5_2 , which is slightly faster in the gel^{12,13}, seems to be *more* likely to occur than the slightly slower 5_1 ³⁹.

The crossing number (i.e. node number) of a knot type also is, initially, a good predictor of relative frequency. However, according to Figure 3.5 of reference 11, for large numbers of segments, the *nine* node composite knot $3_1\#3_1\#3_1$ is more likely to occur than the *seven* node knot $3_1\#4_1$.

Until we have a better understanding of the interplay between knot type and number of segments (i.e. string length), it does not seem possible to predict the relative frequencies of different knot types in random knotting experiments *in vitro* or *in virtuo*⁴⁴ based on one numerical measure of complexity. Predicting relative gel velocity seems a more attainable goal.

As a first sorting, the crossing numbers (see section 2 below) of knot-types seems to work very well as a predictor of relative gel velocity¹⁰, though (Table I, p. 4979 of reference 10, esp. band 6) the separation seems to get less clear as the number of nodes increases. (For homologous series, e.g. of $(2, n)$ torus knots⁴² and twist knots¹⁸, the coincidence of gel bands and crossing numbers seems perfect.) Within the gel region associated with a given crossing number, it may be possible to distinguish bands corresponding to different knot types with the same crossing number, e.g. the ‘granny knot’ vs. the prime six crossing knot 6_1 ¹⁸. Another situation where gel velocity distinguishes between different knots having the same crossing number is the observation^{12,13} that the $(2, n)$ torus knots migrate slightly slower than the corresponding n -crossing twist knots. Of course it is a tautology that we would need a numerical measure of complexity that is finer than crossing number, if we want to distinguish knots of the *same* crossing number; and apparently “nature” does indeed make such finer distinctions. In Table 1 below, we list the first few knot types in order of observed relative gel velocity, as well as the behavior of some homologous families.

The *energy functions* E_2 and U_{MD} for knots, described in sections 3.2 and 3.3 below, seem to make the distinctions we want. In Table 2, we list the computationally estimated “topological ground state energies” for various knots. The numbers \tilde{U}_{MD}^0 are our own calculations. The values of $\min E_{\text{cos}}$ (here E_{cos} is an approximation of $E_2 - 4$) are taken from Table 2 of reference 20, except the value for the square knot and granny, which are taken from elsewhere in that paper. The two energy functions order the knots identically through 7_1 . The more tightly crumpled a knot is, the higher its “energy” and, for all knots through six crossings and most through seven, the energy ranking respects crossing number. The sudden appearance of the $(3, 4)$ torus knot 8_{19}

Table 1: Ranking of DNA knots by relative gel velocity

| knot types (slow to fast) | general observations |
|---|---|
| unknot | All knots are generally ordered by crossing number, especially through six nodes, but less distinctly as number grows ¹⁰ . |
| trefoil knot 3_1 | |
| figure-8 knot 4_1 | |
| (2,5) torus knot 5_1 | |
| twist knot 5_2 | |
| granny knot $3_1\#3_1$ vs. 6_1 ¹⁸ | Twist knots are slightly faster than $(2, n)$ torus knots with the same crossing number ¹³ . |
| prime knots $6_1, 6_2, 6_3$ (no reports that these were separated by gel) | Homologous family of twist knots are ordered by crossing number ¹⁸ . |
| higher crossing knot types | Homologous family of $(2, n)$ torus knots are ordered by crossing number ⁴² . |

early among the seven crossing knots is consistent with its anomalous physical behavior⁹ (and also its anomalously low minimal edge number on the cubic lattice³⁷). The minimum energy for knot type 5_2 is slightly higher than the minimum for 5_1 , so the energy functions would predict that 5_2 moves faster in the gel than 5_1 . Likewise, the energy functions would predict that the composite six crossing knots, i.e. the granny knot or square knot, would move slower in the gel than the twist knot 6_1 . When we get to 7 and 8 crossings, the two energy functions differ slightly in their rankings. Even through all the eight and nine crossing prime knots, the two lists differ only occasionally, and then only by 1–3 places, in how knots are ranked. We plan in the future to obtain better estimates of the various minima (in particular to make more systematic the search for possibly different local minima) and will then be better able to reconcile, or contrast, the two rankings. For now, we accept the overall similarity of the lists, along with conceptual similarities in the definitions of these functions, as a strong statement that the two notions of energy are measuring essentially the same quality of relative knot complexity. Regarding gel velocity of DNA loops, **the qualitative ranking of knots by estimated minimum energy is consistent with all the observed velocity differences noted in Table 1.**

HERE ARE SOME ADDITIONAL NOTES ON TABLE 2.

- The (so far estimated) energy \tilde{U}_{MD}^0 for the granny knot is less than for the square knot, but the difference is less than the likely error in the data, so no difference is reported here.
- The numbers \tilde{U}_{MD}^0 for knots of six or more crossings are rounded to the nearest integer, to remind us that these are rougher approximations than for the simpler knots.
- We have not yet computed a value for the composite $3_1\#4_1$, so that knot should be inserted, perhaps first among the seven crossing knots (the conjecture on additivity of minimum conformal energy²⁰ would place it just after 6_3).
- According to both \tilde{U}_{MD}^0 and $\min E_{\text{cos}}$, the prime seven crossing knots are ordered $7_1, 7_2, 7_3, 7_4, 7_5, 7_6, 7_7$. That is, the same as their numbering in the classical knot tables. This is a remarkable coincidence, inviting speculation on what aesthetic sense, or intuition about complexity, led the early tablers to list the knots in that order.

It is appropriate to say “qualitative” in the last sentence of the paragraph above, because of the pseudo-physical nature of the knot energy functions:

Table 2: Ranking of knot types by estimated “topological ground state” energy

| knot type | \tilde{U}_{MD}^0 | min E_{cos} |
|---------------------------|--------------------|-----------------------------|
| unknot | 0 | 0 |
| trefoil 3_1 | 37.2 | 70.4 |
| figure-8 4_1 | 56.9 | 104.9 |
| (2,5) torus knot 5_1 | 74.9 | 126.8 |
| twist knot 5_2 | 81.6 | 134.6 |
| granny knot $3_1\#3_1$ | 110 | 140.8 |
| square knot $3_1\#3_1^*$ | 110 | 140.8 |
| twist knot 6_1 | 120 | 162.8 |
| 6_2 | 128 | 168.5 |
| 6_3 | 134 | 172.9 |
| (2,7) torus knot 7_1 | 139 | 181.0 |
| (3,4) torus knot 8_{19} | 144 | 197 |
| twist knot 7_2 | 153 | 190.3 |
| 7_3 | 155 | 192.7 |
| 8_{20} | 157 | 203.9 |
| 7_4 | 160 | 197.7 |
| 7_5 | 167 | 199.7 |
| 7_6 | 170 | 203.7 |
| 7_7 | 175 | 207.1 |

these are purely geometric quantities, albeit physically motivated. However, there is a provocative parallel between the energies discussed here, and the calculated free energies of formation in references 40 and 41. While comparing the absolute numbers is meaningless without some agreement on units, if we compare the ratios, we get surprisingly close agreement. In reference 40, the calculated energy of formation of the figure-8 knot at 20 [resp. 30] Kuhn lengths and maximum MgCl_2 concentration (maximum salt \implies polymer is most able to knot) is 1.69 [resp. 1.64] times the energy for a trefoil knot. On the other hand, the ratio of minimum energies \tilde{U}_{MD}^0 of the two knot types is 1.66. Obviously, more data is needed to test whether the energy U_{MD} really can predict so closely the relative energies of formation of different knots *in vitro*.

If the prime six-crossing knots can be distinguished by gel velocity, that would provide a good further test of our assertion that energy functions predict relative velocity. Another source of verification (or challenge, of course) would be to carry out the detailed gel electrophoresis simulations of references 21 and 22 for knots beyond the trefoil. Another project that would seem very interesting would be to try to relate our energy functions to other direct geometric measures of compaction, such as mean square radius of gyration or mean span³⁶. The Monte Carlo study of random polygons in reference 36 can be interpreted to say that if polygonal loops with many segments are viewed from far away, so that only gross features such as overall diameter are discernable, then one won't be able to see any difference between unknots and (even complicated) knots. This seems vaguely similar to the observation one can make from the table of energy minima, especially reference 20, that as the crossing number gets large, there is not much difference between the energy minima for different knot types.

While we are focusing on knots, it should also be noted that DNA catenanes exhibit interesting and useful gel behavior. In reference 1, and the survey reference 12, it is noted that catenanes that differ only in their linking pattern (e.g. *n*-form where four loops are linked in a linear chain vs. iso-form where there are three in a row, with the fourth linked to the one in the middle—see [Figure 2]) are examples of DNA stereoisomers that have the same crossing numbers but different gel velocities. Can knot energies capture this difference? Here it will be necessary to modify the existing minimization algorithms so as to maintain the string length of each loop. For the catenanes in [Figure 2A], we do not (yet) know whether forcing the four loops to maintain identical lengths will produce the same energy ranking as if we allow the loops to change relative lengths. In [Figure 2B], the *only* difference between the two catenanes is the location of the long loop relative to the short ones.

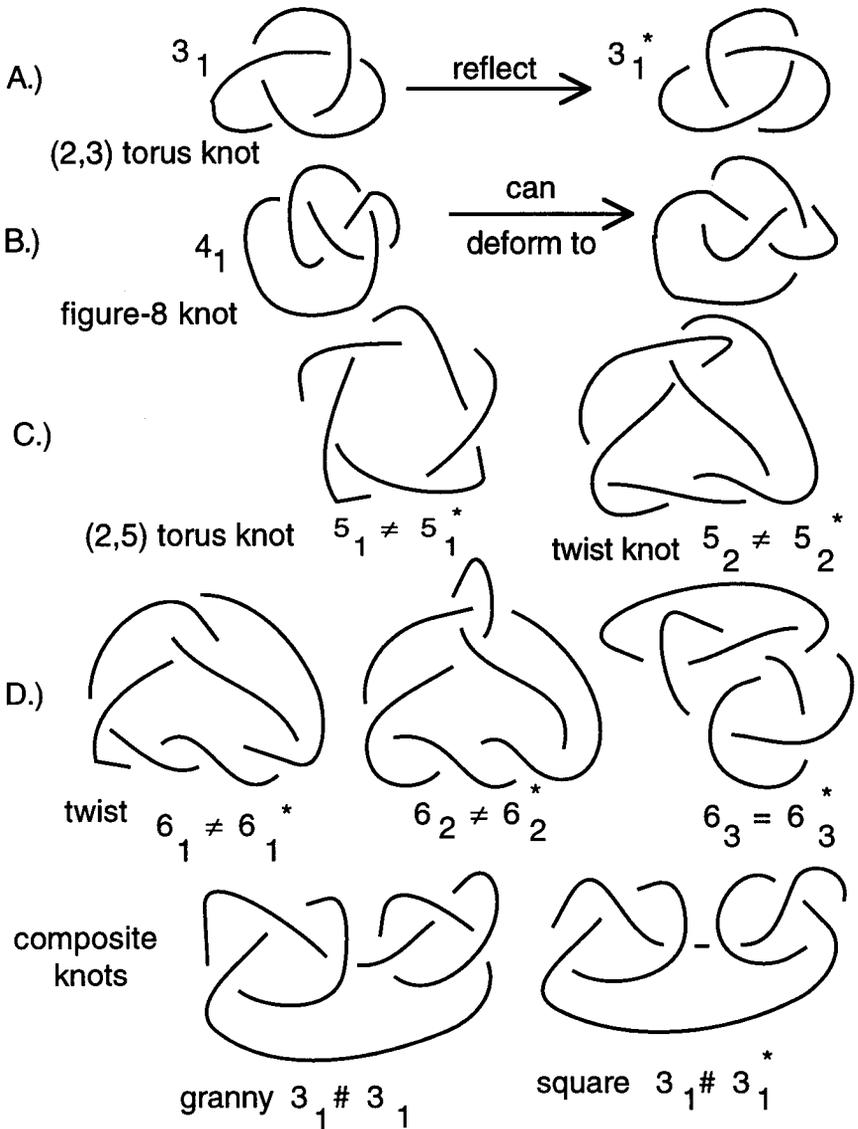
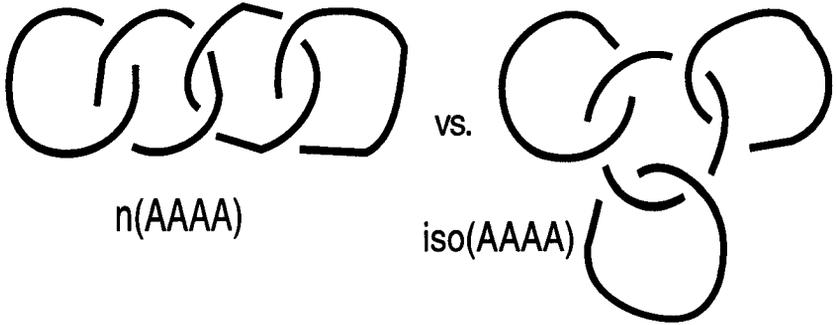


Figure 1. Knots through six crossings

A. Same size loops



B. Different size loops

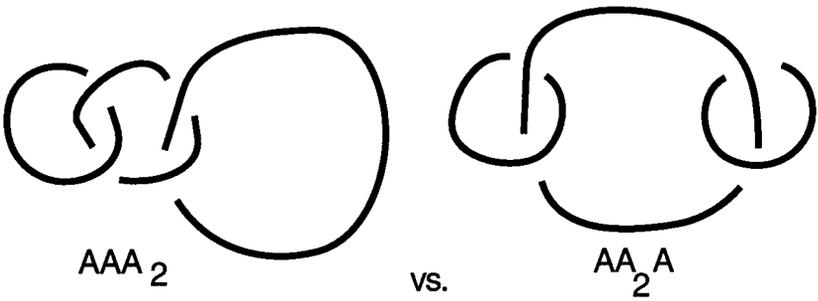


Figure 2. DNA catenanes with different gel velocities

Finally, we may note that the ranking of knots by gel velocity, energy, and (for the simplest knots, or for all knots studied if we don't make very fine distinctions) frequency, even is generally consistent with results in references 26, 34, and 37 on the minimum number of "sticks" needed to represent a given knot type (the studies differ in what restrictions are placed on the geometry of the sticks).

In view of the many correspondences, it seems reasonable to assert that our naively defined "knot energies", numerical measures of the tightness of a knot (from which one obtains a measure of the complexity of a knot-type by taking the minimum over all knots representing a given type) are succeeding in capturing whatever it is that causes the different knot types to have different gel velocities or different behavior in other physical situations.

In the sections below, we provide background for the preceding discussion and, in particular, for Table 2. In Section 2, we give a micro-introduction to knot theory, including terminology we have been using. In Section 3, we present several definitions of *energy* of a knot and describe how the data in Table 2 was obtained.

2 Knots

A *knot* is a simple closed curve in 3-space [Figure 3]. One can talk about smooth knots or about polygonal ones; in the former case, ideas of differential geometry such as total curvature²⁷ can be used to begin measuring how complicated a knot is. For polygons, one might compute total curvature or just count how many sticks are needed to represent the knot; also one might restrict to situations where all the sticks are the same length, or one might allow varying lengths. (For the polygonal energy U_{MD} , we *do* allow lengths to vary.)

Typically, knots are represented and communicated by their projections into a plane, with the graphical convention that an apparent gap in a curve indicates a place where one part of the knot is passing over/under the other relative to the direction of projection [Figure 4]. For a given smooth or polygonal knot, almost all directions yield projections with only finitely many points of singularity, all of which are all double points. The number of double points in a particular projection of a knot is called the *crossing number* of the projection.

Two knots are equivalent if one can be deformed to the other without passing the knot through itself; formally, if one is ambient isotopic to the other; in this case, we say the two knots are of the same *knot type*. So a *knot type* is an equivalence class of knots [Figure 5]. We shall use $\langle K \rangle$ to denote the knot type of a particular knot K . The *crossing number* of a knot type is the minimum, over all knots realizing that type, of the crossing numbers of the

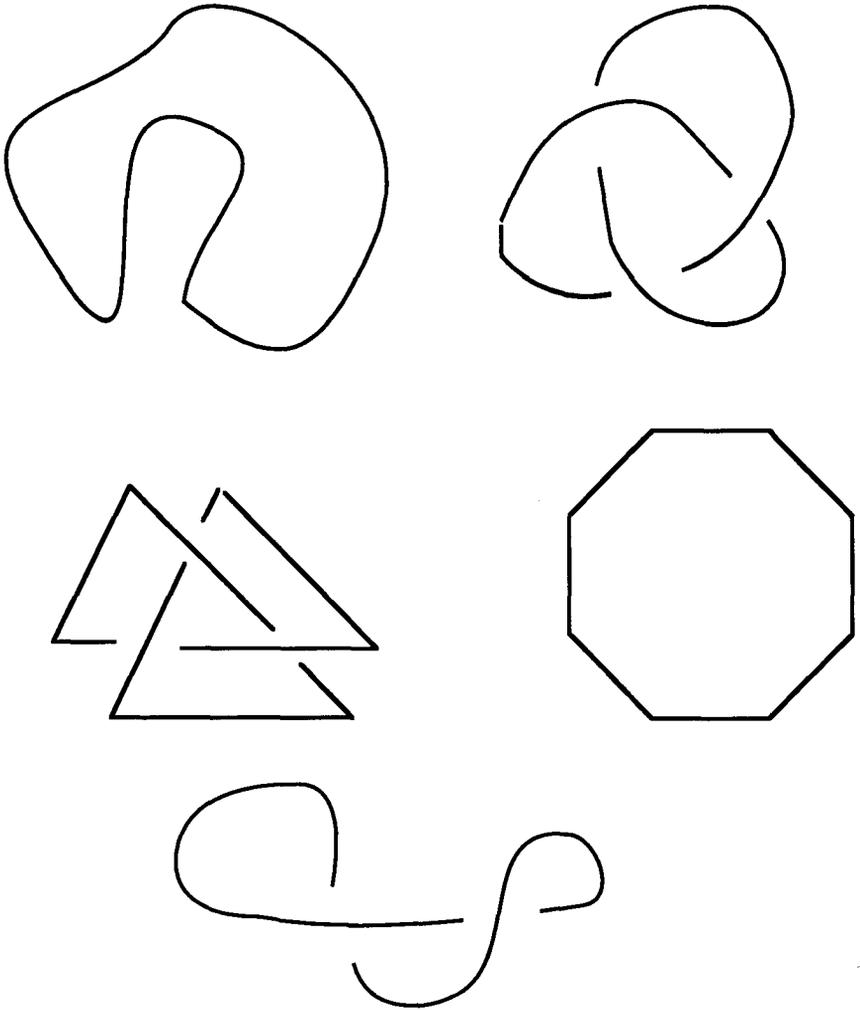
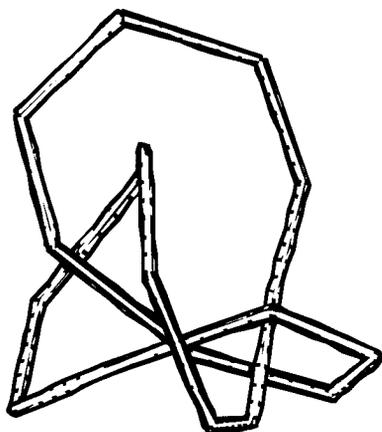
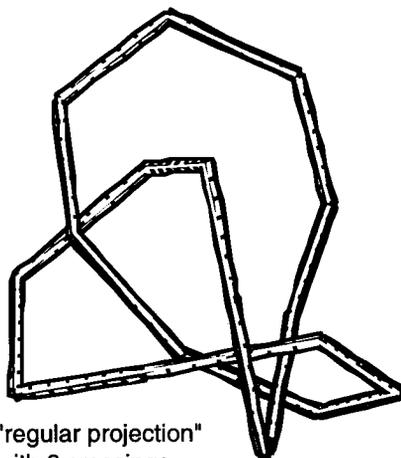


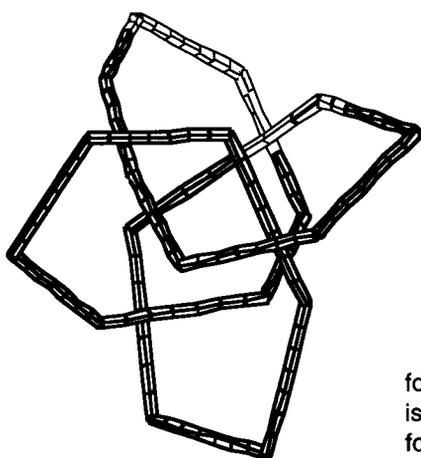
Figure 3. Smooth and polygonal knots



bad: has triple-point



"regular projection"
with 6 crossings



four crossings
is minimum
for this knot

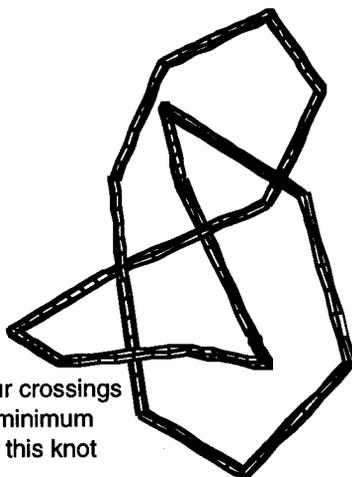


Figure 4. Different projections of the same "figure-8" knot

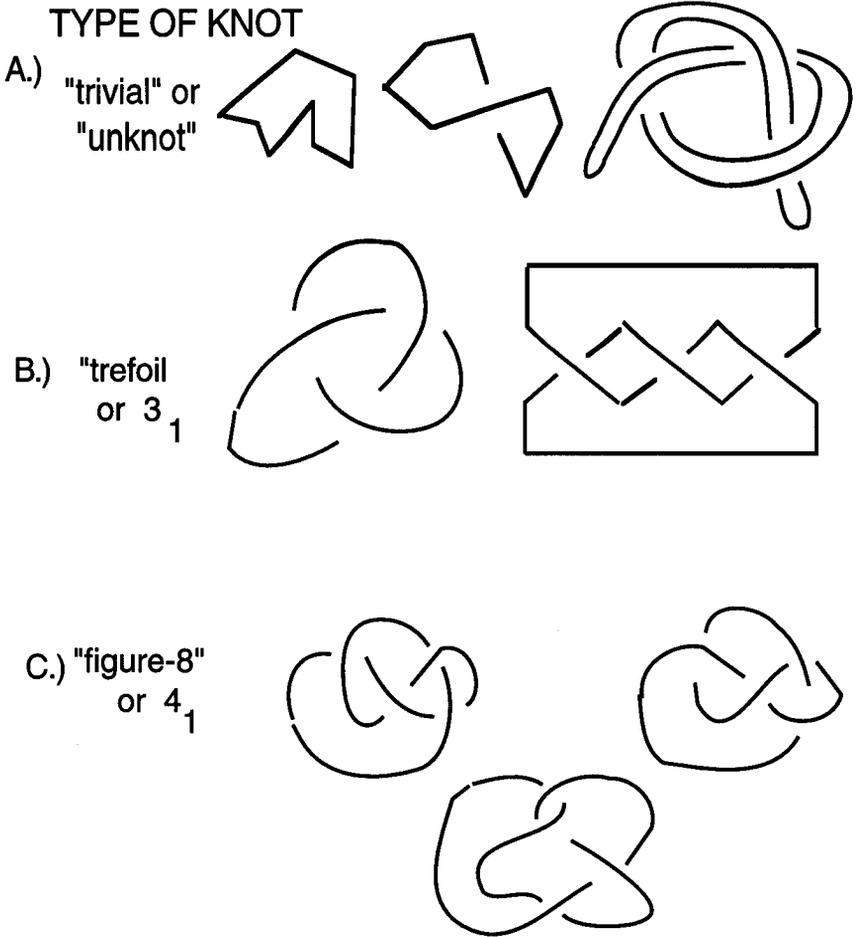


Figure 5. Several types of knots

individual knots. In DNA literature, the term *node* is used instead of crossing.

The only knot that can be drawn with zero, one, or two crossings is the *unknot* or *trivial* knot. (All the knots listed in this and the next paragraph are illustrated in [Figure 1] or [Figure 5].) There are two knot types that can be represented with a diagram having three crossings; one is the *trefoil* knot, the other just the reflection of a trefoil in the plane of projection. The trefoil type also is denoted 3_1 , while its mirror image is 3_1^* . Sometimes this distinction is ignored; we might say that 3_1 and 3_1^* are the same *knot type* but are different *ambient isotopy types*. With four crossings, we can realize knots that are trivial or trefoils, and one new kind, called the *figure-8*, and often denoted 4_1 . It is an interesting exercise to show that a (hence every) knot of the figure-8 type is equivalent to its mirror image, so there's no knot type denoted 4_1^* . At five crossings, there are four knot types, two and their respective mirror images. The knot type denoted 5_1 is one of an infinite class called *torus knots* because they can be drawn on the surface of a standard torus in space. The other five crossing knot type, denoted 5_2 , is also one of an infinite class, called *twist knots*, where the "clasp" at the top remains the same, but the number of "twists" at the bottom varies.

At six crossings, the standard knot tables list three types, 6_1 , 6_2 , and 6_3 (two of these do not interconvert with their mirror images, so we also need to record 6_1^* and 6_2^*). However, a new phenomenon appears at 6 crossings, that is *composition* of knots. The final two knots in [Figure 1] are obtained by tying, in succession, in the same string, first a knot 3_1 and then either a knot 3_1 or else a knot 3_1^* . So some knots are *composite*, others *prime*. The classical knot tables only tried to list prime knots, and just one representative for each chiral pair; even with these restrictions, the number of knot types grows rapidly with crossing number¹⁴.

3 Energy Functions

3.1 Motivation and vertex-energy

It seems most plausible to try to define a self-repelling "energy" of a knot by trying to extend the standard notion of electrostatic repelling from a finite set of charged points to a continuous charge distribution along a curve or polygon. However, this straightforward process *fails* (see e.g. reference 4). If p_1, \dots, p_n are n points in 3-space, each carrying charge $= \frac{1}{n}$, so the total charge = 1,

then the electrostatic potential of the ensemble $\{p_1, \dots, p_n\}$ is

$$\sum_{i=1}^{n-1} \sum_{j=p+1}^n \frac{\frac{1}{n} \frac{1}{n}}{\text{dist}(p_i, p_j)} . \quad (1)$$

If we imagine a single unit charge distributed evenly along a curve or line segment X , we might then try to define the potential energy to be half (because each pair (x, y) appears twice) of

$$\int_X \int_X \frac{dx dy}{|x - y|} . \quad (2)$$

Unfortunately, this integral is *infinite*; equivalently, if the points $\{p_i\}$ are spaced equally along a line segment or a circle, then the limit, as $n \rightarrow \infty$, of the sums (1) is infinite. The divergence of the integral (2) and the limit of the finite sums are on the order of $\int_0^1 \frac{1}{t} dt \approx \sum_1^\infty \frac{1}{n} \approx -\log(0)$.

To define a usable energy function along these lines, one must somehow eliminate the “near-neighbor” effects, which are what force the above integral to diverge. In the case of a polygon, we accomplish this (see below) by ignoring the interactions between a segment and itself or adjacent segments of the polygon. In the case of smooth curves, the divergent integral is “regularized” by subtracting off the (infinite) energy of a standard round circle. In fact, it turns out useful to do a similar reduction of the polygonal energy.

There are several mathematical properties that have generally been accepted as characterizing a “good” energy for knots. Here is part of the list in reference 4:

1. The energy $u(K)$ should be a real number, varying continuously with the position of the polygon or curve K .
2. The function u should be well-behaved under change of scale. (The energies U_{MD} and E_2 described below are both scale-invariant; that is, energy is a property of the *shape* of a knot rather than its absolute *size*.)
3. The function u should be defined in a way that is consistent with the idea that repelling energy varies inversely with distance, or some power of it, between repelling objects.
4. The energy u should respect the topology, in the sense that in order to deform a knot in a way that changes the knot type, that is by passing part of the string through another part, the energy has to become infinite. Alternatively put, the different knot types are separated by infinitely high

potential walls. (A different, perhaps physically more realistic, approach has been suggested by Lomonaco²³, where the energy barrier to changing knot type is finite.)

If we allow knots to flow in their configuration space, we can imagine a knot gradually deforming to reduce energy, heading in the limit to some conformation that is a (local) energy minimum. One implication of Property 4 above is that knot type cannot change during such an energy reducing flow.

Before going on to describe energies for polygons and smooth curves that have the desired mathematical properties, we should cite the seminal paper reference 17, which analyzed vertex-only potentials for polygonal knots. Let K be a polygonal knot with vertices p_1, \dots, p_n . Motivated by equation (1), Fukuhara defined an infinite family of energies,

$$U_d(K) = \sum_{i=1}^{n-1} \sum_{j=p+1}^n \frac{1}{\text{dist}(p_i, p_j)^d} . \quad (3)$$

(This formula is a little simpler than Fukuhara's, which includes a constant of proportionality that depends on the number, n , of edges of the polygon.) Fukuhara obtained the following result:

Theorem. *Consider the space K of polygonal knots of n edges, in which each edge has fixed length = 1. Let K_0 be a polygon in K with vertices p_1, \dots, p_n such that for each $i \neq j$, $\text{dist}(p_i, p_j) > 1$. Then there exists an exponent d (which depends on n) such that the knot type of K_0 remains unchanged as K_0 flows via the gradient of U_d .*

This result is surprising because the energy function only “sees” vertices, so it does not “know” if two edges of a polygon pass through each other at interior points.

Before discussing the energies reflected in Table 2, we might also note that a quite different approach is developed in reference 28, based on idealized fluid flow in knotted tubes.

3.2 Energies for smooth curves

The first knot energy satisfying Properties 1–4 above, E_2 , was proposed by J. O'Hara²⁹. It turned out that this energy has an invariance property that is stronger than being invariant under change of scale; it is, in fact, invariant under conformal transformations of 3-space³ (so E_2 is often called the *conformal energy*). Freedman and his colleagues also established¹⁵ a strong relationship between the energy E_2 of a knot K and the *average crossing number* of K . Here is their formulation of the energy E_2 .

Let K be a smooth knot in 3-space \mathfrak{R}^3 . So we can view K as the image of the standard unit circle S in the plane under a smooth map $\gamma : S \rightarrow \mathfrak{R}^3$. Then

$$E_2(K) = \int_S \int_S \left(\frac{1}{\text{dist}(\gamma(s) - \gamma(t))^2} - \frac{1}{\text{arcdist}(\gamma(s), \gamma(t))^2} \right) |\gamma'(s)| |\gamma'(t)| ds dt \quad (4)$$

Here $\text{arcdist}(s, t)$ denotes the minimum distance long the knot between points s and t .

If we assume the parametrization γ is by arclength, that is the speed $|\gamma'(s)|$ is always 1, then the definition simplifies to

$$E_2(K) = \int_S \int_S \left(\frac{1}{\text{dist}(\gamma(s) - \gamma(t))^2} - \frac{1}{\text{arcdist}(\gamma(s), \gamma(t))^2} \right) ds dt \quad (5)$$

When we compare the above energy with the polygonal energy U_{MD} defined below, we should note that E_2 counts each interaction of $\gamma(s)$ with $\gamma(t)$ twice, whereas U_{MD} counts it just once, so in comparing theorems or computational results, we should compare E_2 with $2 \cdot U_{MD}$.

The conformal energy has been studied theoretically^{3,15,16,29,30,32} and computationally^{2,19,20,31}.

The data presented in reference 20 (called E_{cos} in Table 2 above) is based on a somewhat different approach to regularizing the divergent integral. So, as stated in reference 20, the numbers E_{cos} are the minima so far observed for $E_2(K) - 4$.

Instead of “subtracting the infinity” as is done to regularize the integral in (4), another approach^{5,6} is to view the energy represented by two infinitesimal segments of the curve K as having a vector attribute, and consider only the component normal to the curve. This intuition yields another energy for smooth knots, satisfying Properties 1–4, called the *projection energy*, where the term *projection* refers to the projection of the vector, not a special invariance property of the energy function. To avoid confusing the projection energy with others discussed here, we shall use the notation U_{proj} , which is different from reference 6. As before, let K be a smooth knot in 3-space \mathfrak{R}^3 . At each point x of K there is (since K is smooth) a well-defined tangent line and, hence, a well-defined normal plane N_x . For each point $y \neq x$ of K , let θ_{xy} be the angle between the vector $(x - y)$ and the plane N_x . Then we define

$$U_{\text{proj}}(K) = \int_{y \in K} \int_{x \in K} \left(\frac{\cos(\theta_{xy})}{|x - y|} \right)^2 dy dx \quad (6)$$

The projection energy also is implemented in the program reference 2. The relationship between U_{proj} and the average crossing number of a knot is similar to that of the conformal energy, though the proof is simpler⁷.

3.3 Energies for polygons

For polygonal knots, the first energy satisfying Properties (1)–(4) was defined in reference 4. A simpler polygonal energy, and the one we shall discuss here, is presented in reference 43.

Let X, Y be disjoint line segments in \mathfrak{R}^3 . Then the minimum distance between X and Y is a positive number which we denote $MD(X, Y)$. Letting $\ell(X)$ denote the length of a segment X , define

$$U_{MD}(X, Y) = \frac{\ell(X) \cdot \ell(Y)}{MD(X, Y)^2} .$$

For a polygon K , we define

$$U_{MD}(K) = \sum_{\substack{X, Y \text{ non-consecutive} \\ \text{segments of } K}} U_{MD}(X, Y) .$$

While we are interested in knots, in particular nonplanar polygons with six or more sides, it may be helpful to work out the numbers for some simpler cases.

Example. For a triangle, there are no pairs of nonconsecutive edges, so we define $U_{MD}(\text{any triangle}) = 0$. For a plane quadrilateral, there are two pairs of edges that contribute, and the energy is minimum when the figure is a square; $U_{MD}(\text{square}) = 2 \cdot \left(\frac{a \cdot a}{a^2}\right) = 2$. For a regular pentagon, we have five pairs of edges, each pair contributing 1, so $U_{MD}(\text{regular pentagon}) = 5$. For a regular hexagon, there are six pairs of edges that contribute 1 each and three pairs of edges that contribute $\frac{a \cdot a}{(a\sqrt{3})^2} = \frac{1}{3}$ each, so $U_{MD}(\text{regular hexagon}) = 7$. For regular n -gons with $n \geq 7$, the numbers become irrational.

These examples begin to illustrate the general phenomenon that if K is any polygonal knot with six or more edges, then $U_{MD}(K) > n$.

The energy function U_{MD} satisfies Properties 1–4; in particular, it is invariant under change of scale. The function is not differentiable at some points, so we cannot talk about standard gradient flow. However, it does make sense to do a simulated annealing (“at zero degrees”) type minimization:

- Start with a polygonal knot K_0 ,

- randomly perturb the polygon to a new conformation K_1 (making sure that the perturbation is small enough that the knot type of K_0 is not changed),
- compare the energies $U_{MD}(K_0)$ and $U_{MD}(K_1)$,
- if $U_{MD}(K_0) > U_{MD}(K_1)$ then replace K_0 with K_1 .

In this way, we generate a sequence of knots converging to a knot whose energy is a (local) minimum within the knot type of K_0 . We illustrate in [Figure 6] how an 18 segment figure-8 knot evolves from an initially compressed conformation towards a (local) energy minimizing conformation under U_{MD} .

If a knot type $\langle K \rangle$ can be realized among polygons with, say, 18 edges, then there exists⁴³ a nonsingular polygon K with minimum energy among all 18 segment polygons representing knots of that type. (This is true whether the knot K is prime or composite, in contrast to the conformal energy E_2 , where minimum energy conformations are believed to exist for prime knots.) We may hope to discover (estimate) the minimum energy value for a particular knot type within a given number of segments by starting with any representative knot and letting it flow, as in the preceding paragraphs, to a local minimum. Of course one needs to worry about local vs. global minima. However, in experiments run so far, it appears that for each knot type and given number of segments, there are relatively few distinct local minima, so we have some basis for hoping that the best local minima we have found so far are indeed global minima. There's a danger of circular reasoning here, and we are beginning a more systematic analysis of local vs. global minima for U_{MD} .

In order to obtain *ground state energy* numbers for each knot type, we need to consider what happens when the same knot is made from different numbers of segments. We have two approaches for this.

It appears that as the number of segments increases, the minimum energy first decreases until an "ideal" number of segments is reached, then increases from there on. We know from reference 43 that the energy satisfies the relation $U_{MD}(K) > (\text{number of segments})$, so the energies must eventually grow; but the theorem does not predict the unmodality so far observed. In [Figure 7] we show the observed U_{MD} energy minima for trefoil knots made from 6-to-40 segments. One possible way to define the topological ground state energy of a particular knot type would be to identify the optimum number of segments and use the energy minimum for that number: So for the trefoil knot, we would say that the minimum energy is (approximately) 74.5, attained at 14 segments. This approach may prove to be a useful one for relating energy to knotting probability, since the relative probabilities are sensitive to segment numbers^{11,24,25,35,39,46,47,48}.

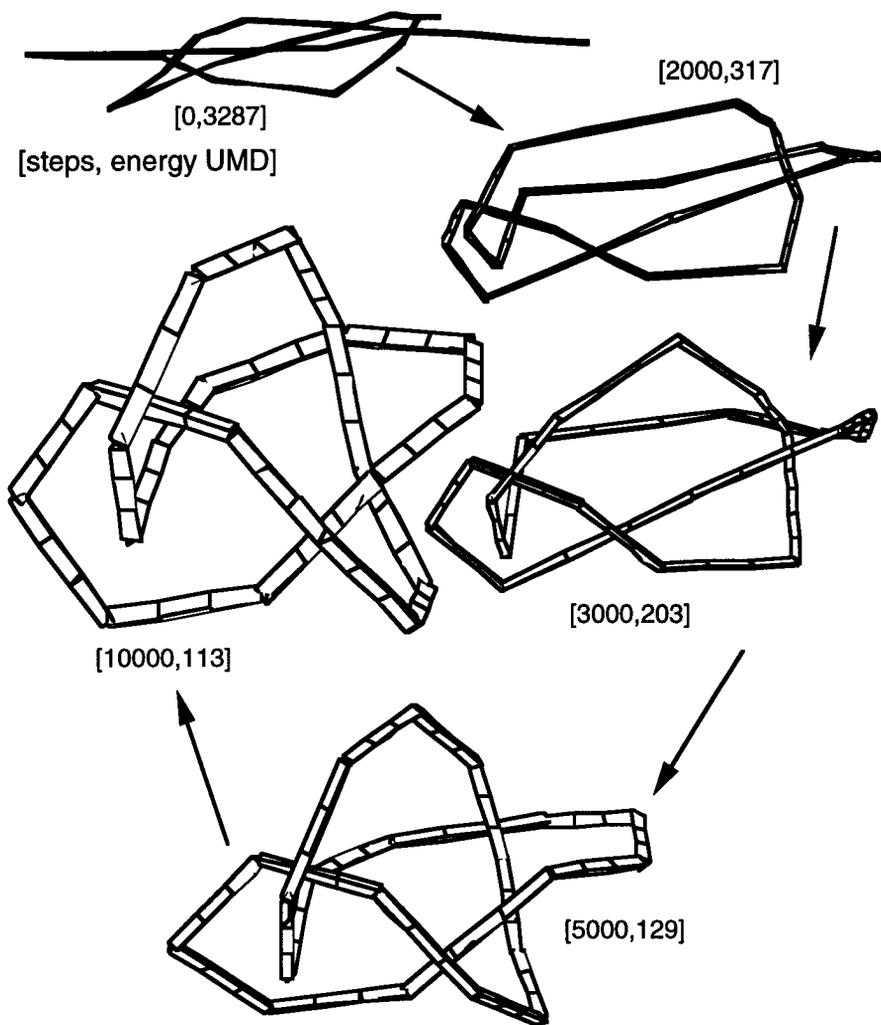


Figure 6. Evolution of a 20-segment figure-eight knot

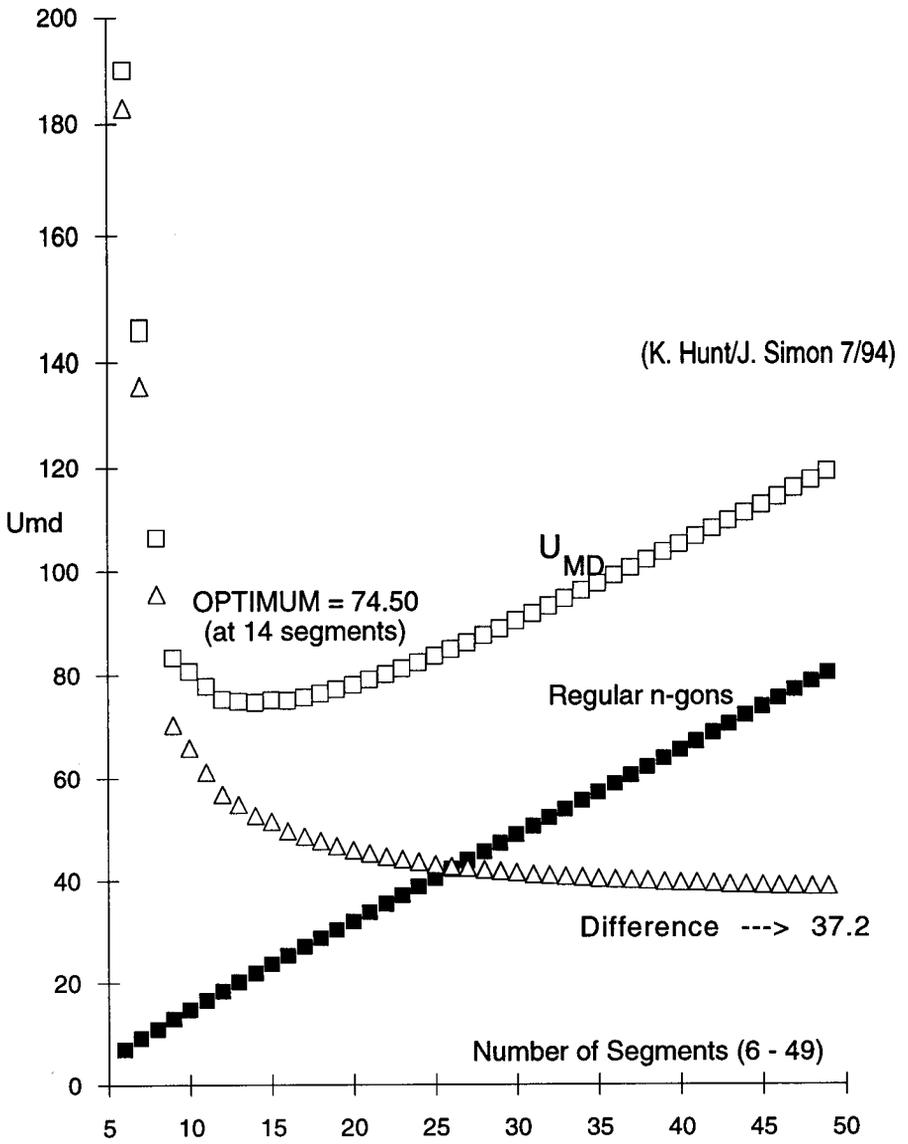


Figure 7. Observed minimum energies for trefoil knots

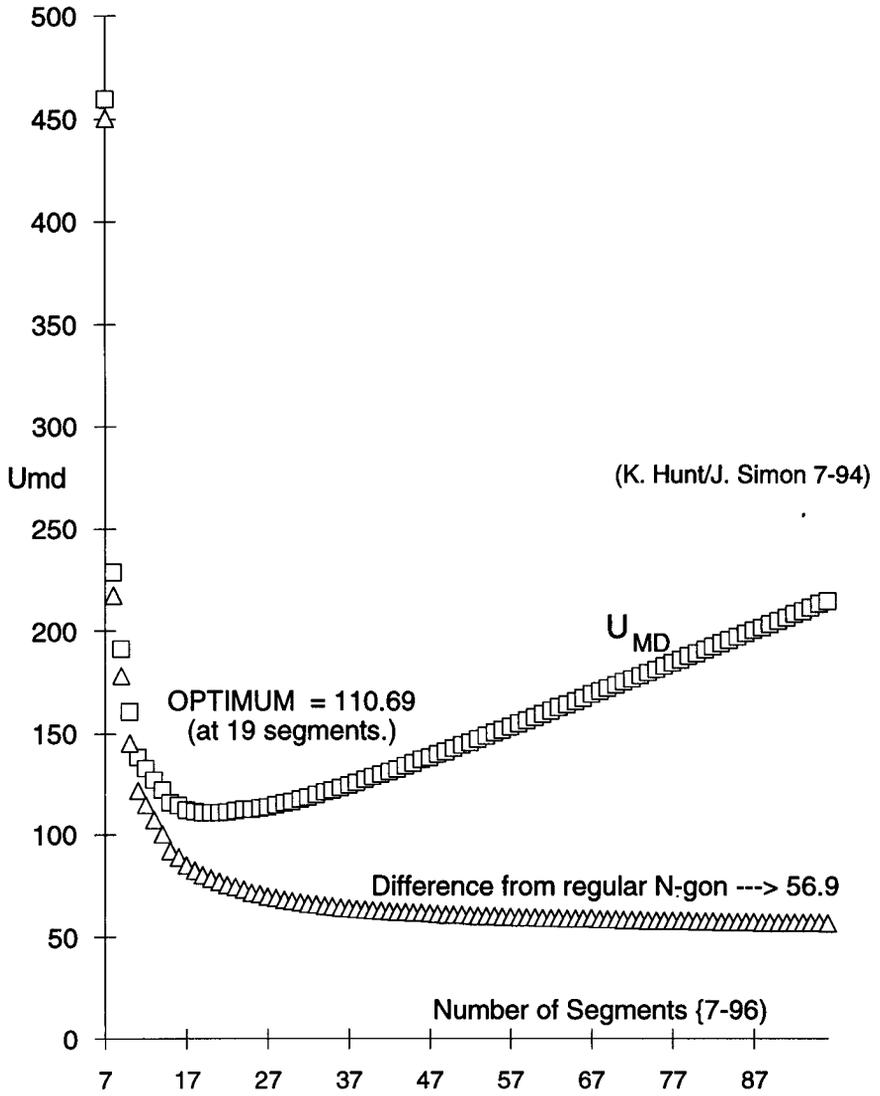


Figure 8. Observed minimum energies for figure-8 knots

Another way we can define the minimum energy for a particular knot type is to somehow remove the phenomenon that causes U_{MD} to grow with the number of segments; this is analogous to the regularization of the divergent integral 4. In [Figure 7], we have plotted, along with the U_{MD} minima for the trefoil knot, the energies of regular planar n -gons, and then the differences between the trefoil energies and the regular n -gon energies. The first observation that seems graphically obvious is that the energies of regular planar n -gons lie on a straight line; actually, there is a bit of wiggle, with odd/even oscillation, but the numbers are asymptotic to a straight line whose slope turns out³⁸ to be $\frac{\pi^2}{6}$. The second observation is that the graph of U_{MD} minima is asymptotically parallel to that straight line. (While it is “obvious” from the graphs for the trefoil knot in [Figure 7] and the figure-8 knot in [Figure 8], the fact that energy minima grow asymptotically linearly in n ($=$ number of segments) is a computational observation, not yet a theorem.) Assuming the graphs of U_{MD} minima and energies of regular n -gons are asymptotically parallel, their difference approaches a constant, and that constant is the value we can use as the topological ground state energy of the knot, \tilde{U}_{MD}^0 . This is how (referring to [Figure 7] and [Figure 8] we obtain the values $\tilde{U}_{MD}^0(\text{trefoil}) \approx 37.2$ and $\tilde{U}_{MD}^0(\text{figure-8 knot}) \approx 70.4$. (Rather than do curve fitting to a model we have yet to postulate and justify, the estimates in Table 2 are simply the values obtained from the largest segment number we have calculated for the given knot type; these segment numbers vary from knot to knot, which is one reason we expect to sharpen the table in the future.)

We need a great deal more computational data, or a theorem, to solidify the belief that all knot types exhibit the kind of convergence described above (i.e. that differences between U_{MD} minima and energies of regular planar n -gons asymptotically decrease towards a constant). But all our data so far is consistent with that assertion, so we shall continue to assume it as a working hypothesis. This approach is very appealing because it allows us to use any one computation of a minimum to estimate the limiting value, along with the belief that if one computes a minimum for a still larger number of segments, the estimate will be more accurate.

In comparing the energy minima for U_{MD} and for $\min E_{\cos}$ in Table 2, it seemed initially plausible to conjecture that $\min E_{\cos}(K) = 2\tilde{U}_{MD}^0(K) - 4$, because one might think of approximating E_2 for some smooth knot K by computing U_{MD} for an inscribed polygon and then doubling. There is excellent agreement for the trefoil, but the data so far do not support the conjecture for other knots. The provocative similarity in how the two energies rank knots remains a challenge to further study.

Acknowledgements

This work was partially supported by NSF DMS-9407132. Kenneth Hunt's programs have been essential for computations; also he was the first to observe the unimodality (with segment number) of energy minima of U_{MD} for a given knot type. Aaron Trautwein performed numerous computational experiments and determined that prime knot types of 6 through 9 crossings are ordered almost identically by \tilde{U}_{MD}^0 and $\min E_2$. Many colleagues have been generous with their enthusiasm, ideas, and knowledge: special thanks are due to C. Benham, G. Buck, N. Cozzarelli, E.J.J. VanRensburg, D. Roseman, S. Spengler, A. Stasiak, N. Stellwagen, and J. Sullivan.

References

1. H.W. BENJAMIN, M.M. MATZUK, M.A. KRASNOW, AND N.R. COZZARELLI, *Recombination site selection by Tn3 resolvase: topological tests of a tracking mechanism*, Cell **40** (1985), 147–158.
2. K. BRAKKE, *Surface Evolver Manual* (v. 1.92), Minnesota Geometry Center, Research Report **GCG55**, July 1993.
3. S. BRYSON, M.H. FREEDMAN, Z.X. HE, AND Z. WANG, *Mobius invariance of knot energy*, Bull. Amer. Math. Soc. **28** (1993) 99–103.
4. G. BUCK AND J. SIMON, *Knots as dynamical systems*, Topology Appl. **51** (1993) 229–246.
5. G. BUCK AND J. ORLOFF, *A simple energy function for knots*, preprint 12/92 (to appear, Topology Appl.).
6. G. BUCK, *Random knots and energy: Elementary considerations*, Journal of Knot Theory and its Ramifications, **3** (1994), 355–364.
7. G. BUCK, *The projection energy bounds crossing number*, preprint 10/93.
8. G. BUCK AND J. ORLOFF, *Computing canonical conformations for knots*, Topology Appl. **51** (1993) 246–253.
9. N. COZZARELLI, pers. comm. 7/94.
10. F.B. DEAN, A. STASIAK, T. KOLLER, AND N.R. COZZARELLI, *Duplex DNA knots produced by escherichia coli topoisomerase I*, Journal of Biological Chemistry **260** (1985), 4975–4983.
11. T. DEGUCHI AND K. TSURUSAKI, *A statistical study of random knotting using the Vassiliev invariants*, Journal of Knot Theory and its Ramifications, **3** (1994), 321–353.
12. P. DRÖGE AND N.R. COZZARELLI, *Topological structure of DNA knots and catenanes*, Methods in Enzymology **212** (1992), 120–130.

13. P. DRÖGE AND N.R. COZZARELLI, *Recombination of knotted substrates by Tn3 resolvase*, Proc. Natl. Acad. Sci. USA **86** (1989), 6062–6066.
14. C. ERNST AND D.W. SUMNERS, *The growth of the number of prime knots*, Math. Proc. Camb. Phil. Soc. **102** (1987), 303–315.
15. M. FREEDMAN AND Z.-X. HE, *Divergence free fields: Energy and asymptotic crossing number*, Ann. Math. **133** (1991) 189–229.
16. M. FREEDMAN, X. HE, [AND Z. WANG], *On the “energy” of knots and unknots*, preprint 12/91 [12/92].
17. S. FUKUHARA, *Energy of a knot*, Fete of Topology (Matsumoto et al., Eds.), Academic Press, 1988, 443–451.
18. R. KANAAR, A. KLIPPEL, E. SHEKHTMAN, J.M. DUNGAN, R. KAHMANN, AND N.R. COZZARELLI, *Processive recombination by the phage Mu Gin system: Implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action*, Cell **62** (1990), 353–366.
19. D. KIM AND R. KUSNER, *Torus knots extremizing the Möbius energy*, Experimental Math. **2** (1993), 1–9.
20. R. KUSNER AND J. SULLIVAN, *Möbius energies for knots and links, surfaces and submanifolds*, preprint 4/94.
21. H.A. LIM AND E.J. JANSE VAN RENSBURG, *A numerical simulation of electrophoresis of knotted DNA*, Supercomputer Computations Research Institute (Florida State Univ.), report #FSU-SCRI-91-163, to appear J. Modelling Sci. Comput., Oxford.
22. H.A. LIM, M.T. CARROLL, AND E.J. VAN RENSBURG, *Electrophoresis of knotted DNA in a regular and random electrophoretic medium*, Biomedical Modeling and Simulation (J. Eisenfeld et al., Eds.), Elsevier Science Pub., 1992, 213–223.
23. S. LOMONACO, *The modern legacies of Thompson’s atomic vortex theory in classical electrodynamics*, Amer. Math. Soc. Proc. in Appl. Math., to appear; also pers. comm., preprint 1994, and talks at several meetings.
24. J.P.J. MICHELS AND F.W. WIEGEL, *On the topology of a polymer ring*, Proc. Roy. Soc. A **403** (1986), 269–284.
25. J.P.J. MICHELS AND F.W. WIEGEL, Phys. Lett. **90A** (1982), 381–384.
26. K. MILLETT, *Knotting of regular polygons in 3-space*, Journal of Knot Theory and its Ramifications, **3** (1994), 263–278.
27. J. MILNOR, *On the total curvature of knots*, Ann. Math. **52** (1950) 248–257.
28. K. MOFFATT, *The energy spectrum of knots and links*, Nature **347** (Sept. 1990) 367–369.
29. J. O’HARA, *Energy of a knot*, Topology **30** (1991) 241–247.

30. J. O'HARA, *Family of energy functionals of knots*, *Topology Appl.*, **48** (1992) 147–161.
31. J. O'HARA, *Energy functionals of knots*, *Topology-Hawaii* (K.H. Doverman, Ed.), (Proc. of 1991 conference), World Scientific, 1992, 201–214.
32. J. O'HARA, *Energy functionals of knots II*, *Topology Appl.* **56** (1994), 45–61.
33. N. PIPPERGER, *Knots in random walks*, *Disc. Appl. Math.* **25** (1989), 273–278.
34. R. RANDELL, *An elementary invariant of knots*, *Journal of Knot Theory and its Ramifications*, **3** (1994), 279–286.
35. E.J. JANSE VAN RENSBURG AND S.G. WHITTINGTON, *The knot probability in lattice polygons*, *J. Phys. A: Math. Gen.* **23** (1990) 3573–3590.
36. E.J. JANSE VAN RENSBURG AND S.G. WHITTINGTON, *The dimensions of knotted polygons*, *J. Phys. A: Math. Gen.* **24** (1991) 3935–3948.
37. E.J.J. VANRENSBURG AND S.D. PROMISLAW, *Minimal knots in the cubic lattice*, preprint 12/93.
38. At the July 1994 IMA conference, after the talk which included the data and computationally estimated slope of 1.647, several colleagues (including E.J.J. VanRensburg, A. Stasiak, and J. Sullivan) suggested that the limiting slope was, in fact, $\frac{\pi^2}{6}$, and this has been verified.
39. V.Y. RYBEKOV, N.R. COZZARELLI, AND A.V. VOLOGODSKII, *Probability of DNA knotting and the effective diameter of the DNA double helix*, *Proc. Natl. Acad. Sci. USA* **90** (1993), 5307–5311.
40. S.Y. SHAW AND J.C. WANG, *Science* **260** (1993) 533.
41. S.Y. SHAW AND J.C. WANG, *DNA knot formation in aqueous solutions*, *Journal of Knot Theory and its Ramifications*, **3** (1994), 287–298.
42. S. SPENGLER, A. STASIAK, AND N.R. COZZARELLI, *The stereostructure of knots and catenanes produced by phage λ integrative recombination: implications for mechanism and DNA structure*, *Cell* **42** (1985), 325–334.
43. J.K. SIMON, *Energy functions for polygonal knots*, *Journal of Knot Theory and its Ramifications*, **3** (1994), 299–320.
44. In a conversation during the July 1994 IMA conference, Andrzej Stasiak suggested the appealing term *in virtuo* to describe computer experiments.
45. DEWITT SUMNERS AND S.G. WHITTINGTON, *Knots in self-avoiding walks*, *J. Phys. A: Math. Gen.* **21** (1988), 1689–1694.
46. M.C. TESI, E.J. JANSE VAN RENSBURG, E. ORLANDINI, AND S.G. WHITTINGTON, *Knot probability for lattice polygons in confined geometries*, *J. Phys. A: Math. Gen.* **27** (1994) 347–360.
47. M.C. TESI, E.J. JANSE VAN RENSBURG, E. ORLANDINI, D.W. SUMNERS, AND S.G. WHITTINGTON, *Knotting and supercoiling in circular*

- DNA: A model incorporating the effect of added salt*, Phys. Rev. E: **49** (1994) 868–872.
48. A.V. VOLOGODSKI, A.V. LUKASHIN, M.D. FRANK-KAMINETSII, AND A.V. AHSHELEVICH, *The knot probability in statistical mechanics of polymer chains*, Sov. Phys. JETP **39** (1974), 1059–1063.
49. S. WASSERMAN AND N.R. COZZARELLI, *Supercoiled DNA-directed knotting by T4 topoisomerase*, Journal of Biological Chemistry **266** (1991), 20567–20573.
50. S. WHITTINGTON, *Topology of polymers*, in *New Scientific Applications of Geometry and Topology* (D.W. Sumners, Ed.), Amer. Math. Soc. PSAM **45**, 1992, 73–95.

CHAPTER 11

PHYSICALLY-BASED STOCHASTIC SIMPLIFICATION OF MATHEMATICAL KNOTS

Robert P. Grzeszczuk
Silicon Graphics, Inc., rg@sgi.com

Milana Huang
MultiGen, Inc., milanah@multigen.com

Louis H. Kauffman
*Department of Mathematics, University of Illinois at Chicago,
kauffman@uic.edu*

The article describes a tool for simplification and analysis of tangled configurations of mathematical knots. The proposed method addresses optimization issues common in energy-based approaches to knot classification. In this class of methods, an initially tangled elastic rope is “charged” with an electrostatic-like field which causes it to self repel, prompting it to evolve into a mechanically stable configuration. This configuration is believed to be characteristic for its knot type. We propose a physically-based model to implicitly guard against isotopy violation during such evolution and suggest that a robust stochastic optimization procedure, simulated annealing, be used for the purpose of identifying the globally optimal solution. Because neither of these techniques depends on the properties of the energy function being optimized, our method is of general applicability, even though we applied it to a specific potential here. The method has successfully analyzed several complex tangles and is applicable to simplifying a large class of knots and links. Our work also shows that energy-based techniques will not necessarily terminate in a unique configuration, thus we empirically refute a prior conjecture that one of the commonly used energy functions (Simon’s) is unimodal. Based on these results we also compare techniques that rely on geometric energy optimization to conventional algebraic methods with regards to their classification power.

1 Background

The focus of this paper is the knot classification problem, which can be intuitively formulated as follows: take a piece of string tied in an arbitrary manner whose ends were then pasted together and simplify it as much as possible. It has been conjectured that such canonical configuration exists and is unique for the knot’s class and can thus be used to distinguish among different knot types. Specifically, we are interested in determining if a particular tangle can be simplified to a circle, in which case it can be shown not to have been knotted in the first place (the unknot detection problem). Similarly, given a tangled

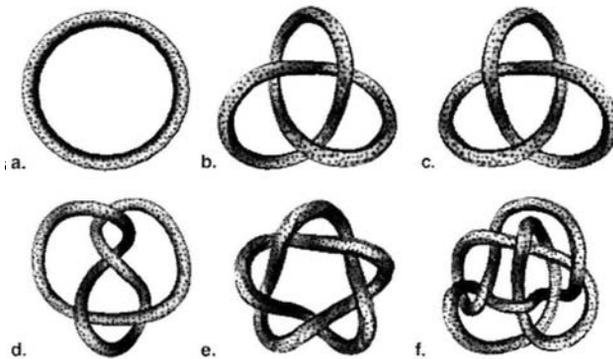


Figure 1: Knot gallery: a) unknot b) left trefoil c) right trefoil d) figure eight knot e) star knot f) a knot identified as 9_{33} in the Reidemeister's table.

initial configuration that is knotted, we would like to be able to reduce it to a state that is characteristic for the knot's type (the general knot classification problem). Finally, given two tangled configurations, we would like to be able to establish if they represent a knot of the same type (the knot equivalence problem). In the course of the work, features of space curves such as energy distribution and their local stability are also visualized. Aside from its theoretical appeal, the method promises to solve practical problems common in genetic research and polymer design.

1.1 Knot Theory

A mathematical knot is a closed curve in 3-dimensional space that does not self-intersect and has no thickness¹². Two knots, K_1 and K_2 , are said to be equivalent if there exists an ambient isotopy that continuously deforms K_1 into the shape of K_2 without K_1 self-intersecting or breaking. If no such deformation exists then K_1 and K_2 are of two distinct types. The simplest type of knot is the unknot, or trivial knot. This type of knot is equivalent to the unknotted circle, $x^2 + y^2 = 1, z = 0$, and is shown in Fig. 1a which also displays a selection of common knot types. A knot is said to be chiral if it is topologically distinct from its mirror image; otherwise, it is achiral. For example, a trefoil knot T (Fig. 1b) is chiral and thus distinct from its mirror T^* (Fig. 1c), while a figure-eight knot (Fig. 1d) is achiral.

1.2 Knot Classification Methods

Despite being the center of attention for many years, the problem of establishing knot equivalence still lacks a computationally infallible technique to handle the general case. Several solutions varying in the ability to distinguish among knot types have been proposed. These solutions can be roughly divided into three broad categories: 2D transformation techniques (in which a planar projection of the knot is manipulated isotopically), algebraic methods (which build an homomorphic abstract model of the knot from its planar projection and attempt to simplify that model instead), and 3D transformation techniques (where the knot is manipulated in its native space).

An approach based on Reidemeister moves²¹ is representative of the first class of methods. A Reidemeister move is one of three rules that changes the crossings of a planar knot projection in a local area while preserving the knot type. Reidemeister proved that two knots in three-dimensional space can be deformed continuously, one into the other, if and only if any diagram obtained from by projection to a plane of one knot can be transformed into any diagram of the other via a sequence of type-preserving 2D moves he proposed. However, while many heuristics have been proposed, a definitive method to determine the minimal, or even correct, sequence of moves to deform one knot to another is still an open problem.

An alternative approach to knot classification is not to transform the knot representation (e.g., crossing diagram) directly but instead to cast the problem into a more familiar domain. For example, it is possible to construct a polynomial that preserves some topological properties of the knot as it undergoes algebraic manipulations. Such a polynomial is calculated from the planar knot projection. A set of variables with powers and coefficients are produced from the arrangement of crossings of the projection. Different projections of the same knot produce the same polynomial, therefore such a polynomial is an invariant of ambient isotopy. Such an invariant can be used to classify knots. If two knots are of the same type then the polynomials calculated for two knots are the same. One of the first polynomial invariants to gain popularity was developed by J.W. Alexander¹. Alexander's polynomial fails to distinguish between all types of knots. Therefore it is not an invariant in a pure sense. For example, it does not detect chirality—it cannot tell a trefoil from its mirror image. Since that time, a number of more powerful polynomial invariants have been discovered and include the Jones and Kauffman¹² polynomials. However, these do not distinguish all knot types either.

The third class of techniques establishes knot equivalence by directly transforming one knot onto the other through a series of isotopic, type-preserving,

deformations in their native, three-dimensional, space. A scalar value, energy, is typically associated with a knot as a measure of similarity between distinct configurations. Typically, the energy will depend on the geometry of the knot and will vary as the knot undergoes deformations and as such is not useful as an invariant of ambient isotopy. However, it has been suggested⁴ that the value of energy of the configuration that minimizes the given functional may be such an invariant and that the minimal energy configuration is a standard form of the knot. If this is indeed true, the problem of knot classification can be mapped to an optimization problem in which a configuration that minimizes the given energy functional (or the minimal energy itself) is sought. It is worth noting that such configuration may depend on the functional chosen as well as the knot representation used (e.g., polygonized or continuous).

A number of energy functions have been proposed for knot invariance. The idea of associating an electrostatic energy to a knot was proposed by Fukuhara⁷. Electrons are placed at equal intervals along a polygonal knot thus causing it to self-repel into a configuration that is a simplification of the original knot. O'Hara and Freedman propose an energy function based on electrostatic repulsion and prove that, given such energy, there exists a finite number of knot types below a given energy value¹⁸. Buck and Simon show with their purely geometric "minimal distance" energy function there exists a global energy minimum which is an invariant of knot type. Freedman, He, and Wang⁶ give an energy function and prove that their energy bounds the average crossing number⁶. Therefore, some energy bounds the number of knot types. They also show that the minimum energy configuration of any closed curve is a circle. Kusner and Sullivan¹⁵ and also Bryson et.al.³ experiment with repulsive energy potentials between pairs of line segments or vertices of the knot. Fig. 6 shows how their energy is invariant under conformal (Möbius) transformations of space.

1.3 Optimization Techniques

Optimization problems, where a minimum (or maximum) of some scalar cost functional is desired, are often solved by iterative improvement. In this approach, at any stage of the process, a change to the current state of the system is proposed (e.g., one of the independent variables is perturbed). If such a change results in an improvement of the evaluated cost function, then the change is accepted as the new state. This is iteratively applied until changes no longer give lower cost values. This is a greedy algorithm, allowing only transitions that result in improvement. The changes to the system can be introduced methodically or at random. Gradient descent¹⁹ and conjugate

gradient¹⁹ methods are examples of deterministic methods that follow this concept and are the method of choice for the problem at hand. For example, Fukuhara⁷ and Buck and Simon⁴ both evolve the knot along the gradient of the energy potential. On the other hand, Simon²⁴ uses a nondeterministic descent method by performing small random perturbations of the knot and accepting the move if it lowers the energy. Opportunistic methods such as these do not perform well in the presence of local minima as they will stop at the first minimum found. In addition, gradient based methods tend to converge at a prohibitively slow pace for functionals which are locally flat (i.e., contain plateaus of nearly zero gradient strength).

It is possible to relax the strict opportunistic strategy of greedy iterative improvement. This strategy is the approach favored by Wu²⁵ who employs noisy function evaluations (i.e., a user-specified fixed percentile of uphill moves are accepted). Such *ad hoc* schemes permit escaping from some local minima. However, selecting a suitable fraction of such uphill moves is rather difficult. If too few are permitted, the system can only escape very shallow local minima; too many, and the search becomes a purely random sampling of a prohibitively large search space. In contrast, the research presented here and earlier in⁹, employs an optimization technique that is resilient to getting stuck in local minima through its methodical use of stochastic sampling. An important theoretical advantage of simulated annealing (SA) over pure stochastic sampling is that the former has global convergence proofs to attest to its ability to find the global minimum. An even more critical practical advantage that sets simulated annealing apart from hill climbing techniques is that it performs a “blind” search that does not depend on the strength (or even existence) of the energy gradient. For example, flat plateaus are explored quite vigorously and constitute much less of a problem.

Applying simulated annealing to minimize knot energies has been independently researched by Ligocki¹⁶. Ligocki’s work differs from the work presented here in three significant ways. First, the energy functions evaluated are different. More importantly, the perturbation methods differ. Most notably, Ligocki uses a fixed move size distribution (his geometric perturbation technique allows only one vertex to be moved at a time) requiring a prohibitively slow logarithmic cooling schedule. This forces Ligocki to compromise the theoretical requirements of the conventional annealing and resort to “simulated quenching,” which no longer guarantees convergence to the global minimum. Finally, while Ligocki performs explicit geometric checks in order to avoid isotopy violation, we devised a physically-based model which implicitly prevents the curve from crossing over itself.

The remaining part of the paper is structured as follows. Section 2 de-

scribes the implementational details of the proposed algorithm, including two physically based methods for randomly perturbing curves. Then, a series of experiments are presented in order to illustrate the behavior of the method as well as to verify our findings empirically. Two complex unknots and a complex trefoil are simplified to their canonical configurations. Based on these results we show that one of the commonly used functionals (Simon’s minimal distance energy²⁴) is not unimodal and we comment on the power of the energy-based approach when it comes to distinguishing among arbitrary knots. Although, as we point out, there may not completely disqualify energy-based approaches. Further, we show that Simon’s energy has analytical and numerical properties that make it difficult to optimize with deterministic techniques that are based on gradient descent.

2 Methods

In the solution that we propose, the curve representing the knot evolves in a self-repelling potential field generated by the functional associated with it under the guidance of the simulated annealing algorithm. Simulated annealing¹⁴ is a non-deterministic technique that operates within an iterative framework similar to methods described earlier. The system evolves from state to state through a series of local perturbations. Configurations that lower the energy of the system are always accepted, resulting in consistent downhill behavior. However, unlike the case of “greedy” iterative improvement, occasional uphill moves are also accepted. The probability p of accepting such an uphill move as the new configuration varies during the course of the procedure. p is related to a user-adjustable parameter of the system called temperature through the Boltzmann probability factor:

$$p(\Delta E) = \exp(\Delta E/kT) \quad (1)$$

where ΔE is the energy change of the system due to perturbation, k is Boltzmann’s constant and T is temperature. The equation comes from thermodynamics and describes the energy state of atoms at thermal equilibrium and their probable behavior as they fluctuate about the average behavior of the system. Initially, the temperature of the system is high (we say that the system is “melted”) and uphill transitions are accepted often, but as the system “cools down,” fewer such transitions are accepted. Finally, when the system “freezes,” only down-hill transitions are accepted. Therefore, the system explores the search space in a nearly unconstrained fashion during the early stages of the search slowly narrowing its focus to the more promising areas. In classical annealing¹⁴ the temperature of the system was lowered logarithmically.

A much more efficient algorithm was suggested by Shu²³, who proved that an exponential schedule can be used if the moves are drawn from Lorentzian rather than Gaussian distribution. In either case, such sequence of stochastic perturbations forms a Markov Chain⁵, which has a proof of statistical convergence to a global minimum¹⁴. That is, one can estimate the probability of convergence to a global minimum as a function of the annealing parameters.

To implement simulated annealing, one needs the following four elements. First, a concise description of possible states of the system is required. The second element is a scalar objective function that is applied to the system state. This function is the energy function to be minimized. Third, one needs a procedure to introduce random perturbations to the system. Such a procedure, called a “move generator,” has to produce valid states and be ergodic (i.e., there has to exist a sequence of perturbations that transforms any two given configurations onto each other). In addition, a move generator has to provide some control over the sizes of the moves so that rapidly converging algorithms can be used. Lastly, an annealing schedule is needed that regulates the changes in system state. The starting (melted) temperature of the system, rate of temperature decrease (cooling), and number of iterations per time step are part of the schedule.

2.1 State Description and Visualization

In the proposed method, the knot is manipulated in its native 3D space and the system configuration is described by a list of n vertices. For the sake of visualization the knot is represented by a generalized cylinder with a spline-interpolated curve as the axis. An alternative to fully rendering such a model in 3D is a diagrammatic representation of all the crossings. Unfortunately, both methods tend to result in very cluttered scenes for all but the simplest configurations. A number of ways to aid in discerning complex geometry and structure were recently suggested by, among others, Interrante¹⁰ and Rheingans⁷. In their approaches, a highly structured texture is applied to the model in order to provide additional features as an aid in scene interpretation. While originally applied to resolving ambiguities inherent in transparent surface rendering, these methods are also applicable to opaque scenes where clutter deemphasizes the effects of shading and occlusion. This technique was adopted in Fig. 1 and 6.

In some situations, we may be interested in a distribution of certain features along the curve rather than the structure of the the curve itself. In such situations, we will use a model that is based on the knot’s Gauss code (or trip code)^{8, 20}, and¹² which is a particularly simple transcription of the knot’s

structure in a plain linear form. In our adaptation, a Gauss model of a knot will be defined to be a circle that is created by enumerating all the vertices of the polygonized knot in the order of traversal (see Fig. 8). Note that any space curve can be represented this way even though it may be knotted. Also, this type of “space unfolding” has a very different meaning than “knot un-tangling” which is the subject of this paper. In addition, we will use color to encode selected local properties of knot configurations. For example, we will assign colors based on the potential energy of a link which is useful for identifying “tight spots” (edges in tight areas will necessarily have high energies). Alternatively, we will color the edges by the strength of the energy gradient (i.e., the magnitude of the self-repulsive force). This method will be useful in investigating local stability issues as edges which are in mechanically stable configurations have zero-strength resultant forces acting upon them.

2.2 Energy Functional

There has been a number of simplifying energy functions that were proposed for knot classification. In our implementation we decided to use a function due to Simon²⁴ which we will refer to as E_{MD} here. The energy is defined on a polygonal knot K of n consecutive edges e_1, e_2, \dots, e_n . It is a summation over all non-neighboring pairs of edges, defined as follows. Let $len(e_i)$ be the length of edge e_i ; and let $md(e_i, e_j)$ be the minimum distance between non-neighboring edges e_i, e_j :

$$E_{MD}(K) = \sum_{i,j <= n}^n \frac{len(e_i)len(e_j)}{md(e_i, e_j)^2}. \quad (2)$$

The energy function defined by (2) has a number of advantageous mathematical properties. Simon states that for each knot type represented as a polygonal knot, there does exist a configuration that minimizes $E_{MD}(K)$. $E_{MD}(K)$ is scale invariant causing energy to be dependent on the knot’s shape, not its size. On the other hand, $E_{MD}(K)$ is characterized by a non-linear, asymptotic dependency on the number of nodes²⁴ which makes it necessary to assure that knots be represented by the same number of vertices for the sake of comparison. One (admittedly, fairly academic) drawback of the above functional is that it does not have a continuous derivative and can potentially cause problems with gradient-based approaches. Conveniently, this has no effect on simulated annealing which has non-analytic nature.

It is important to notice that Simon’s energy is just one of many such functions commonly used for the purpose of knot classification, and, arguably,

not the most promising one. In particular, it is possible that the results presented here would be different if a different energy function was tested (i.e., it is possible that there exists a unimodal energy function). The main practical consideration for choosing Simon's energy for this work was availability of tools by Wu²⁵ that were used to evaluate our results with respect to deterministic methods. However, because our method in no way depends on the properties of the functional being optimized it can be used as a general purpose tool and can thus be applied to investigate other functions as well.

2.3 *Perturbation Methods*

One of the most critical technical components of the simulated annealing method is the move generator whose purpose is to propose a change to the system by taking the current system configuration and perturbing it slightly (i.e., it will select a random section of the knot consisting of a random number of vertices and replace it with a slightly different section). It is important that such a move generator maintains isotopic property (i.e., it does not change the type of the knot by allowing cross-overs). One way to perturb a knot is to use purely geometric transformations (e.g., rotate segments randomly). Because such deformations tend to result in crossovers, explicit checks are used in order to reject invalid configurations. This is likely to result in significantly increased computational effort, particularly in tight areas of the knot. An alternative that we adopt in this method employs a physically-based model in order to deform the knot. In this paradigm, the knot is modeled as a physical system whose properties do not allow it to self-penetrate.

Two different perturbation methods are introduced here. During the annealing process the system randomly chooses which perturbation method to apply with equal probabilities. They both treat the knot as a physically based model by associating forces such as electrostatic repulsion and elasticity with the knot. These forces are chosen for their opportunistic qualities. Electrostatic repulsion amongst the edges of the polygonal knot helps push the knot apart, while also maintaining ambient isotopy by not allowing the knot to cross over itself. Spring forces help keep the length of each edge from becoming infinitely long due to the electrostatic repulsive force that is also applied to the edges. It should be noted that our use of electrostatic and elastic forces for the sake of implementing the move generators in no way determines the actual energy associated with the knot, which may not have a physical interpretation (compare section 2.2).

Efficient implementation of the simulated annealing process requires that some form of control over the size distributions for the proposed perturbation

methods is provided²³. Both methods presented here allow for such control. The lengths (i.e., number of nodes) of the perturbed segments as well as the time of their physically-based evolution are random variables with Lorentzian distributions. The Lorentzian distribution, also referred to as a Cauchy distribution, is defined as:

$$P(x, \mu, \Gamma) = \frac{1}{\pi} \frac{\Gamma/2}{(x - \mu)^2 + (\Gamma/2)^2} \quad (3)$$

where μ is the mean and Γ is the half-width² of the move size distribution. The Lorentzian distribution is chosen over the Gaussian distribution because probability does not diminish to zero away from the mean as quickly. Therefore, the rapid cooling schedules, which tend to strand the process in local energy minima, can be balanced by higher probability of crossing over high energy barriers thus maintaining convergence of the resulting Markov Chain.

Physically Based Model

The physically based model applies forces to the knot which are calculated by integration of ordinary differential equations. Two types of forces are used: elastic and electrostatic. Elastic force is implemented with Hook's law. For a pair of particles at positions a and b , spring force between them is given by:

$$f_a = \left[k_s (|l| - r) + k_d \frac{(v_a - v_b) \cdot l}{|l|} \right] \frac{l}{|l|}, f_b = -f_a \quad (4)$$

where f_a is the force on a , f_b is the force on b , $l = a - b$, r is rest length, k_s is spring constant, k_d is damping constant v_a is the velocity of a , v_b is the velocity on b . It is applied between all pairs of neighboring vertices.

Electrostatic force is based on Coulomb's law where force is inversely proportional to the distance between two edges. Let a and b be the two edges. Electrostatic force between them is approximated by the equation:

$$f_a = \frac{k(q_a q_b)}{d_{min}} u, f_b = -f_a \quad (5)$$

where f_a is the force on a , f_b is the force on b , q_a is the charge of a , q_b is the charge of b , d_{min} is the shortest distance between a and b , u is a unit vector pointing from b to a , k is a constant of proportionality. Electrostatic repulsion is applied between all pairs of edges, with the exception of neighboring edges, due to singularities that are present there. Standard techniques are used to integrate equations of motion.

Charge Drop

In the charge drop method a number of electrostatic point charges are randomly dropped in the vicinity of the knot which is also charged with a uniformly distributed charge of the same sign (Fig. 2). A random section (i.e., a set of consecutive edges) of the knot of random length is selected and it is allowed to move while the rest of the curve remains fixed. In this scenario, the point charges will distort the selected section by repelling the similarly charged line segments. At the same time, the forces of repulsion between all edges of the knot will implicitly prevent it from violating the isotopy condition during this process. Line segment repulsion is weighted heavier than random charge repulsion to insure that the force from charge repulsion will not overpower line segment repulsion, resulting in a crossover. The system is advanced a random number of iterations with the variable step size midpoint Euler method.

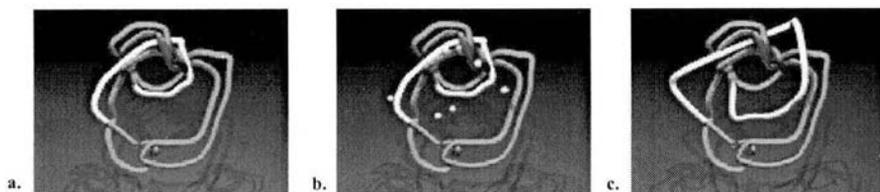


Figure 2: Drop charge perturbation: a. choose random segment, b. drop random charges, c. evolve the system deterministically.

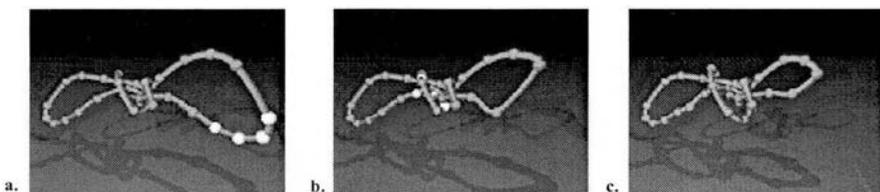


Figure 3: A sequence of operations during the “Shift Node” perturbation method: a. delete a sequence of nodes b. insert a sequence of nodes c. evolve the knot according to their equations of motion.

Node Shift

The second perturbation method is the node shift method which tightens and loosens different parts of the knot. Nodes, or vertices, are shifted along the polygonal knot followed by the application of spring force and edge repulsion

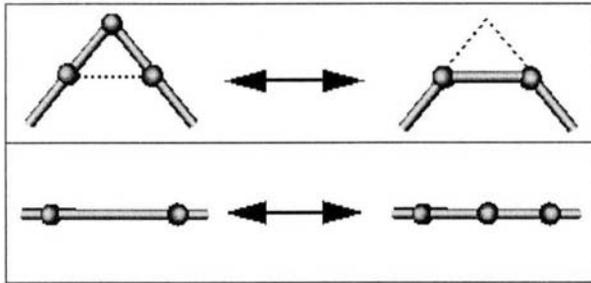


Figure 4: Two operations are used by the “Shift Node” method: remove vertex by a triangle move (top panel), add a vertex (bottom panel).

forces to the knot (Fig. 3). Nodes are shifted by first selecting vertex positions to delete. A position can be deleted if the new knot configuration K' is ambient isotopic to the original knot K . This will be true if the move from K to K' does not allow any part of the knot to pass through itself. This is determined by creating a triangle from the vertices of the selected node and its two neighbors. This triangle represents the smooth path the line segment would travel through in order to move to K' (Fig. 4). If any edge intersects this triangle, then this movement would create a crossover, possibly changing the knot type. Such crossover would cause the move to be rejected. However, if no edge intersects this triangle, the node is a candidate for deletion. A list of nodes to delete are selected by beginning at a start node chosen on the knot then walking along the knot performing triangle tests on consecutive 3 vertices. Similarly, a sequence of nodes is added to the new configuration by creating a new vertex position at the midpoint of edges (see Fig. 4). This simple operation cannot affect the knot type. Once the nodes are moved around, the system is allowed to evolve in the unbalanced potential thus generated under the influence of electrostatic and elastic forces.

A similar method proposed in¹⁵ adds vertices to a polygonal knot in places where potential energy is high. The node shift method presented in this paper maintains a consistent number of vertices in the knot. Vertices are deleted in some areas and added to other areas, which can be viewed as a shift of vertices along the polygonal knot. The number deleted and the number added are equal which maintains a constant problem size making the optimization problem manageable. It also gives consistency to any mapping the user may have defined of attributes to each node.

2.4 Annealing Schedule

The annealing schedule is the component of the system that regulates the temperature and the convergence of the method. Before the proper annealing process can be initiated, it is crucial to assure that the system is thoroughly “melted.” In the “melted” state the temperature of the system must be high enough so that the thermal energy of the particles is large enough to jump over the highest energy barriers, allowing for unconstrained exploration of the entire search space. The system is at an effective melting temperature when at least 80% of moves are accepted¹³. One can automatically “melt” the system by doubling temperature until the prescribed fraction of the moves is accepted. From this point the system is slowly cooled taking care that at each temperature the system should reach equilibrium before the temperature is further decreased. In this work, the temperature is decreased when either the system runs *max* moves or the number of accepted moves exceeds $0.1 \times \text{max}$, where *max* is equal to a constant times the number of vertices in the knot. Because our moves are Cauchy-distributed, we can use a fairly rapid exponential annealing schedule in which the temperature T decreases by a constant ratio²³: $T_n = (T_1/T_0)^n T_0$, where T_0 is starting temperature and $0.90 < (T_1/T_0) < 0.99$ ¹⁴. The system is considered frozen and annealing stops when less than a fixed fraction (e.g., 1%) of attempted moves are accepted at any given temperature. We can terminate the process early in case of the unknot, which has the lowest characteristic energy of all n -vertex knots¹⁶, by comparing to within a tolerance the energy of the knot with energy calculated for a regular n -gon.

3 Results

3.1 Energy Unimodality

In an initial series of experiments, we considered examples of knots that appear to have distinct local energy minima, in the sense of an empirically stable end state to the process initiated by the dynamics of the program. In the first experiment, we compared the minimal energies for the trefoil and its mirror image and found them, not surprisingly, to be identical. Similar results were obtained for the figure-eight knot and its mirror image.

In a more involved experiment¹¹, we considered the Kinoshita-Terasaka Knot *KT*. This knot, shown in Fig. 6, is famous as the one of the smallest knots that is undetectable by the classical Alexander polynomial^{1, 12}. In Fig. 6 we show a 12 crossing version of *KT*, labeled *KT12* and its 11 crossing equivalent *KT11*. From the point of view of a knot theorist, the 12 crossing

configuration is quite interesting, as it shows that the knot KT is a *ribbon knot*, and consequently KT bounds a smooth disk in upper four space. In our experiment we wanted to determine the minimal configuration of the KT knot given either of the configurations shown in Fig. 6 as the starting point. In order to eliminate dependency of the energy on the number of edges²⁴, we introduced additional vertices along the edges of both configurations to assure that both were represented by a polygon with the same number of vertices ($n=56$). These were used as the initial configurations for a deterministic optimization procedure. The final configurations, which were arrived at after less than 4000 iterations, were visually distinct, yet quite similar to the initial states. The corresponding energy values were identical to within 0.05% (336.303 and 336.458 respectively).

3.2 SA vs. Gradient Descent

Having determined empirically that multiple distinct minima exist for the Simon's energy, we proceeded with a set of experiments that would evaluate the performance of the method as compared to deterministic methods. Three tangled configurations shown in the top row of Fig. 7 were the subjects of the tests described below. They were all executed on an SGI Indigo 2 system configured with a 200 MHz MIPS R4400 processor and 96 MB of RAM.

Case 1

In our first experiment we used a tangled unknot represented by a polygonized curve with 67 vertices. First, we applied a deterministic optimizer which found a stable solution after 60,000 steps, going from the initial energy of 1561.85 to the final energy of 116.83, within a threshold value of 0.06 of the global minimum, within 5 hours. Subsequently, the same initial configuration was subjected to the stochastic optimization method and reached minimal energy configuration within 115 moves, which took approximately 2 hours. Temperature started at 150.0 degrees and was decreased exponentially by a constant factor of 0.95. The annealing was considered complete (and the system deemed frozen) when the temperature dropped below 1.0 degrees or when the system's energy E was within $\epsilon = 0.5$ of the global minimum E_{min} of an unknot of 67 vertices. The actual comparison normalized the energy by dividing energy by the number of vertices $vnum$ ($|\frac{E-E_{min}}{vnum}| < \epsilon$). Final energy was 112.84 as compared to energy of a 67 vertex unknot of 109.96.

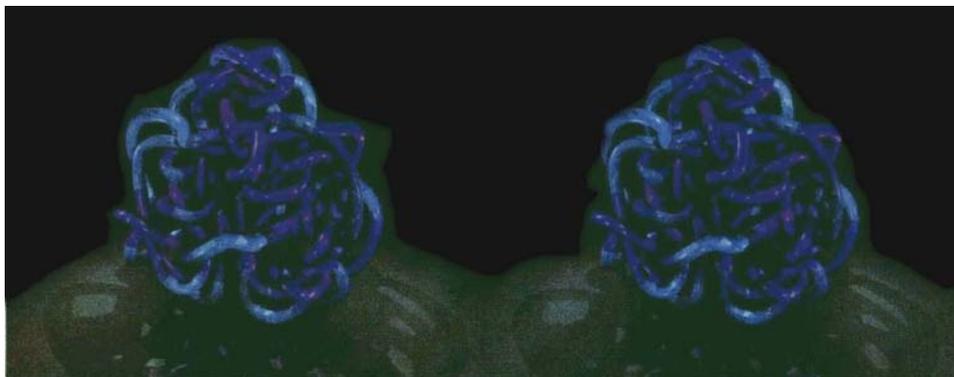


Figure 5: A stereo pair of the Ochiai knot which was used as the starting configuration for case 2 (see text).

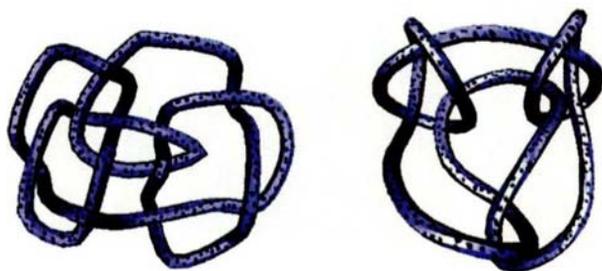


Figure 6: Kinoshita-Terasaka knot: 11 crossing version and 12 crossing version.

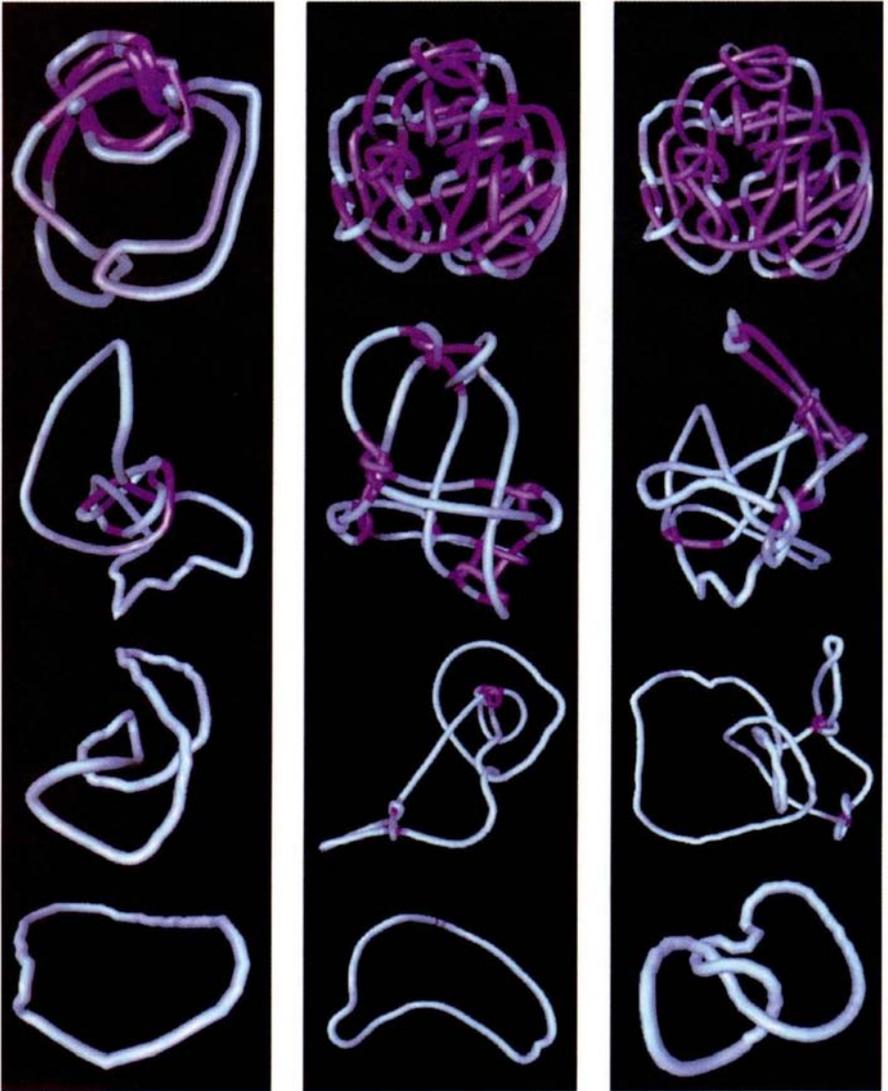


Figure 7: The three test cases (see text) evolve under guidance of a simulated annealing process from the initial configuration (top row), through melting and cooling, to freezing (bottom) stages. The curves are color-coded by the potential energy with shades of red corresponding to high, and blue to low energy.

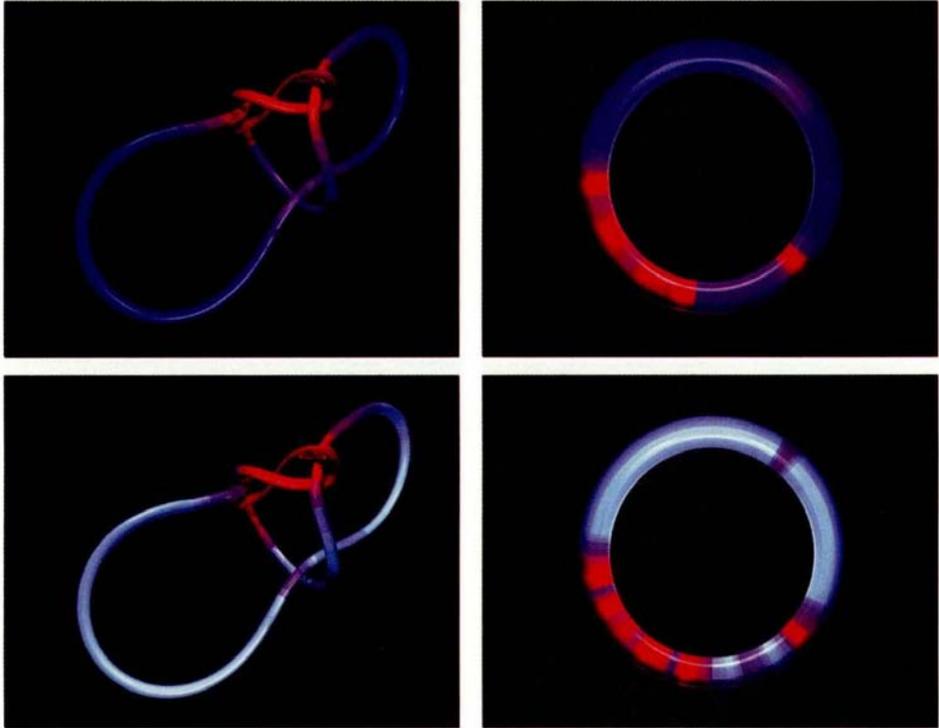


Figure 8: In situations where distribution of a feature along the curve rather than the geometry of the curve itself is of interest, a Gauss model of the curve is used (see text). The top row is colored by the potential energy (red corresponds to tight, and blue to loose areas). The bottom row is colored by the magnitude of the energy gradient (red corresponds to strong, blue to weak force).

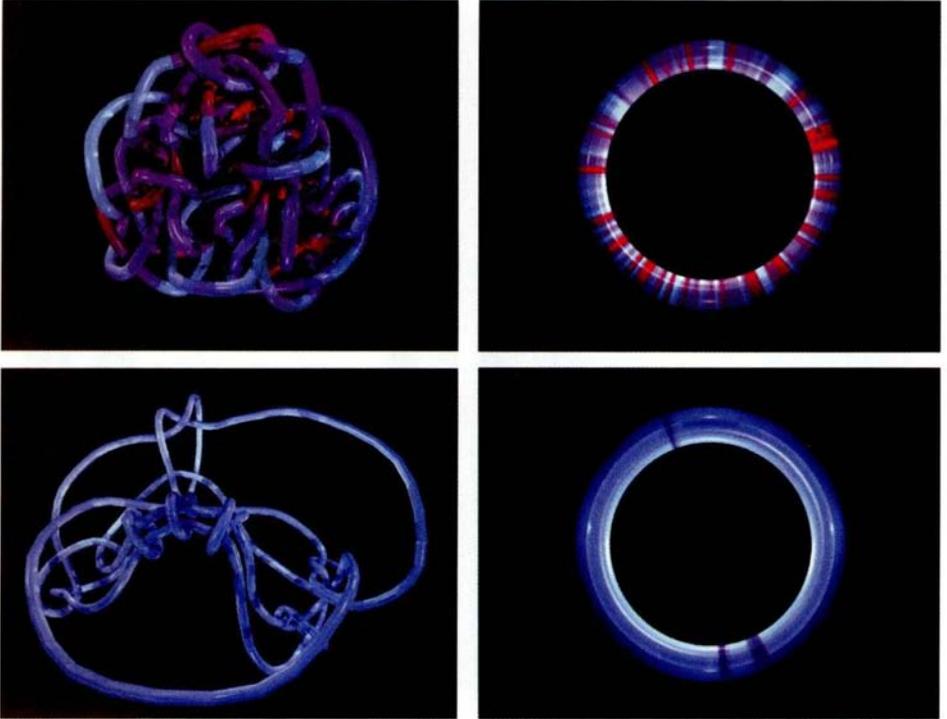


Figure 9: Distribution of the energy gradient magnitude (i.e., strength of the self-repulsive force) along the knot. The initial and intermediate configurations for the iterative improvement approach are shown in the top and bottom panels respectively. The right column shows the configuration on the left after it was unfolded into its Gauss model. Red corresponds to high and blue to low gradient magnitude. Nearly zero force for the second configuration indicates a mechanically stable state that gives little guidance to gradient descent methods.

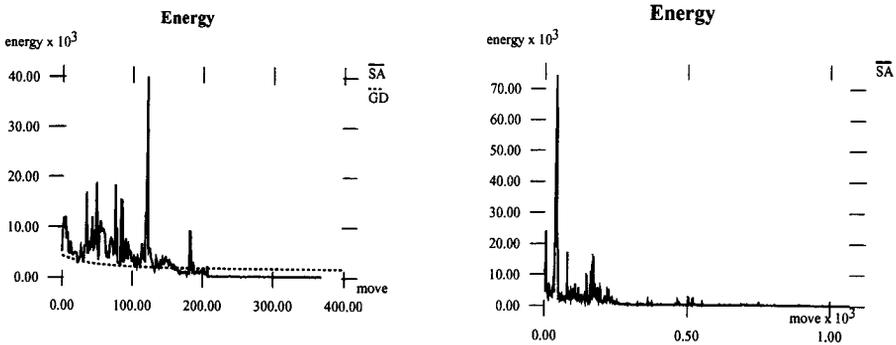


Figure 10: Graphs of energy vs. number of iterations during Ochiai unknot (left) and Ochiai trefoil (right) evolution (cases 2 and 3 in text) for simulated annealing (solid line) and gradient descent (dashed line).

Case 2

In our second experiment we used an unknot represented by a polyline with 139 vertices whose tangled configuration was described by Ochiai¹⁷. A stereo pair depicting the starting configuration is shown in Fig. 5. The energy of the initial configuration was 4464.47. As before, gradient descent technique was used first to untangle this configuration. Care had to be taken in selecting a suitably small step size, otherwise, the system would attempt energy-increasing moves. Sufficiently small step size would eventually produce a circular configuration. However, the convergence rate of the system was fairly slow: it took 292,000 steps and 107 hours of computation to simplify the knot. Attempts to speed up the process via adaptive step adjustment based on the maximum gradient strength did not produce satisfactory results. Fig. 9 shows the distribution of the energy gradient strength along the curve for the initial and an intermediate configuration. Simulated annealing reached the minimal energy configuration within tolerance of $\epsilon = 0.5$ after 368 moves. Fig. 7 illustrates the knot's starting configuration and a number of intermediary configurations during simulated annealing. Fig. 10 shows the associated energy graph. Temperature started at 377 degrees, and was allowed to decrease by a factor of 0.95, until temperature reached 1.0 degrees or the energy was within ϵ of the energy of a 139 vertex unknot. The end energy of 235.30 was reached after approximately 48 hours of wall clock time. Energy of a 139 vertex unknot is 228.52.

Case 3

Case 3 was a 139-vertex trefoil knot obtained from Ochiai's unknot by replacing one of the over-crossings in the upper right part of the planar projection with an under-crossing (compare top row of the center and right columns in Fig. 7). The initial energy was 4493.9. Gradient descent method exhibited prohibitively slow convergence and was terminated after 150,000 moves when it failed to lower the energy of the system by more than 0.1% in 1,000 consecutive iterations. Simulated annealing was terminated after 840 moves during which the temperature was lowered exponentially at the rate of 0.95 from the melting point of 340 degrees to near freezing at 3.3 degrees. Fig. 10 shows the associated energy graph. The final configuration shown in Fig. 7 has an energy of 354.9 (as compared to the energy of a 139-node trefoil of 265.3).

4 Conclusions

An important contribution of this work is an empirical identification of cases for which Simon's energy is not unimodal. Namely, it fails to detect chirality as well as to distinguish between distinct configurations of the KT knot. Therefore, we were able to determine that the minimal configuration is not unique to the knot type, even though the minimal energy may be. This determination provides the rationale for using a robust optimization technique for the purpose of analyzing knot energies. It also shows Simon's approach to knot classification to be no more powerful than Alexander's polynomial, and less powerful than Jones', *HOMFLY*, and Kauffman's polynomials. It is important to note that multimodality of Simon's energy function may not necessarily invalidate gradient-based methods as long as all the distinct local minima correspond to the same energy level. Obviously, the above result has to be viewed in the context in which it was obtained. It is still possible that a unimodal energy function exists and the tool proposed here can be used to search for it.

In order to validate the proposed method, we applied it to three different cases. The method was able to determine the minimum energy level for each case and to classify the initial tangles as either the unknot or the trefoil. From the energy graphs (Fig. 10) and the evolution sequences (Fig. 7) one can see that the knots were allowed to freely search the solution space and take uphill energy moves. The second and third cases illustrated configurations that pure gradient descent had difficulty minimizing in an efficient manner, whereas simulated annealing successfully found the energy minimum. The main conclusion drawn from these results is that Simon's energy is plagued by numerous, nearly flat plateaus and is thus difficult to optimize with gradient descent methods,

which use slope to guide them in their search. This conclusion is verified by the bottom row of Fig. 9 which depicts the distribution of the gradient strength along the curve in an intermediate configuration arrived at with the iterative improvement method for case 2. Nearly zero-strength resultant forces acting on all the edges of the knot are indicative of the mechanical stability of this configuration, which causes slow convergence of the gradient-based methods.

The approach proposed here shows promising results when compared to competing methods. Many current methods use gradient descent techniques⁴ which will find a minimum, but not necessarily the global minimum. Even if all proposed self-repelling potentials are proven to have a unique minimal energy, our method may produce more rapid convergence than conventional techniques. Further, it allows for a much wider class of functions to be used as estimates of knot energies. For example, the energy for very large models can be approximated with stochastic sampling. The resulting energy is noisy and thus unsuitable for gradient based methods.

An important characteristic of the proposed approach is our use of a physically-based model to implicitly enforce the isotopy requirement, which in no way determines the function being optimized. Therefore, while the results presented here refer specifically to one individual energy (i.e., Simon's minimal distance) the same technique can be useful in a wide variety of other approaches.

Acknowledgments

Ying-Quing Wu is acknowledged for kindly providing us with a copy of his "ming" program which was used to evaluate the iterative improvement approach. POV-Ray was used for ray-tracing. Dan Sandin is acknowledged for helpful discussion. Michael Jones and Guy Russel are gratefully acknowledged for their editorial comments. The second author wishes to acknowledge funding for the Electronic Visualization Laboratory at the University of Illinois at Chicago from the National Science Foundation (NSF), grant CDA-9303433, which includes major support from the Defense Advanced Research Projects Agency.

1. J.W. Alexander. Topological invariants of knots and links. *Trans. Amer. Math. Soc.*, 20:275–306, 1923.
2. Philip Bevington. *Data reduction and error analysis for the physical sciences*. McGraw-Hill, 1992.
3. Steve Bryson, Michael Freedman, Zheng-Xu He, and Zhenghan Wang. Mobius invariance of knot energy. *Bull. Amer. Math. Soc.*, 28:99–103, 1993.

4. G. Buck and J. Simon. Knots as dynamical systems. *Topology Appl.*, 51, 1993.
5. G.C. Fox, R.D. Williams, and P.C. Messina. *Parallel Computing Works*. Morgan Kaufmann Publishers, Inc., 1994.
6. M.H. Freedman, Z.X. He, and Z. Wang. On the energy of knots and unknots. *Annals of Math.*, 139:1–50, 1994.
7. S. Fukuhara. Energy of a knot. In Y. Matsumoto, T. Mizutani, and S. Morita, editors, *A Fete of Topology*, pages 443–451. Academic Press, Inc., 1988.
8. C.F. Gauss. *Werke*. Teubner, Leipzig, 1900.
9. M. Huang, R. Grzeszczuk, and L. Kauffman. Untangling knots by stochastic energy optimization. In *Proceedings of IEEE Conference on Visualization*, pages 279–286, 1996.
10. V. Interrante, H. Fuchs, and S. Pizer. Illustrating transparent surfaces with curvature-directed strokes. In *Proceedings of Visualization '96 Conference*, pages 211–218, 1996.
11. L. Kauffman, M. Huang, and R.P. Grzeszczuk. Self-repelling knots and local energy minima. In *Proceedings of the conference on Random Knotting*, 1996.
12. L.H. Kauffman. *Knots and Physics*. World Scientific Publishing, 1993.
13. S. Kirkpatrick. Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34:975–986, 1984.
14. S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
15. R.B. Kusner and J.M. Sullivan. Mobius energies for knots and links, surfaces and submanifolds. In Willam H. Kazez, editor, *Geometric Topology*, pages 570–604. American Mathematical Society, International Press, 1997.
16. Terry Ligocki and James Sethian. Recognizing knots using simulated annealing. *Journal of Knot Theory and Its Ramifications*, pages 477–495, 1994.
17. M. Ochiai. Non-trivial projections of the trivial knot. *S.M.F. Asterisque*, 192:7–9, 1990.
18. J. O'Hara. Energy of a knot. *Topology*, 30:241–247, 1991.
19. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.
20. R.C. Read and P. Rosenstiehl. On the gauss crossing problem. *Colloquia Math. Soc. J. Bolyai, Combinatorics*, 18:843–876, 1976.
21. K. Reidemeister. *Knotentheorie*. Chelsea Publishing Co., 1948.

22. P. Rheingans. Opacity-modulating triangular textures for irregular surfaces. In *Proceedings of Visualization '96 Conference*, pages 219–225, 1996.
23. H. Shu and R. Hartley. Fast simulated annealing. *Phys. Lett. A.*, 122:157–162, 1987.
24. J. Simon. Energy functions for polygonal knots. In *Random Knotting and Linking*. World Scientific Publishing, 1994.
25. Ying-Qing Wu. Personal correspondence, Jan 1996.

This page is intentionally left blank

COLOUR PLATES

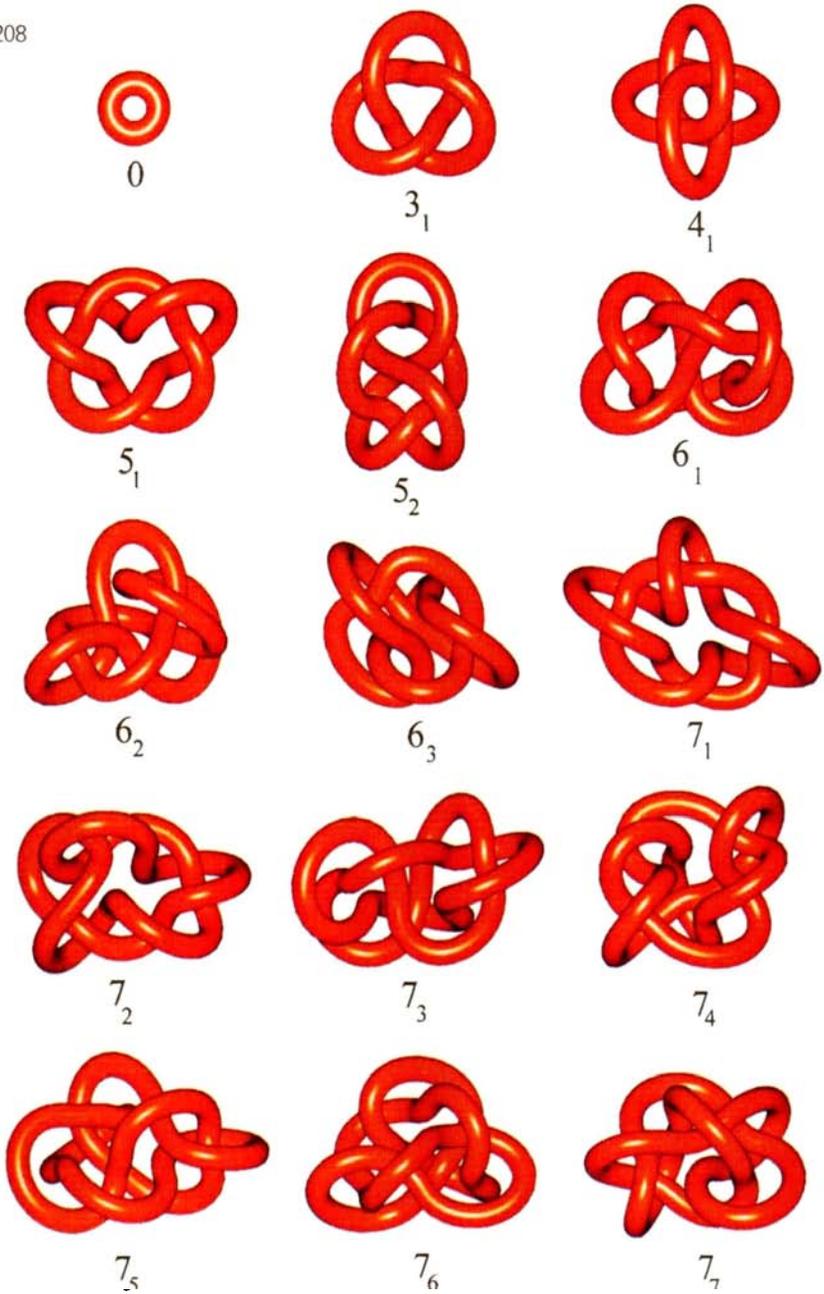
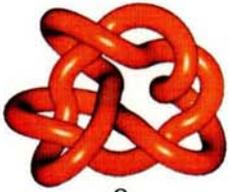
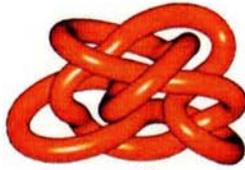
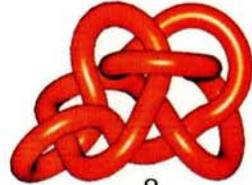
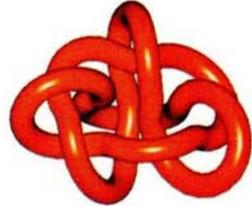
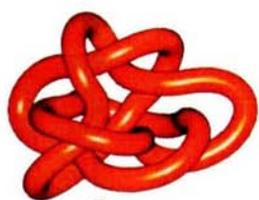
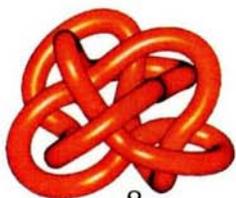
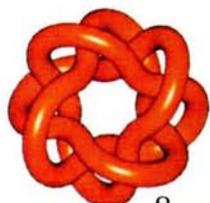
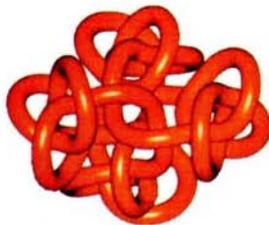
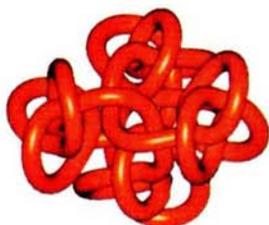
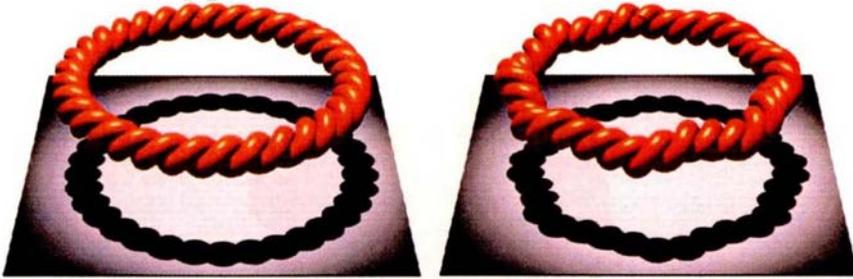


Plate 1-3. Ideal geometrical representations of various types of knots as obtained by computer simulations. We show all prime knots with up to 8 crossings and three examples of composite knots. The last composite knot which is composed of five 4_1 knots is presented as a stereo pair.

8₁8₂8₃8₄8₅8₆8₇8₈8₉8₁₀8₁₁8₁₂8₁₃8₁₄8₁₅

 8_{16}  8_{17}  8_{18}  8_{19}  8_{20}  8_{21}  $3_1 \# 3_1$  $+3_1 \# -3_1$  $4_1 \# 4_1 \# 4_1 \# 4_1$



(a)



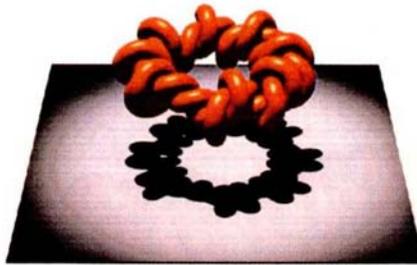
(b)



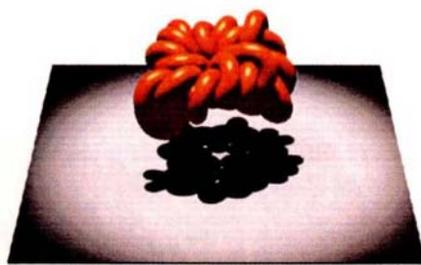
(c)



(d)



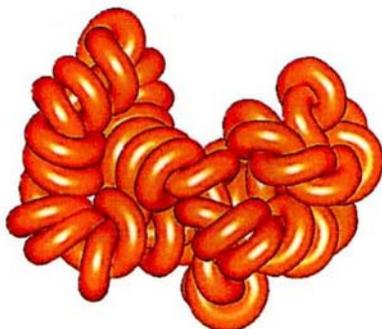
(e)



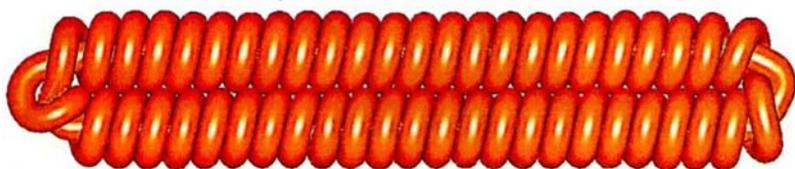
(f)

Plate 4. Symmetry breaking evolution of the $(2, 33)$ torus knot during the tightening process simulated with the SONO algorithm. (a) the circular tightly twisted form. (f) the collapsed discoidal form.

(a)



(b)



(c)

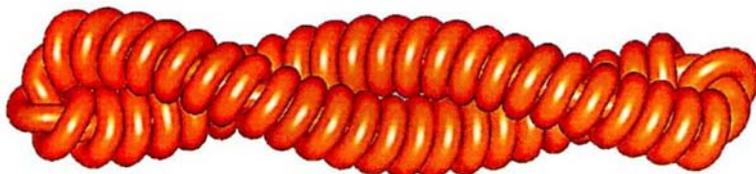


Plate 5. Three different forms of the $(2, 99)$ torus knot: (a) the discoidal form elapsed from the symmetry breaking evolution of the circular initial conformation, (b) the elongated tight form, (c) the twisted elongated tight form.

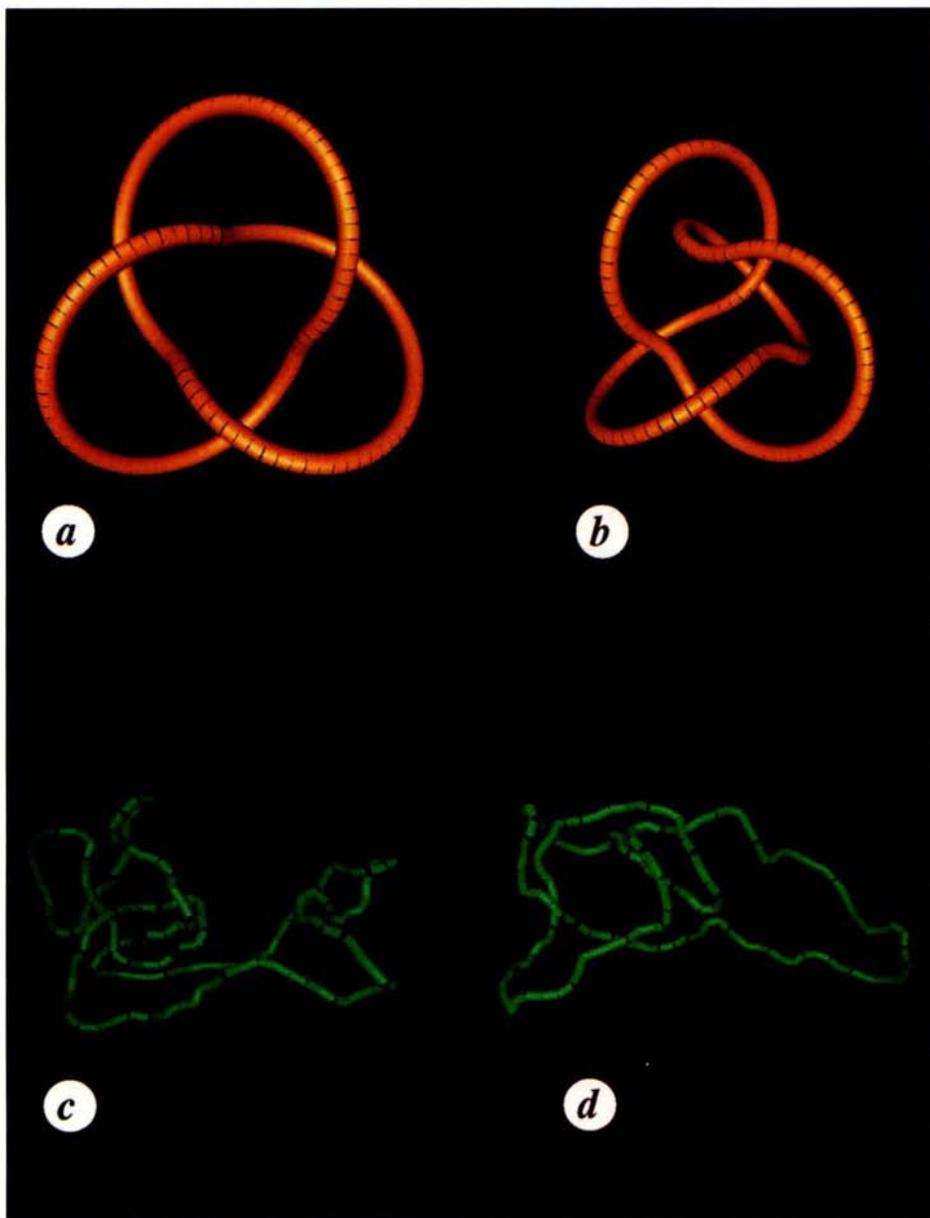
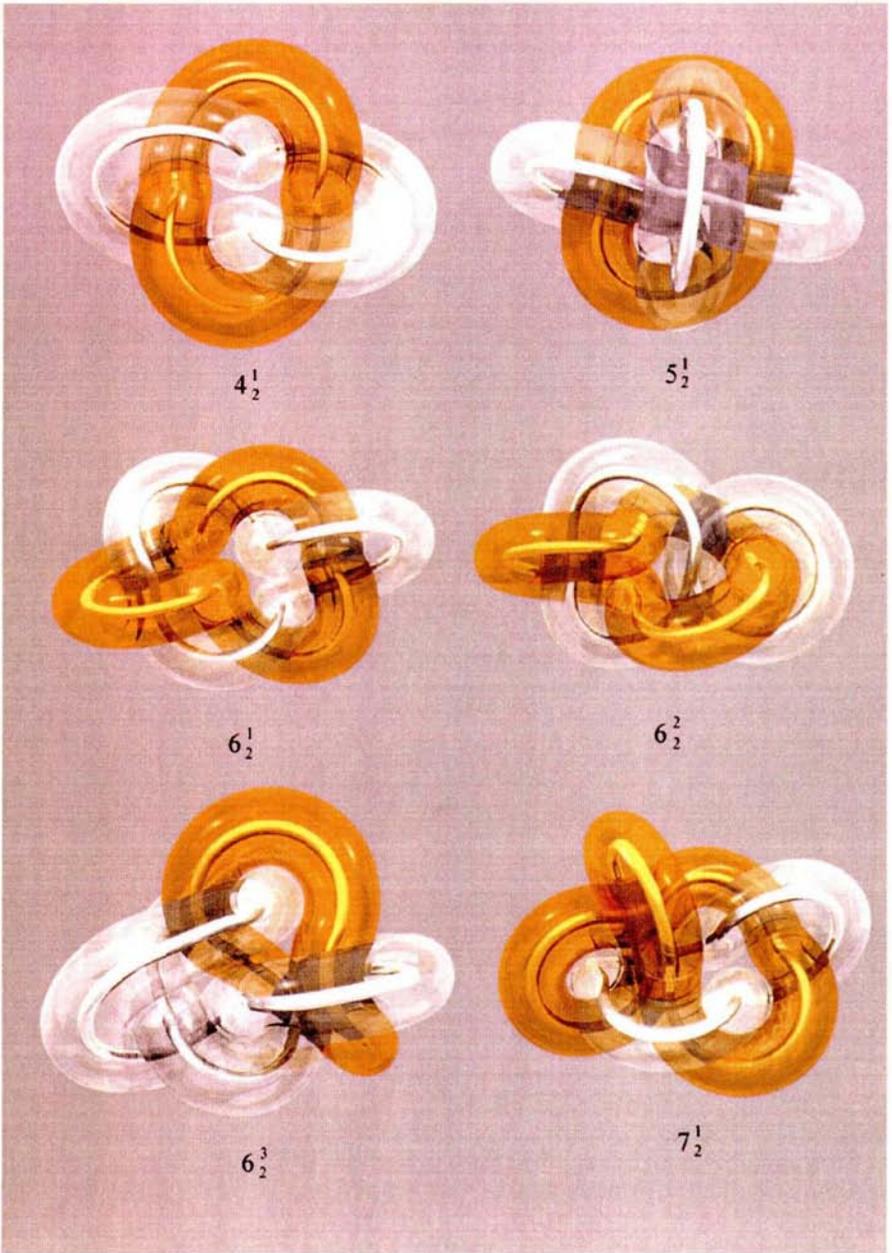
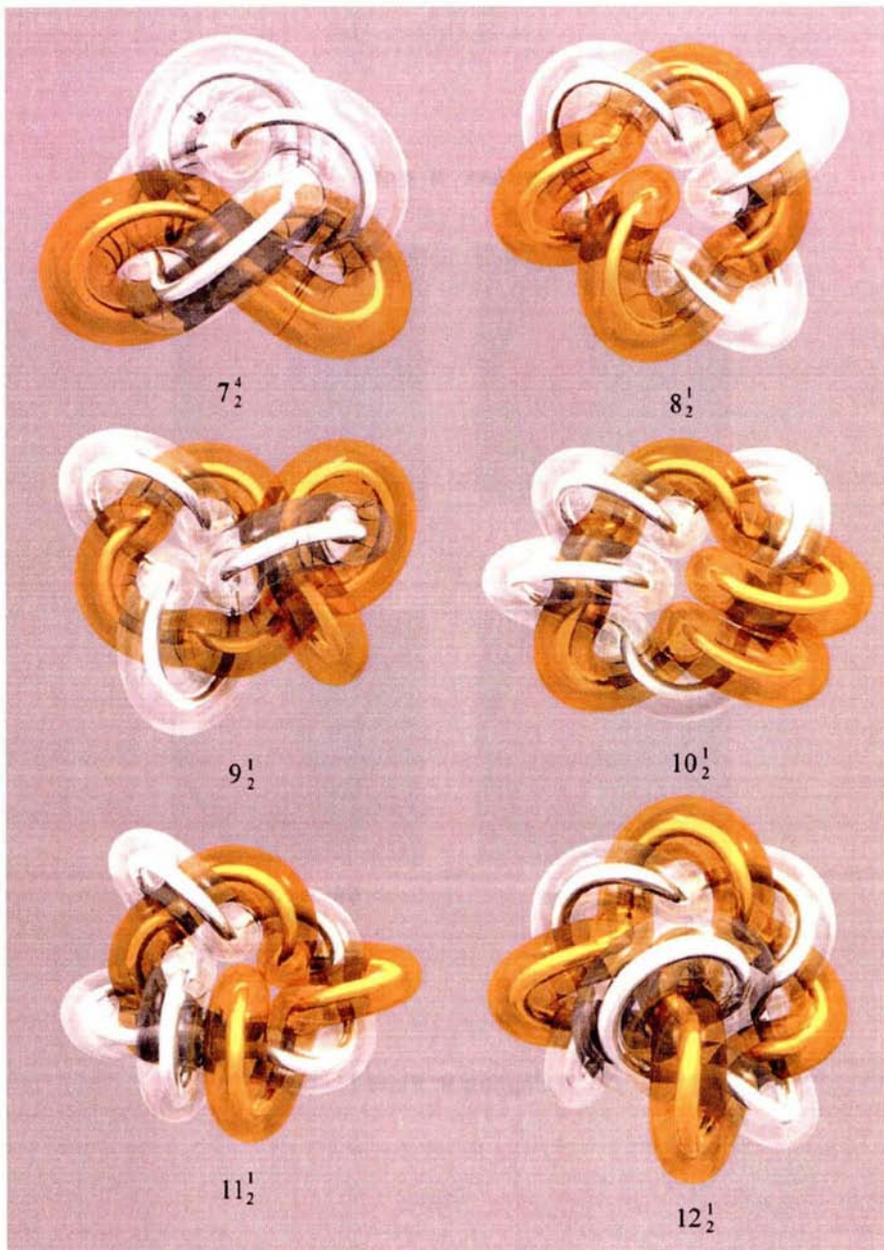
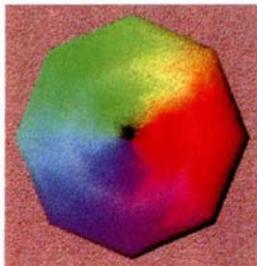


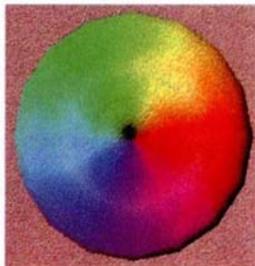
Plate 6. Comparison between ideal trajectories of knots and of randomly distorted configurations: (a), (b) ideal axial trajectories of trefoil and four-noded knot; ©, (d) one of random configurations, simulated for DNA molecule 5400 bp long undergoing thermal motion.



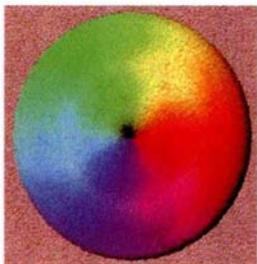


Regular n gons

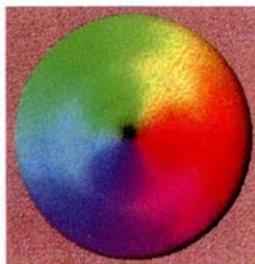
8 sides
.9481



16 sides
.9871



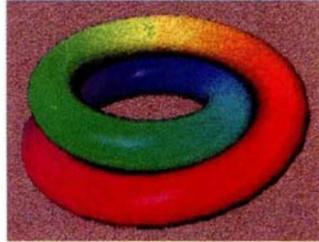
32 sides
.9968



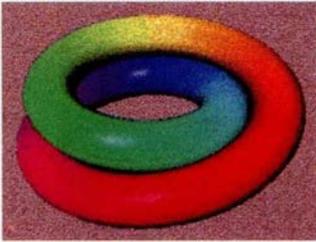
64 sides
.9992

(2,1) Torus Knot

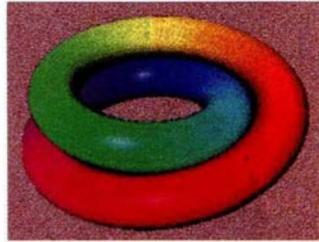
25 sides
.9184



50 sides
.9902



100 sides
.9975



200 sides
.9994

Figure 8 Knot

Figure 8 Knot



25 sides
96.36



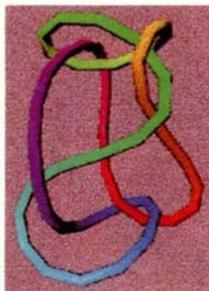
50 sides
87.43



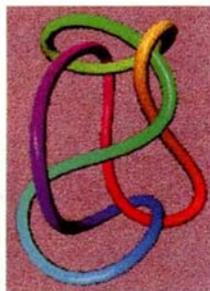
100 sides
86.45



200 sides
86.36

8_12 Knot

50 sides
312.61



100 sides
276.17



200 sides
270.89



400 sides
269.04

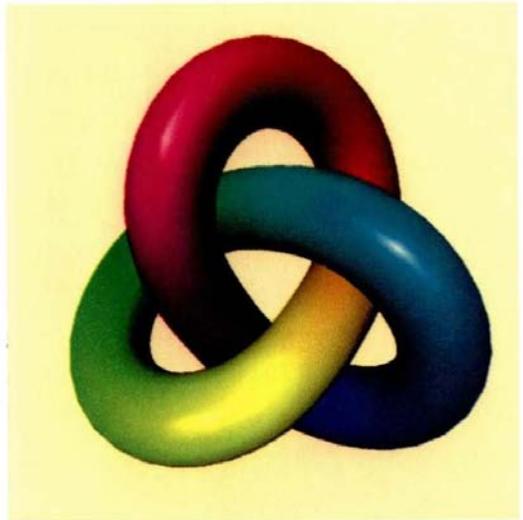
Which knot is more "ideal"?

This conformation of the trefoil knot has maximum thickness - for a given length of string, it admits a uniform tube of maximum radius. (The tube shown here is thinner than maximum, to show the shape of the knot.)



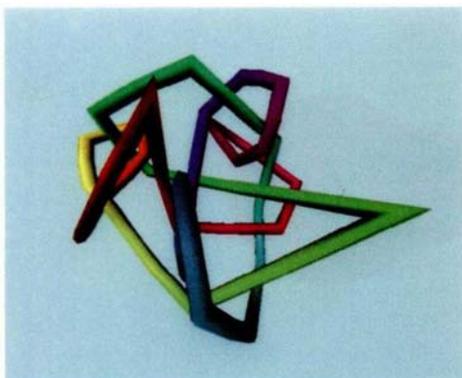
This conformation was obtained from the above one by lowering the *minimum distance energy*. The curvature is now more uniform, but the maximum tube-thickness is smaller.

(E. Rawdon, J. Simon)

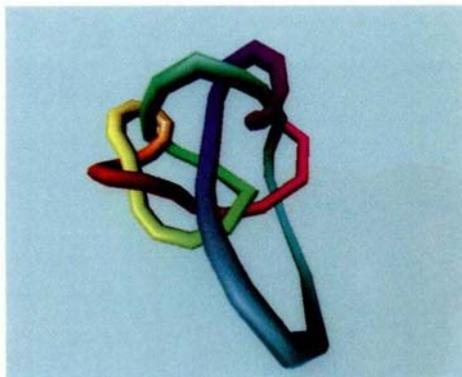


Evolution of a knot using Y.Q. Wu's program *MING*

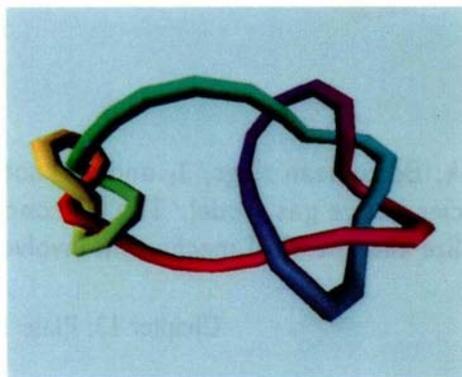
We begin with a complicated polygon, and tell the computer to move the vertices so as to reduce the *minimum distance energy*.



The actual knot type begins to emerge.



And we see that the knot actually is a composite of two familiar knots.



(J. Simon)



Hopf link, Borromean rings, 3_1 and 4_1 knots observed in the study of a two species lattice gas model. The iso-concentration surfaces are drawn to visualize the chemical mechanism involved in the pattern formation.

CHAPTER 12

KNOTS AND FLUID DYNAMICS

H.K. Moffatt

*Isaac Newton Institute for Mathematical Sciences
Clarkson Road, Cambridge, UK.*

1 Introduction

Fluid mechanics provides a very natural setting for the consideration of knotted or linked structures, and the manner in which these may ‘relax’ to an equilibrium configuration of minimal energy. For we may imbed within a fluid medium a solenoidal vector field $\mathbf{B}(\mathbf{x}, t)$ ($\nabla \cdot \mathbf{B} = 0$) which represents the knotted or linked structure in an appropriate way and which is ‘frozen’ in the fluid under any continuous fluid motion, so that its topological properties are conserved. If moreover the field \mathbf{B} is assumed to impart a force to the fluid medium, then an associated energy W stored in the field \mathbf{B} may be defined. Under the action of the force, this energy is converted to kinetic energy of motion and hence dissipated via the agency of viscosity. There is thus a natural mechanism for the decrease of W under the constraint of conservation of field topology; and we are then faced with the mathematical problem of determining the equilibrium field configuration, $\mathbf{B}^E(\mathbf{x})$ say, which minimises W subject to prescribed field topology. The fluid dynamical relaxation process provides a natural route towards this minimum energy equilibrium state.

In the following, we shall first review the general theory of relaxing fields, as developed by Moffatt (1985), and then particularise to fields defined within a tubular neighbourhood of a given knot K (Moffatt 1990). The internal twist of the field within this neighbourhood (related to the ‘framing of the knot’) affects the relaxation process, and the minimum energy attained depends upon this twist. If the twist is zero, then the minimum energy configuration is closely related to the ‘ideal’ configuration of Katritch *et al* (1996) in which the tube length L is minimised (and cross-section S maximised) for given volume $V = LS$; however the relaxation process does not constrain the cross-section to remain circular.

If the twist is large, then the behaviour is quite different: the dominant field component B_θ (where r, θ are polar coordinates on the tube cross-section) now tends to *decrease* S with consequential *increase* of L ; at the same time, the tube is subject to kink instabilities which decrease the internal twist, but at

the expense of increase in the *writhe* of the tube axis. What is conserved in this process is the *helicity* $\mathcal{H} = h\Phi^2$ of the tube field, where Φ is the (conserved) axial flux; and we shall show that the minimum energy scales like $|h|^{4/3}$ when $|h| \gg 1$, irrespective of knot type.

We shall adopt in what follows the language of magnetohydrodynamics (MHD), in which $\mathbf{B}(\mathbf{x}, t)$ is interpreted as the magnetic field in a perfectly conducting fluid medium. We need not, however, feel constrained by this interpretation; from an abstract point of view, $\mathbf{B}(\mathbf{x}, t)$ is to be regarded simply as a vector field, convected with the fluid motion in such a way that its flux across any *material* element of area (i.e. consisting always of the same fluid particles) is conserved.

2 The general theory of relaxation

The fundamental equation satisfied by any such 'frozen' field is well-known in the MHD context, namely

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \wedge (\mathbf{v} \wedge \mathbf{B}) \quad (1)$$

where $\mathbf{v}(\mathbf{x}, t)$ is the fluid velocity. We shall suppose that the flow is volume-preserving (incompressible) so that $\nabla \cdot \mathbf{v} = 0$. The flow determines a time-dependent mapping

$$\mathbf{x} \rightarrow \mathbf{X}(\mathbf{x}, t) \quad (t > 0) \quad (2)$$

for each fluid particle starting at \mathbf{x} at time $t = 0$; and the solution of (1) is given in terms of this mapping by

$$B_i(\mathbf{X}, t) = B_j(\mathbf{x}, 0) \frac{\partial X_i}{\partial x_j}. \quad (3)$$

The antisymmetric part of the deformation tensor $\partial X_i / \partial x_j$ is associated with rotation of a fluid volume element initially centred on \mathbf{x} , while the symmetric part represents irrotational deformation (whereby a small sphere deforms to an ellipsoid).

Let us define the energy density of the field \mathbf{B} as $\frac{1}{2}\mathbf{B}^2$; the rate of change of this energy density is, from (1),

$$\frac{\partial}{\partial t} \frac{1}{2} \mathbf{B}^2 = \mathbf{B} \cdot \frac{\partial \mathbf{B}}{\partial t} = \mathbf{B} \cdot \nabla \wedge (\mathbf{v} \wedge \mathbf{B}). \quad (4)$$

By standard vector manipulation, this may be written

$$\frac{\partial}{\partial t} \frac{1}{2} \mathbf{B}^2 = -\mathbf{v} \cdot (\mathbf{j} \wedge \mathbf{B}) - \nabla \cdot [(\mathbf{v} \wedge \mathbf{B}) \wedge \mathbf{B}] \quad (5)$$

where $\mathbf{j} = \nabla \wedge \mathbf{B}$ is the *current* associated with \mathbf{B} . If, as we may suppose, \mathbf{B} is a localised field, the divergence term vanishes when integrated over the fluid domain, and we have

$$\frac{dM}{dt} = - \int \mathbf{v} \cdot \mathbf{F} dV \quad (6)$$

where $M = \frac{1}{2} \int \mathbf{B}^2 dV$, and

$$\mathbf{F} = \mathbf{j} \wedge \mathbf{B} \quad (7)$$

is the effective force per unit volume exerted by the field \mathbf{B} on the fluid. In electrodynamics, this is recognized as the Lorentz force, and is associated with the Maxwell stress tensor

$$T_{ij} = B_i B_j - \frac{1}{2} B^2 \delta_{ij}$$

representing both tension in the lines of force of the \mathbf{B} -field, and pressure between adjacent lines of force. The field energy decreases through contraction of lines of force (thus relieving tension), but this can proceed only for so long as topological constraints present in the initial field $\mathbf{B}_0(\mathbf{x}) = \mathbf{B}(\mathbf{x}, 0)$ permit. If every \mathbf{B} -line is an unknotted closed loop which can be shrunk to a point in the fluid domain without cutting any other \mathbf{B} -lines, then there is no topological impediment to decrease of the field energy towards zero. This case (of *trivial topology*) is however exceptional; in general even simple fields for which all field lines are closed curves exhibit linkages that present topological barriers to this decrease of energy. Three examples are shown in figure 1: a flux linkage, a Whitehead link (in which the total flux trapped is zero), and a Borromean link (in which the total flux trapped is again zero and it is the relative position of different strands of flux that provides the topological barrier).

In each of these cases, if a closed field line C is shrunk to a point via a volume-preserving fluid motion, then the energy of the trapped field necessarily increases without limit. Thus for example, if C is taken to be the circle $x^2 + y^2 = a^2$, it may be shrunk to a point via the incompressible strain flow

$$\mathbf{v} = (-\alpha x, -\alpha y, 2\alpha z) \quad (\alpha > 0) \quad (8)$$

under whose action the radius of the convected circle C is $r = ae^{-\alpha t}$ at time t . This flow involves persistent stretching in the z -direction under which the z -component of any trapped flux *increases* exponentially, with consequent increase of field energy. In a freely evolving situation, the contraction of C will presumably be arrested when an equipartition of the field energy is established between the trapping component (in a flux tube centred on C) and the trapped component normal to C .

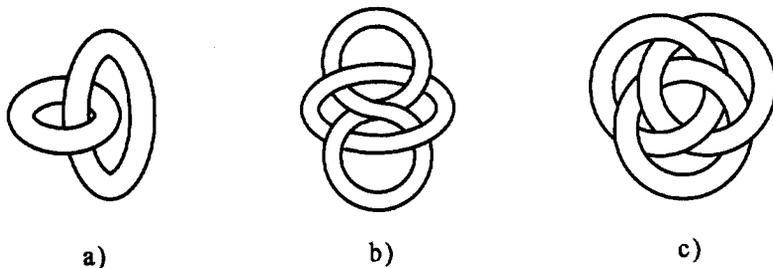


Figure 1: a) Flux linkage with helicity $\mathcal{H} = 2\Phi_1\Phi_2$ (see §3); b) the Whitehead link, with zero helicity; c) the Borromean link for which each pair of flux tubes is unlinked.

This picture has been formalised by Freedman (1988) who has proved that in any topologically nontrivial field of closed \mathbf{B} -lines, the field energy is indeed bounded away from zero under any volume-preserving diffeomorphism. This means that, if we can construct a process by which the energy of such a field decreases monotonically, then that energy must tend to a positive limit.

Such a process may indeed be easily constructed. We simply suppose that the velocity field \mathbf{v} is driven in the fluid by the force $\mathbf{F} = \mathbf{j} \wedge \mathbf{B}$ starting from rest. If the fluid has (uniform) density ρ and viscosity μ , it then flows according to the Navier-Stokes equation

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla p + \mathbf{j} \wedge \mathbf{B} + \mu \nabla^2 \mathbf{v}. \quad (9)$$

If the fluid is assumed to fill \mathbb{R}^3 , the localised force $\mathbf{j} \wedge \mathbf{B}$ generates a localised *vorticity* field $\boldsymbol{\omega} = \nabla \wedge \mathbf{v}$; the corresponding velocity field has a quadrupole character (the total impulse imparted by $\mathbf{j} \wedge \mathbf{B}$ being zero) and is therefore $O(r^{-4})$ as $r = |\mathbf{x}| \rightarrow \infty$. The kinetic energy of motion

$$K = \frac{1}{2} \int \rho \mathbf{v}^2 dV \quad (10)$$

is therefore certainly finite. Its rate of change is

$$\frac{dK}{dt} = \int \mathbf{v} \cdot (\mathbf{j} \wedge \mathbf{B}) dV - \mu \int (\nabla \wedge \mathbf{v})^2 dV \quad (11)$$

the second term on the right representing viscous dissipation of energy (into heat). Coupling this with (6) gives

$$\frac{d}{dt}(K + M) = -\mu \int \omega^2 dV \quad (12)$$

so that the total energy certainly decreases monotonically. Since M is bounded below and K is positive, this means that $K + M$ must tend to a limit, and the enstrophy of the flow must tend to zero:

$$\int \omega^2 dV \rightarrow 0. \quad (13)$$

In this limit, if we ignore the (unphysical) possibility of the appearance and persistence of singularities in vorticity in a viscous fluid, the flow is irrotational; again excluding singularities, the only possibility compatible with the boundary condition $\mathbf{v} \rightarrow 0$ as $r \rightarrow \infty$ is that $\mathbf{v} = 0$ everywhere. Thus, as $t \rightarrow \infty$, the field structure does relax to an equilibrium minimum energy structure $\mathbf{B}^E(\mathbf{x})$, in which the force field $\mathbf{j}^E \wedge \mathbf{B}^E$ is in equilibrium with the pressure gradient ∇p^E , i.e.

$$\mathbf{j}^E \wedge \mathbf{B}^E = \nabla p^E, \quad (14)$$

and the fluid is at rest.

The dynamical model adopted above (eqn. 9) is not the only possibility, if the only requirements of the relaxation process are that (i) the topology of the \mathbf{B} -field be conserved; and (ii) energy be dissipated. The requirement (i) is guaranteed by equation (1). The requirement (ii) may be satisfied by the simpler 'porous medium' model in which \mathbf{v} is directly related to $\mathbf{j} \wedge \mathbf{B}$ via the equation

$$k\mathbf{v} = -\nabla p + \mathbf{j} \wedge \mathbf{B}. \quad (15)$$

The pressure term is still needed in order to ensure that \mathbf{v} remains solenoidal (i.e. $\nabla \cdot \mathbf{v} = 0$). It is easy to show that the energy equation associated with (15) is now

$$\frac{dM}{dt} = -k \int \mathbf{v}^2 dV. \quad (16)$$

Kinetic energy does not appear since, in effect, in the dynamical model (15) inertia forces are neglected. Equation (16) implies monotonic decrease of field energy M until $\mathbf{v} \equiv 0$ (again neglecting the possibility of the appearance of unphysical singularities of \mathbf{v}). The model (15) has been adopted by Linardatos (1993) and by Chui & Moffatt (1996) in the determination of a variety of two-dimensional nonlinear magnetostatic equilibrium states.

3 Conservation of field helicity

There is a global invariant associated with the frozen field equation (1), which plays an important part in the subsequent application to knotted structures. This is the field *helicity* \mathcal{H} . To be specific, let us assume that the field \mathbf{B} has compact support in \mathbb{R}^3 , and let \mathbf{A} be a vector potential satisfying

$$\mathbf{B} = \nabla \wedge \mathbf{A}. \quad (17)$$

Then the helicity of \mathbf{B} is defined by

$$\mathcal{H} = \int \mathbf{A} \cdot \mathbf{B} \, dV, \quad (18)$$

the integral being over the support of \mathbf{B} (or equivalently, over all space). As shown by Woltjer (1958), this helicity is invariant, a result interpreted by Moffatt (1969) in terms of the conserved linkages of magnetic field lines in a frozen-field situation. Note that \mathcal{H} does not depend upon the choice of gauge for the field \mathbf{A} .

The interpretation of \mathcal{H} for the case that is of particular interest here, when \mathbf{B} is confined to a tubular neighbourhood \mathcal{T} of a given knot K , has been established by Moffatt & Ricca (1992). Suppose that the knot K is itself a \mathbf{B} -line of the field, and that every \mathbf{B} -line in \mathcal{T} is a closed curve which cuts any section of \mathcal{T} normal to K only once and which has winding number h around K . Let Φ be the flux of \mathbf{B} across each section of \mathcal{T} normal to K . Then

$$\mathcal{H} = h\Phi^2. \quad (19)$$

Moreover, as shown by Călugăreanu (1959, 1961), h may be expressed as the sum of 'writhe' and 'twist' components:

$$h = Wr + Tw. \quad (20)$$

Here, Wr is defined as a double integral round K :

$$Wr = \frac{1}{4\pi} \oint_K \oint_K \frac{(d\mathbf{x} \wedge d\mathbf{x}') \cdot (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}, \quad (21)$$

and is a number, determined solely by the geometry of K , and which varies continuously under continuous distortion of K .

If K is deformed to be nearly in a plane, then Wr tends to an integer equal to the number of positive crossings minus the number of negative crossings in

the projected knot. More generally, Wr is the average over all projections of this difference between positive and negative crossings.

The twist Tw depends not only on the geometry of K , but also on the *relative geometry* of any two lines of force within the tube \mathcal{T} . Let s be arc length on K and let $\mathbf{N}(s)$ be a unit vector normal to K in the direction from K to a given \mathbf{B} -line in \mathcal{T} . Then, by definition,

$$Tw = \frac{1}{2\pi} \oint_K (\mathbf{N}' \wedge \mathbf{N}) \cdot \mathbf{t} ds \quad (22)$$

when \mathbf{t} is a unit tangent vector on K . This can be thought of as the twist of a ribbon whose boundaries are K and the given \mathbf{B} -line, this twist being assumed the same for all \mathbf{B} -lines. The twist Tw can be further decomposed in the form

$$Tw = \mathcal{T} + n$$

where

$$\mathcal{T} = \frac{1}{2\pi} \oint_K \tau(s) ds \quad (23)$$

with $\tau(s)$ the *torsion* of K , and n is the number of rotations of the vector $\mathbf{N}(s)$ relative to the Frenet triad $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ on K in one passage around K . \mathcal{T} and n are well-defined only if K has no inflexion points (i.e. points of zero curvature). If K is subjected to continuous distortion that takes it through an 'inflexional configuration' (i.e. a configuration having an inflexion point) then both \mathcal{T} and n jump by an integer, the sum however varying continuously. This behaviour was recognized by Călugăreanu (1961); it was shown to be generic by Moffatt & Ricca (1992). Details of the mechanism by which twist may be converted to writhe have been recently investigated by Longcope & Klapper (1997).

4 Relaxation of knotted fields

Suppose now that the initial field $\mathbf{B}_0(\mathbf{x})$ of §2 is taken to be a 'tubular field', each pair of field lines in the tube having winding number h as described above. Under the relaxation process of §2, each field line 'wishes' to contract in length. During this process, the helicity $\mathcal{H} = h\Phi^2 = (Wr + Tw)\Phi^2$ is constant, and there may have to be some kind of trade-off between writhe and twist contributions. The volume V of the tube is also constant (under the assumption of incompressibility). Under these conditions, the minimum (or equilibrium) energy M^E is determined by the flux Φ , the volume V and the helicity \mathcal{H} , and by no other parameters. On dimensional grounds, it must take the form

$$M^E = m(h)\Phi^2 V^{-1/3} \quad (24)$$

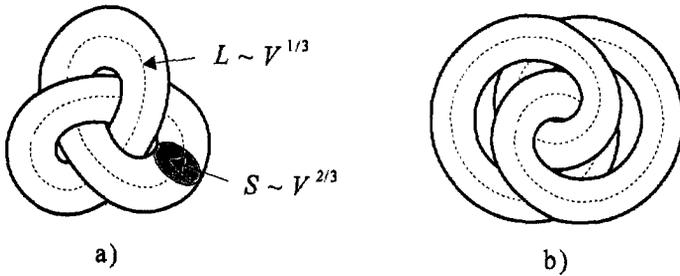


Figure 2: Two distinct minimum energy states that may be expected when K is the trefoil knot.

where $m(h)$ is a (dimensionless) function of the dimensionless parameter h (Moffatt 1990). We have until now restricted h to be an integer; however we may now allow h to be any real number; if h is rational, then the \mathbf{B} -lines in \mathcal{T} are closed curves, while if it is irrotational they do not close, but lie on nested toroidal surfaces with axis K . Note that, by Dehn surgery (cut, twist and reconnect the tube) the value of h may be set to any prescribed value for the initial field. This prescription of h is equivalent to prescribing a ‘framing’ of the knot. The particular choice $h = 0$ constitutes ‘zero-framing’. This for example is the choice adopted by Berger & Field (1984).

Suppose we start with this natural choice $h = 0$. As noted in §4, relaxation proceeds through contraction of \mathbf{B} -lines; for an untwisted tube, this can be best accomplished by contraction of the axis of the tube and corresponding increase of cross-section (to conserve volume). This process is arrested when the tube gets so ‘fat’ that it makes contact with itself in such a way that no further increase of cross-section is achievable. In this sense, the attainment of a minimum has clear parallels with the concept of ‘ideal’ knot configuration, developed by Katritch *et al* (1996). Of course, in the present context, the tube cross-section is not constrained to remain circular, and some flattening of the cross-section in regions where transverse contacts are made is to be expected.

A final caveat is needed: even when $h = 0$, there is no guarantee of uniqueness for the ultimate minimum energy state attained under a relaxation process. There may be more than one minimum (each being a ‘local’ minimum in the function space of accessible configurations), and in this case we should speak of the ‘energy spectrum’ of the knot, this being the whole set of energy levels of minimal states. Figure 2 shows two distinct minimum energy

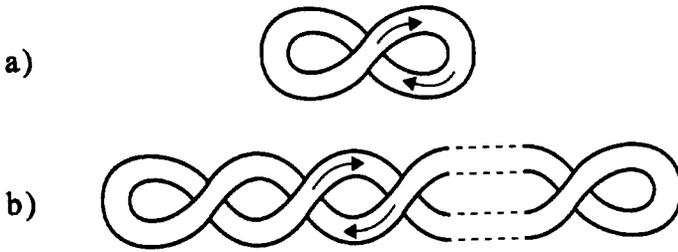


Figure 3: Preferred configurations of an unknotted but twisted tube, a) for $h = 1$, b) for $h = n > 1$.

states expected when K is the trefoil knot. This type of configuration has been studied in detail by Chui & Moffatt (1995).

5 Relaxation of strongly twisted tubes

If $|h|$ is large, then a very different behaviour is to be expected. In this situation, the twist component of field B_θ in the tube \mathcal{T} is strong compared to the axial component B_s (θ being an angular coordinate in the plane of cross-section of \mathcal{T}), and relaxation is achieved by shrinkage of the cross-section S (with then B_θ decreasing like $S^{1/2}$) and associated *increase* in length L of the tube axis (with $V = LS = \text{cst.}$, and $B_s \sim L \sim S^{-1}$). This process reaches equilibrium when B_θ and B_s are of the same order of magnitude, with $B_s = \Phi/S$, $B_\theta \sim |h|\Phi S^{1/2}/V$, giving $S \sim |h|^{-2/3}$, $L \sim |h|^{2/3}$. The associated magnetic energy is of order $B_s^2 V$, i.e.

$$M^E \sim |h|^{4/3} \Phi^2 V^{-1/3}, \quad (25)$$

and this result apparently holds irrespective of knot type K .

There is however a further effect that leads to reduction of magnetic energy. This reduction can be achieved by a 'kink instability', analogous to the instability of an elastic wire subject to twist. This instability occurs even for an unknotted configuration; for example if the tube \mathcal{T} has circular axis, and it is subjected to unit twist ($h = 1$), then it will 'prefer' to deform to a figure-of-eight configuration for which the twist Tw is reduced to zero, and the writhe Wr is increased to 1 (figure 3a); the increase in writhe energy in this deformation is more than compensated by the complete loss of twist energy.

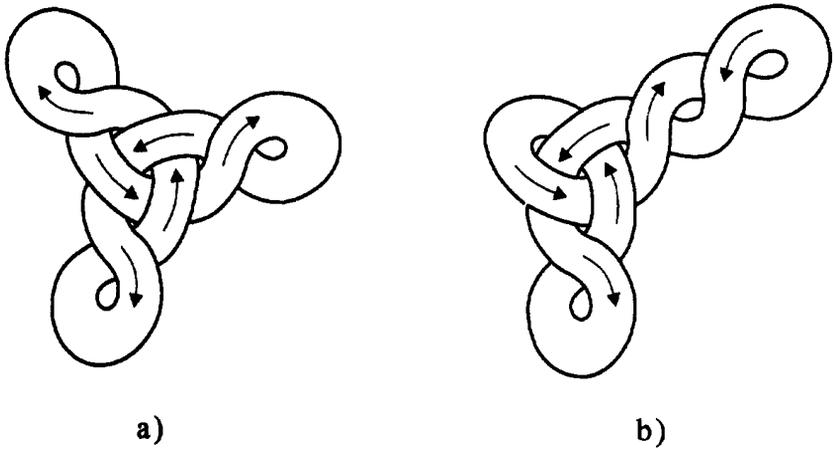


Figure 4: Conjectured minimum energy configurations for the trefoil knot when $h = 6$. a) symmetric state; b) non-symmetric state.

These energy changes have been calculated for the case of a twisted wire with standard elastic properties by Wadati *et al* (1989) (see also Calladine 1980), and the calculation for a magnetic flux tube is very similar.

Likewise, if h is any positive integer n , then again twist Tw may be reduced to zero via the deformation indicated in figure 3b. In this relaxed configuration, $Wr = n$ and the tube length is $L \sim 2nS^{1/2} = 2hS^{1/2}$; the magnetic energy $M = \frac{1}{2}B^2V$ with $B = \Phi/S$ and $V = LS$ is still given in order of magnitude by (25); the reduction in energy that can be achieved by the kink instability does not therefore change its order of magnitude. Its precise calculation involves determination of the dimensionless constant of proportionality in (25) for the two competing configurations (one with $Tw = n$, $Wr = 0$; the other with $Tw = 0$, $Wr = n$); this remains an open problem at present, although one may conjecture with confidence that the second configuration (depicted in figure 3b) wins. Similarly, for any knotted tube, twist can be converted to writhe through kink instabilities in disjoint sections. Figure 4a illustrates what may happen for the trefoil knot with $h = 6$. The tube can adjust itself so that $Tw = 0$, and $Wr = 3 + (3 \times 1)$ is accounted for by the writhe ($= 3$) of the trefoil knot itself and the additional writhe ($= 1$) of each of the three twisted arms in a symmetrical configuration. The possibilities for multiple (non-symmetric) minimum energy states become apparent here (figure 4b).

References

1. M.A. Berger and G.B. Field 1984 The topological properties of magnetic helicity. *J. Fluid Mech.* **147**, 133–148.
2. C.R. Calladine 1980 Toroidal elastic supercoiling of DNA. *Biopolymers* **19**, 1705–1713.
3. G Călugăreanu 1959 L'intégral de Gauss et l'analyse des noeuds tridimensionaux. *Rev. Math. pures appl.* **4**, 5–20.
4. G Călugăreanu 1961 Sur les classes d'isotopie des noeuds tridimensionaux et leurs invariants. *Czechoslovak Math. J.* **11**, 588–625.
5. A.Y.K. Chui and H.K. Moffatt 1995 The energy and helicity of knotted magnetic flux tubes. *Proc. R. Soc. Lond. A* **451**, 609–629.
6. A.Y.K. Chui and H.K. Moffatt 1996 Instability of magnetic modons and analogous Euler flows. *J. Plasma Physics* **56**, 677–691.
7. M.H. Freedman 1988 A note on topology and magnetic energy in incompressible perfectly conducting fluids. *J. Fluid Mech.* **194**, 549–551.
8. G Katritch, J Bednar, D Michoud, R.G. Scharein, J Dubochet and A Stasiak 1996 Geometry and physics of knots. *Nature* **384**, 142–145.
9. D Linardatos 1993 Determination of two-dimensional magnetostatic equilibria and analogous Euler flows. *J. Fluid Mech.* **246**, 569–591.
10. D.W. Longcope and I Klapper 1997 Dynamics of a thin twisted flux tube. *Ap. J.* **488**, 443–453.
11. H.K. Moffatt 1969 The degree of knottedness of tangled vortex lines. *J. Fluid Mech.* **36**, 117–129.
12. H.K. Moffatt 1985 Magnetostatic equilibria and analogous Euler flows of arbitrarily complex topology. Part 1. Fundamentals. *J. Fluid Mech.* **159**, 359–378.
13. H.K. Moffatt 1990 The energy spectrum of knots and links. *Nature* **347**, 367–369.
14. H.K. Moffatt and R.L. Ricca 1992 Helicity and the Călugăreanu invariant. *Proc. R. Soc. Lond. A* **439**, 411–429.
15. M Wadati and H Tsuru 1986 Elastic model of looped DNA. *Physica* **21D**, 213–226.

CHAPTER 13

KNOTS IN BISTABLE REACTING SYSTEMS

ANATOLY MALEVANETS

*Department of Physics, Theoretical Physics, 1 Keble Road,
Oxford OX1 3NP, England*

RAYMOND KAPRAL

*Chemical Physics Theory Group, Department of Chemistry,
University of Toronto, Toronto M5S 3H6 Canada*

Bistable chemically reacting media can segregate into domains of two stable phases that differ in their chemical composition and are separated by chemical fronts. For activator-inhibitor kinetics, when the diffusion coefficient of the inhibitor is greater than that of the activator, stable localized chemical patterns can form. These localized structures arise from the interaction between chemical fronts as a consequence of the fast inhibitor diffusion. In three dimensions one can build linked and knotted chemical dissipative structures whose stability is a consequence of their topology. In the general case, the evolution equations that govern the pattern formation are not of gradient type but the stable chemical patterns assume nearly ideal forms.

1 Introduction

Open chemically reacting systems forced out of equilibrium by flows of reagents may evolve to patterned stationary or time-dependent states.¹ We shall be concerned with bistable reacting media where the system may evolve to either of two stable states that differ in their chemical compositions. In a spatially-distributed system these two stable states may exist in different regions of space separated by interfaces across which the chemical concentrations change sharply, akin to a mixture of oil in water. In three dimensions a tubular domain containing one of the stable phases may have the topology of a link or knot embedded in a "sea" of the other stable phase.^{2,3} These compact stable structures rely on their topology for their existence: in the absence of a linked or knotted geometry the tubular domains that comprise these structures shrink to balls. Such compact stable structures are termed topologically stabilized patterns.

These topologically stabilized patterns are seen in multi-component reaction-diffusion systems which, in general, are not of gradient form: the patterns are dissipative structures. It is then interesting to consider the factors that determine the final geometries of the linked and knotted patterns. We shall show that the links and knots are approximately ideal⁴ and this fact can be ascribed to the interactions that are responsible for their existence.

Section 2 describes the FitzHugh-Nagumo (FHN) reaction-diffusion system used to illustrate the formation of topologically stabilized patterns. This model, which has its origins in nerve impulse propagation, is believed to capture the essential features of real bistable media which support localized chemical patterns, the antecedents of the knotted topologies discussed here. Our simulations are actually carried out on a microscopic reaction model whose mean field limit is the FitzHugh-Nagumo equation. We describe the construction of the stochastic model in Sec. 3. As a result, the chemical patterns we find arise from the random reactive and diffusive motion of millions of particles and their existence demonstrates the robustness of the chemical patterns to molecular fluctuations. The localized linked and knotted patterns are analyzed in Sec. 4 on the basis of their ideal structures, while the conclusions of this study are given in Sec. 5.

2 Competing Interactions in Bistable Media

The chemical patterns we shall describe arise from competing activator-inhibitor kinetics. The FitzHugh-Nagumo reaction-diffusion equation,^{5,6}

$$\begin{aligned}\frac{\partial u}{\partial t} &= -u^3 + u - v + D_u \nabla^2 u, \\ \frac{\partial v}{\partial t} &= \epsilon(u - \alpha v - \beta) + D_v \nabla^2 v,\end{aligned}\quad (1)$$

is such an activator-inhibitor system. An increase in the concentration of the activator u leads to an increase in its own production and that of v ; an increase in the concentration of the inhibitor v suppresses the production of chemical species. The interplay between these two “species” is responsible for the effects we shall describe.

For the moment suppose the system is spatially homogeneous and the diffusion terms in Eq. (1) are absent. The u and v nullclines of the resulting ordinary differential equation, the equations for $\dot{u} = 0$ and $\dot{v} = 0$, are shown in Fig. 1. As drawn, there are three intersections between the u and v nullclines. Linear stability analysis shows that the outer two fixed points are stable while the central fixed point is unstable. This is the bistable regime discussed above: depending on the initial values of u and v , the system will evolve to either of these two stable fixed points.

Next, suppose we reintroduce the diffusion terms and consider a spatially-distributed system described by Eq. (1). Inhomogeneous initial states will evolve to form domains of the two stable phases. The structure and dynamics of these domains, which depend on the interplay between the activator-inhibitor

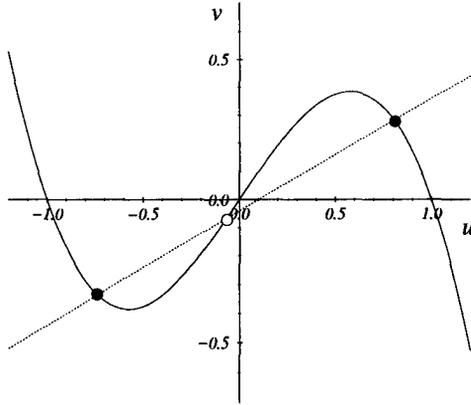


Figure 1: Nullclines of the FitzHugh-Nagumo equation for parameters in the bistable regime. The linear v nullcline intersects the cubic u nullcline in three fixed points; the outer two denoted by filled circles are stable while the central fixed point denoted by an open circle is unstable.

kinetics and the relative values of the diffusion coefficients of these species, are the topics of this chapter.

The FHN reaction-diffusion equation is not of gradient form and there is no free energy functional that controls the evolution to the final attracting states. In order to examine the structure of Eq. (1) it is convenient to formally solve the linear equation for the inhibitor v and write the equation for the activator u as,

$$\frac{\partial u(\mathbf{r}, t)}{\partial t} = -u^3 + u + D_u \nabla^2 u - \int_V d\mathbf{r}' \int dt' G(\mathbf{r} - \mathbf{r}', t - t') \epsilon(u(\mathbf{r}', t') - \beta), \quad (2)$$

where the Green function $G(\mathbf{r}, t)$ is given by

$$G(\mathbf{r}, t) = (4\pi D_v t)^{-3/2} e^{-\alpha t - r^2/4D_v t}. \quad (3)$$

From this form we can see that the activator field at space-time point (\mathbf{r}, t) is influenced by its value at point (\mathbf{r}', t') due to coupling with the inhibitor field. In this general form the evolution equation cannot be written as the gradient of a free energy functional. However, if ϵ is large so that v varies on a much

faster time scale than that of u , we may let $u(\mathbf{r}', t') \approx u(\mathbf{r}', t)$ in the integral in Eq. (2) and, after integration, obtain,

$$\frac{\partial u(\mathbf{r}, t)}{\partial t} = -u^3 + u + \frac{\beta}{\alpha} + D_u \nabla^2 u - \int_V d\mathbf{r}' G(\mathbf{r} - \mathbf{r}') u(\mathbf{r}', t), \quad (4)$$

with

$$G(\mathbf{r}) = (4\pi \tilde{D}_v r)^{-1} e^{-\kappa r}. \quad (5)$$

Here $\kappa = (\alpha/\tilde{D}_v)^{1/2}$ and the scaled diffusion coefficient of the activator is $\tilde{D}_v = D_v/\epsilon$. Now Eq. (4) can be written in gradient form

$$\frac{\partial u(\mathbf{r}, t)}{\partial t} = -\frac{\delta \mathcal{F}}{\delta u}, \quad (6)$$

where the nonlocal free energy functional \mathcal{F} is

$$\mathcal{F} = \int d\mathbf{r} \left(\mathcal{V}(u) + \frac{D_u}{2} |\nabla u|^2 \right) + \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' u(\mathbf{r}) G(\mathbf{r} - \mathbf{r}') u(\mathbf{r}'). \quad (7)$$

The potential \mathcal{V} is defined as $\mathcal{V}(u) = u^4/4 - u^2/2 - \beta u/\alpha$. With these results in hand we may review some of the features of this type of activator-inhibitor kinetics. First, we note that if the scaled diffusion coefficient of the inhibitor tends to zero, Eq. (4) reduces to the time-dependent Ginzburg-Landau equation (model A for a non-conserved order parameter of critical phenomena⁷). It is well known that this equation exhibits domain coarsening: if the system is prepared in the unstable state domains of the two stable phases will form and grow. The growth law and the nature of the domains depends on whether the potential \mathcal{V} is symmetric or asymmetric. For an asymmetric free energy potential with $\beta \neq 0$, if two phases are separated by a planar interface, the more-stable phase will consume the less-stable phase at a rate determined by the planar front velocity. For more complicated phase geometries the front velocity and domain curvature determine the coarsening rate.⁸

The situation changes if $D_v \gg D_u$. As indicated above in the fast inhibitor limit the additional long-range term in the free energy functional gives rise to a competing interaction that leads to the possibility of modulated phases or localized patterns. These localized patterns are observed both in the fast inhibitor limit (Eq. (4)) and in the general case or slow inhibitor limit (Eq. (2)).⁹⁻¹⁴ When the inhibitor diffuses rapidly compared to the activator, the long-ranged v field influences the interactions among the u -field fronts, leading to front repulsion in certain parameter regimes. Such front repulsion can give rise to isolated structures of the less-stable phase in a sea of the more

stable phase.^{9,10} In the absence of such front repulsion the small domains of the less-stable phase would be consumed by the more-stable phase.

We also note that for a different model with a nonlocal energy functional it was demonstrated^{15,16} that compact stable patterns exist. The stable patterns are defined by the local minima of the functional and their form is defined by the tendency of the system to maximize surface contact area at a fixed distance and minimize the volume of the minority state. These structures are closely related to the two dimensional patterns that exist in the FHN system in the large ϵ limit^{11,12} and their shape can take rather complicated forms.

We shall be concerned with the limit where the inhibitor diffuses rapidly compared to the activator and localized structures are possible. The basic building block for the three-dimensional chemical patterns discussed here is a tubular domain of the less-stable phase in a “sea” of the more-stable phase. Depending on the system parameters, such tubular domains may either grow, with fingering instabilities in some parameter ranges or shrink to balls as shown in Fig. 2.³ Recall that because chemical reactions can change the numbers of

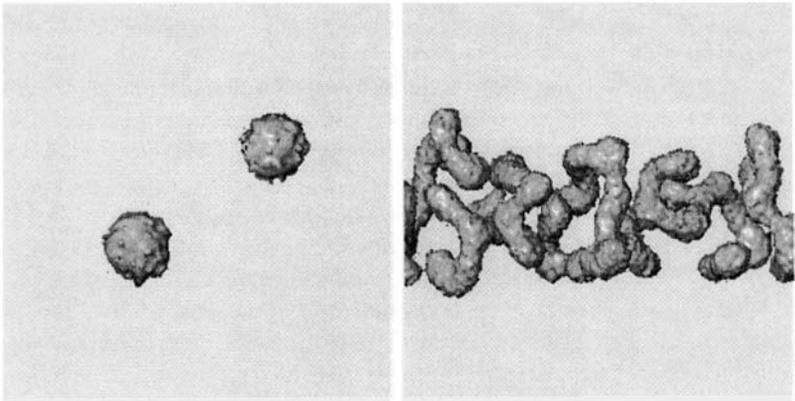


Figure 2: Contraction of a tubular domain to a ball (left panel); growth of a tubular domain by fingering instabilities (right panel). System parameters are: $\alpha = 5.2$, $\beta = 0.33$, $\epsilon = 0.0055$ and $\alpha = 5.4$, $\beta = 0$, $\epsilon = 0.0085$, respectively. The diffusion ratio is $D_b/D_a = 4$.

chemical species we have non-conserved dynamics and the amount of a given phase is not fixed by the initial preparation of the system.

Before presenting a discussion how the above considerations lead to the possibility of stable linked and knotted patterns, we give an account of the stochastic dynamics that is actually used to simulate the chemical patterns. The use of a stochastic model, rather than the FHN reaction-diffusion equation,

not only allows us to probe the mesoscale origin of chemical pattern formation, but also demonstrates that the chemical patterns we find are robust to molecular fluctuations.

3 Stochastic Model

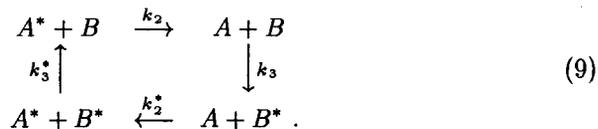
Microscopic Markov chain models utilize stochastic rules that simulate both the reactive collision events that are responsible for the interconversion of chemical species and the random walk dynamics that lead to diffusive transport in the medium. The u and v variables in the FHN model do not directly correspond to the concentrations of chemical species, nor is a chemical mechanism usually associated with this kinetic equation.

In order to formulate a microscopically-based stochastic model it is first necessary to provide a mechanism whose mass action law is the FHN kinetic equation. Some features of the FHN kinetics seem to preclude such a mechanistic description; for instance, the production of u is inhibited by a term linear in v , a contribution not usually encountered in mass action kinetics. However, if each local region of space is assumed to be able to accommodate only a maximum number N of each chemical species, then such a mechanism may be written.^{2,3} We assume that the chemical reactions depend on the local number of molecules of the species as well as the number of vacancies corresponding to each species, in analogy with the dependence of some surface reactions on the number of vacant surface sites or biochemical reactions involving complexes of allosteric enzymes that depend on the number of vacant active sites.

The reaction mechanism consists of the cubic autocatalytic steps involving species A and its corresponding vacancy species A^* ,



This part of the mechanism is responsible for the bistability of the system but, in the absence of other coupling, will simply give rise to stable states corresponding to full occupancy or complete vacancy at a site. A second species B involved in bimolecular reactions with A is needed to account for both the linear and cubic parts of the FHN model. The following cyclic reaction steps accomplish this:



Many mechanisms may give rise to the same chemical rate law and this mechanism may be considered as the definition of the microscopic reactive dynamics. Assuming mass action kinetics, we may write the following chemical rate law from a knowledge of this mechanism:

$$\begin{aligned}\frac{da}{dt} &= \{k_1a - k_1^*(1-a)\}a(1-a) + k_2(1-a)b - k_2^*a(1-b), \\ \frac{db}{dt} &= k_3^*(1-a)(1-b) - k_3ab,\end{aligned}\quad (10)$$

where a and b are the average fractional occupancies of a site. For $k_2^* = k_2$ and $k_3^* = k_3$, these equations can be converted to FHN form with the linear change of variables, $a = c_a u + a_0$ and $b = c_b v + b_0$ where

$$\begin{aligned}c_a^2 &= \frac{1}{3} \left(\frac{k_1 + 2k_1^*}{k_1 + k_1^*} \right)^2 - \frac{k_2 + k_2^*}{k_1 + k_1^*}, \quad a_0 = \frac{1}{3} \frac{k_1 + 2k_1^*}{k_1 + k_1^*}, \\ c_b &= -\frac{k_1 + k_1^*}{k_2} c_a^3, \quad b_0 = a_0 \frac{c_b}{c_a} \left\{ 1 - \left(\frac{a_0}{c_a} \right)^2 \right\},\end{aligned}\quad (11)$$

and use of the scaled time variable $\tau = t/\tau_s$ with $\tau_s^{-1} = (k_1 + k_1^*)c_a^2$. The parameters in (1) are related to the rate constants by

$$\alpha = -\frac{c_b}{c_a}, \quad \beta = \frac{1 - a_0 - b_0}{c_a}, \quad \epsilon = \frac{k_3}{k_2} \left(\frac{c_a}{c_b} \right)^2.\quad (12)$$

Using these relations one may choose desired values of the FHN parameters α , β and ϵ by tuning the values of the rate constants in the mechanism. In this way we can select parameter domains that are likely to favor isolated domain structures where linked and knotted patterns are expected to be found.

To construct a Markov chain model we suppose the dynamics takes place on a regular lattice and that the local spatial regions are associated with sites on this lattice. The Markov chain model then requires that transition probabilities be defined for the local reactive transformations at a lattice site and for the hops from a given site to neighboring sites in order to simulate the random walks that lead to diffusive motion of the species. Here we give a brief account of the strategy used to build the stochastic model and its relation to the mass action rate law. In the Appendix we give a more detailed analysis of the Markov chain to show how the reaction-diffusion equation can be obtained.

Once a reaction mechanism has been specified, the reaction transition probabilities are easily constructed by implementing birth-death stochastic rules¹⁷ for the reactions. At each time step one of the six reactions in the

mechanism (labeled by the index j) is chosen at random and takes place with a probability given by

$$p_j = \gamma \bar{k}_j N \prod_{\kappa=A}^B \frac{n_\kappa!}{(n_\kappa - \nu_\kappa^{(j)})!} \frac{n_\kappa^*!}{(n_\kappa^* - \nu_\kappa^{*(j)})!} \frac{(N - \nu_\kappa^{(j)} - \nu_\kappa^{*(j)})!}{N!}, \quad (13)$$

where $n_\kappa^* = N - n_\kappa$, $\nu_\kappa^{(j)}$ and $\nu_\kappa^{*(j)}$ are the stoichiometric coefficients for species κ and κ^* , respectively, in reaction j and $\bar{k}_{2i-1} = k_i$ and $\bar{k}_{2i} = k_i^*$ for $i = 1, 2, 3$. Here γ is a constant that determines the overall time scale of the reactive transformations. These reaction probabilities simply state that the probability of a given reaction step is proportional to the macroscopic rate constant for that step, times the number of possible ways that reaction can occur for the given numbers of A and B molecules at the site.

To simulate diffusion, random walk transition probabilities are chosen so that the probability of a hop to a neighboring node $p(n_\kappa)$ is linear in the fractional occupancy of the site and $p(0) = 0$ and $p(N) = 1$ to insure that the occupation number never falls below zero or is greater than N (exclusion principle is always satisfied).^{2,3} This choice of transition probabilities guarantees that the diffusion coefficient is independent of n_κ . One may show that this diffusion rule leads to a binomial stationary distribution (cf. Appendix),

$$p_{n_A}^B = \binom{N}{n_A} a^{n_A} (1-a)^{N-n_A}, \quad (14)$$

with a similar equation for species B . If diffusion can maintain the a local binomial distribution during the course of slow reactive interconversions between A and B species, the time rate of change of of the average particle densities due to reactions is given by

$$\frac{d\bar{n}_\kappa}{dt} = \sum_{j=1}^6 \sum_{n_A, n_B} (\Delta n_\kappa)_j p_j p_{n_A}^B p_{n_B}^B, \quad (15)$$

where $(\Delta n_\kappa)_j$ is the change in particle number of species κ in the j th reaction. Direct evaluation of the r.h.s. of this equation leads to the mass action law Eq. (10). This confirms that the Markov chain model yields the mass action law in the limit of perfect diffusive mixing.

In the Appendix we show that the reaction-diffusion equation, Eq. (1), is obtained in the limit of slow reaction and small spatial gradients. Consequently, we may use the Markov chain model in suitable parameter regimes to simulate the reaction-diffusion system (this is the limit with which we shall be primarily concerned in this chapter) or investigate the breakdown of such macroscopic models due to reactive correlations.²

3.1 Simulation details

In the simulations reported in this chapter the random walk dynamics is realized by use of an auxiliary “excited” particle lattice. At each step at most one particle per site is transferred to the excited state with a probability depending on the site occupation number as described above. Next, the excited particles are translated one lattice unit in a random direction chosen from a set $V = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. Finally, the excited particles are placed in new lattice positions. For the three-dimensional simulations, rather than using the simple cubic lattice, we have taken the set V to represent the projections of the normals of the four-dimensional FCC Wigner-Seitz cell on the three-dimensional space.¹⁸ The diffusion rule leads to mixing of particles on the lattice and, to a good approximation, it establishes a local binomial probability distribution of particle numbers.² The use of a 24-direction diffusion rule is motivated by symmetry considerations. The FCC set of directions given by the six permutations of $(\pm 1, \pm 1, 0, 0)$ leads to a spherically symmetric discrete Laplacian with fourth-order corrections in the lattice spacing.

The kinetic and diffusion parameters were the same for all simulations with the exception of the growing patterns in Fig. 2 (right panel) whose parameters are given in the caption. The parameters were selected to be: $\tilde{k}_1 = 0.862$, $\tilde{k}_1^* = 0.76$, $\tilde{k}_2 = 0.04$ and $\tilde{k}_3 = 0.006$. The maximum occupancy per node was taken to be $N = 7$ and the diffusion coefficient ratio was $D_b/D_a = 4$. The kinetic parameters correspond to the following FHN parameters: $\alpha = 5.21$, $\beta = 0.33$ and $\epsilon = 0.0055$. The calculations were performed on $256 \times 256 \times 256$ lattices for the Borromean rings and 4_1 knot, on $256 \times 256 \times 128$ lattices for the 3_1 knot and on $256 \times 128 \times 128$ lattices for the Hopf link. The simulations were carried out on a CAM-8 machine¹⁹ that is designed for lattice-gas models of the type used here.

4 Links and Knots

In this section we show how the domain dynamics and existence of localized structures described earlier can lead to stable links and knots. We consider the regime where tubular domains shrink. A tubular domain shrinks by contracting at its free ends, maintaining a fixed diameter until the final stages of the contraction where a ball is formed as in Fig. 2. Suppose that the tubular domain has no free ends but is instead a completely connected domain in three-dimensional space. Now the entire continuous tubular domain will contract with diffusion providing an effective mean force that tends to reduce the curvature. Such contraction may again yield a ball but it is possible that the

local repulsion between fronts that arises from the rapidly diffusing inhibitor species will be sufficiently strong that collapse is prevented and a stable compact structure results which retains the linked or knotted topology of the initial condition.^a That this is indeed the case is demonstrated by the Hopf and Borromean links and 3_1 and 4_1 knots shown in Figs. 3 and 4. In these figures

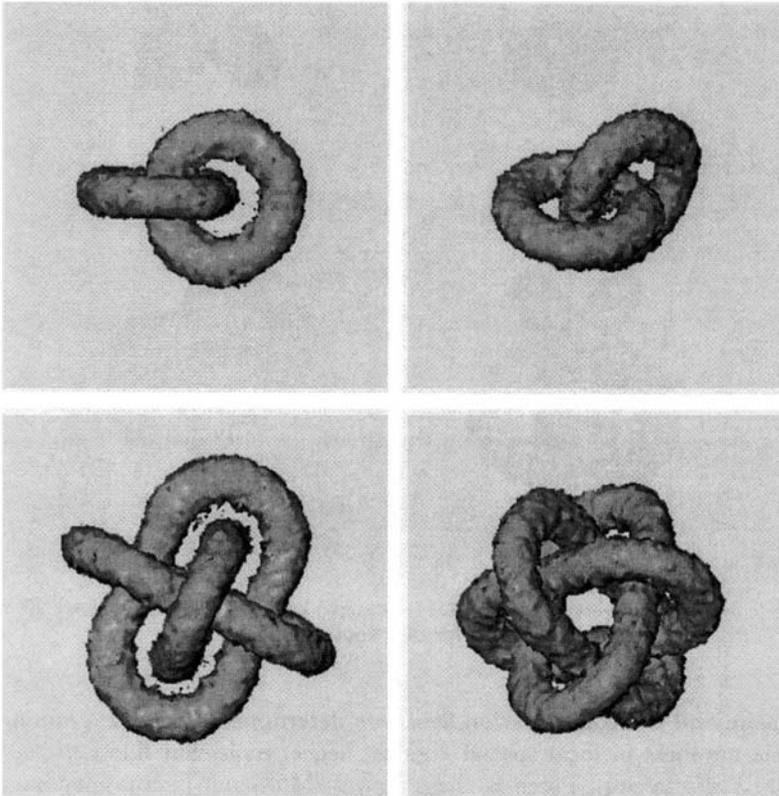


Figure 3: Hopf (800K time steps) and Borromean (> 1M time steps) links. Two different views are shown for each link.

isoconcentration surfaces corresponding to the activator are shown. The imperfections in these surfaces are due primarily to the fact that the dynamics is

^aThe stable localized three-dimensional patterns described in Refs.^{15, 16} are akin to the stable ball-like configurations in Fig. 2 and are not topologically stabilized like the structures considered in this chapter.

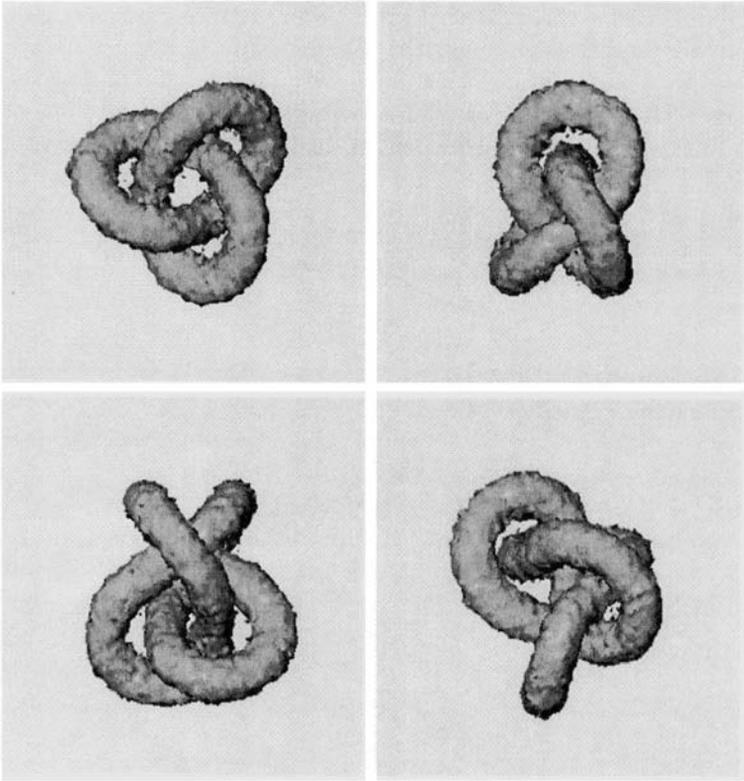


Figure 4: 3_1 (2.6M time steps) and 4_1 (4M time steps) knots. Two different views are shown for each knot.

stochastic and the concentration fields are determined by coarse graining the particle numbers in local spatial regions; hence, molecular fluctuations exist and their effects can be seen in these figures. Molecular fluctuations have another consequence: none of these patterns is truly stable but rather metastable since fluctuations can occasionally lead to breaking of a tubular domain. This fluctuation effect can be controlled by changing the spatial scale by tuning the diffusion relative to reaction. The parameters in our simulations are such that fluctuation effects are almost negligible and the simulation results approximate those of the FHN reaction-diffusion equation. We have never observed fluctuation breaking of the Borromean rings or 3_1 or 4_1 knots on the time scale of our simulations; we have observed breaking of the Hopf link for sufficiently long times for some realizations of the stochastic evolution.

Plots of the local curvature and torsion provide quantitative information about the geometrical structure of the patterns. The curvature $\kappa(\sigma)$ is defined as,²⁰

$$\kappa = \frac{|\mathbf{R}_\sigma \times \mathbf{R}_{\sigma\sigma}|}{|\mathbf{R}_\sigma|^3}. \quad (16)$$

To compute this quantity the curve $\mathbf{R}(\sigma)$ was extracted from an analysis of the concentration fields of the links and knots. This curve was constructed so that the parameterization of the line σ was approximately proportional to the natural parameterization s . Equation (16) is invariant with respect to the change of variable $\sigma = \sigma(s)$. The curvatures of the 3_1 and 4_1 knots are shown in Fig. 5. From this figure one can see that the 3_1 knot has the expected three-

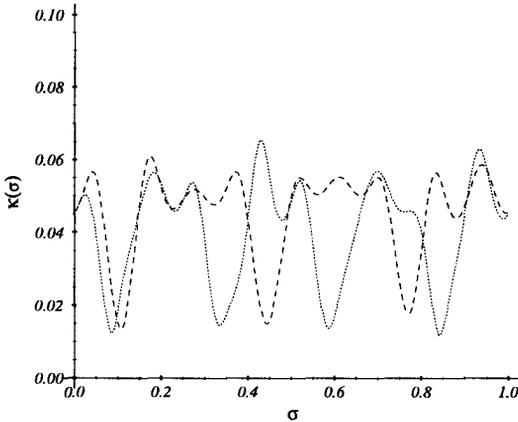


Figure 5: Local curvatures as a function of σ for the 3_1 knot (dashed line) and the 4_1 knot (dotted line).

fold symmetry to a good approximation. Its curvature is nearly constant at a value of around $\kappa = 0.05$ except for three regions of σ where it assumes small values. In these regions the tubular domain passes through a knot loop and the tube has a nearly linear segment. More significant variations of the curvature with σ are seen for the 4_1 knot. Its more complex form leads to additional regions where a tubular domain must pass through a knot loop leading to stronger variations in the curvature.

An examination of the local torsion, $\tau(\sigma)$, defined by

$$\tau = \frac{(\mathbf{R}_\sigma \times \mathbf{R}_{\sigma\sigma}) \cdot \mathbf{R}_{\sigma\sigma\sigma}}{|\mathbf{R}_\sigma \times \mathbf{R}_{\sigma\sigma}|^2}, \quad (17)$$

provides information on the planarity of the structures. The local torsion is shown in Fig. 6. One can see large segments of σ where the torsion is near

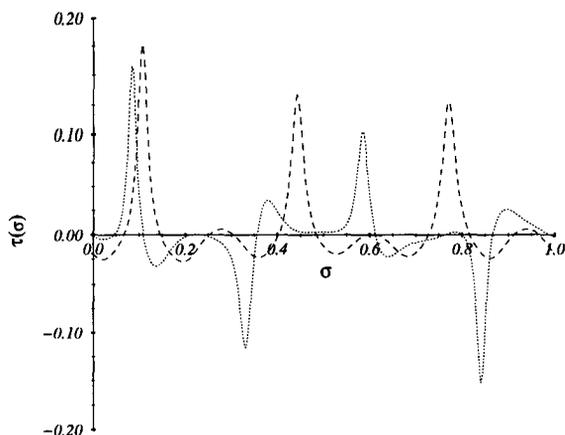


Figure 6: Local torsion as a function of σ of the 3_1 (dashed line) and 4_1 (dotted line) knots.

zero for both knots, indicating planarity. The 3_1 knot is built from three such planar regions as expected (one can see some deviation from perfect symmetry likely due to transient effects in the relaxation to the final stable knot), while the 4_1 knot is built from four such planar regions with smaller extent.

4.1 Ideal links and knots

It is difficult to predict the final stable forms of the links and knots in the bistable reacting medium. As discussed above, in general there is no free energy functional whose minimization yields the stable structures. The main feature that is responsible for the formation of these chemical patterns is the front repulsion that arises through the mediation of the fast diffusion of the inhibitor field. Some insight into the concentration fields that underlie these linked and knotted structures can be gained by examining the concentration

profiles of both the activator and inhibitor species. In the colour plate we display the composite picture of the two iso-surface fields: B -field concentration 0.16 and A -field concentration 0.5. The A field is sharp and its iso-surface is insensitive to the exact value of the concentration. From the form of the iso-surface profiles we deduce that the filament repulsion is mediated by the B field and the thickness of the resulting tube is approximately independent of the filament surroundings. A two-dimensional picture of these concentration profiles for a cut through a three-dimensional Borromean ring pattern is shown in Fig. 7. One sees the features described above: the A field is very sharp while the B field concentration that couples the fronts is diffuse.

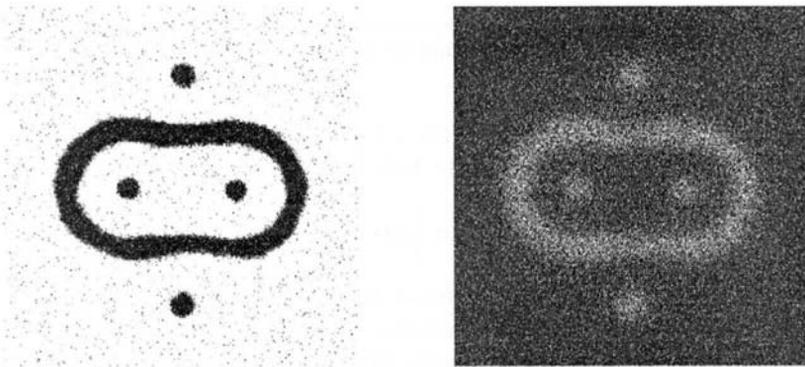


Figure 7: Concentration fields in a two-dimensional cut through the Borromean rings.

We now discuss the ideality of the linked and knotted chemical patterns. An ideal form of a knot or link, constructed from a tube of uniform diameter, has the highest ratio of volume to surface area.⁴ The knotted patterns in the bistable chemical medium appear as iso-surfaces of the concentration fields and can be approximated to a good degree by uniform tubes drawn around closed filaments. Some deviations from ideality are to be expected since we can see that the tubular domains corresponding to the iso-concentration surfaces do not have uniform diameters. However, these deviations are not large and it is interesting to examine the extent to which such ideality applies.

To determine the ideal configurations we extracted the curves $\mathbf{R}(\sigma)$ from the three-dimensional concentration patterns as discussed above. Tubular domains with given radii were then constructed with the curves $\mathbf{R}(\sigma)$ at their centers. The ideal configurations of the 3_1 and 4_1 knots are shown in Fig. 8

Deviations from uniformity of the knotted patterns make it preferable to use integral characteristics of the patterns for the extraction of the system



Figure 8: Approximately ideal representations of the 3_1 and 4_1 knots. A tubular domains of radius 17 are built around the middle line of the patterns.

parameters. The Weyl formula for the volume of the tube in three dimensions relates the volume of the tube to the tube length and diameter:

$$V = \frac{1}{4}\pi D^2 L, \quad (18)$$

where V , D and L are the tube volume, diameter and length, respectively. In Table 1 we compare the computed volume to diameter ratio for two iso-surface concentrations and compare the results with the data for the ideal knots.⁴

| Pattern | length(L) | volume $\times 1/64$ | | L/D | | |
|------------|----------------|----------------------|--------|--------|--------|-------|
| | | (0.13) | (0.16) | (0.13) | (0.16) | ideal |
| 3_1 knot | 666 | 12968 | 8437 | 16.7 | 20.5 | 16.4 |
| 4_1 knot | 887 | 17983 | 11392 | 21.8 | 27.4 | 21.2 |
| Borromean | 427×3 | 25909 | 16265 | 10.5 | 13.3 | |
| Hopf | 264×2 | 10291 | 6715 | 6.6 | 8.2 | 6.1 |

Table 1: Length to diameter ratios for various knots and links.

From this Table we observe that the radii of the tubular domains, determined using Eq. (18) and the measured lengths and volumes, are the same, regardless of the type of link or knot. The diameter of the tubular domain is determined by the front repulsion, which depends on the local kinetics and

diffusion coefficients, and the curvature of the front. Since the diffusion coefficients and kinetic parameters are the same for all knots and links, we expect the diameters of the tubular domains to be the same. The nearly ideal forms are achieved through evolution where the volume of the less-stable phase shrinks, maintaining a constant diameter, until the final stable pattern is reached. We find $D \approx 20$ and $D \approx 16$ for the 0.13 and 0.16 iso-concentration surfaces, respectively. This indicates that a consistent tubular representation is possible for all structures. The theoretical radii of the patterns, computed using the length of the filaments and the theoretical length to diameter ratios^b, are close to those of the 0.13 iso-concentration surface, indicating the ideal character of the observed knots and links.

5 Conclusion

The linked and knotted patterns we have described in this chapter should be observable in experiments on chemical and other systems. Experiments on the iodide-ferrocyanide-sulfite system²¹⁻²³ in quasi two-dimensional geometries have shown the existence of labyrinthine and patterns and localized structures which involve front repulsion for their existence. The main difficulty in finding the three-dimensional topologically stabilized structures is the experimental preparation of the proper initial conditions. Perhaps one way to find such structures is to start from random initial conditions. Since the compact patterns lie in the parameter regime where tubular domains shrink to balls, any structures that are not topologically stabilized will shrink and the structures of interest will survive.

The chemical system we have considered provides an example where the final linked or knotted geometry is the result of the evolution of a nonlinear, dissipative system where there is no free energy functional. The links and knots are compact attracting states of the reaction-diffusion equation (or more specifically the Markov chain dynamics in our stochastic model). In this sense the factors governing the evolution and forms of the patterns are quite different from other cases where potential energy functions have been associated to knots to determine their geometry²⁴ or applications that use the non-dissipative Euler equations of fluid mechanics²⁵. As discussed above, the stable structures arise from the fact that one is in a parameter regime where the tubular domains containing one of the phases shrink, but the geometry of the tubular domain in conjunction with the repulsion between fronts prevents collapse and leads to a stable pattern.

^bWe use the value 2π for the Hopf link.

The structures we have described are also quite distinct from the linked and knotted vortex filaments seen in excitable and oscillatory media, even for the same FitzHugh-Nagumo model explored here.^{26,27} The excitable regime corresponds to the case where the system is monostable. The stable fixed point is susceptible to finite amplitude perturbations which lead to large excursions in phase space before return to the fixed point. The singular filaments are the cores of three-dimensional spiral waves. Since our tubular domains are composed of one of the two bistable states and have no phase field associated with them that can give rise to twist in the filaments, their structure is actually much simpler than that of vortex filaments.

The fact that the links and knots assume nearly ideal forms seems natural in view of the factors that determine their structure: tubular domain contraction and front repulsion. It would be interesting to explore further the factors that determine the geometries of these stable structures.

Acknowledgments

This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix: System evolution in the Boltzmann approximation

The analysis in Sec. 3 can be made rigorous by analyzing the system in the Boltzmann approximation. In this case the particle distributions at different nodes are considered to be independent so that the state probability distribution is approximated by the product of the reduced 1-node probability distributions:

$$P(\{n_{\mathbf{r}}\}, \{m_{\mathbf{r}}\}, t) = \prod_{\mathbf{r}} p(n, m, \mathbf{r}, t). \quad (19)$$

The evolution of the node probability distribution is described by the Markov process where the transition matrix is the composition of diffusion and reaction operators:

$$p(n, m, \mathbf{r}, t + 1) = \widehat{\mathbf{W}}^C \widehat{\mathbf{W}}^{D_a}(\bar{a}) \widehat{\mathbf{W}}^{D_b}(\bar{b}) p(n, m, \mathbf{r}, t). \quad (20)$$

The operators $\widehat{\mathbf{W}}^{D_a}$ and $\widehat{\mathbf{W}}^{D_b}$ are non-local and depend on the states at the surrounding nodes and have the following form:

$$\begin{aligned} W_{n,n'}^{D_{i=(a,b)}}(c) = & \left[1 - \bar{c} \left(1 - \frac{n'}{N} \right) + (1 - \bar{c}) \frac{n'}{N} \right] \delta_{n',n} \\ & + \bar{c} \left(1 - \frac{n'}{N} \right) \delta_{n',n-1} + (1 - \bar{c}) \frac{n'}{N} \delta_{n',n+1}, \end{aligned} \quad (21)$$

where \bar{c} is the average particle density of the neighbouring sites:

$$\bar{c}(\mathbf{r}, t)N = \frac{1}{2d} \sum_{\mathbf{r}' \in \mathcal{N}(\mathbf{r})} \sum_n np(n, m, \mathbf{r}', t) = \frac{1}{2d} \sum_{\mathbf{r}' \in \mathcal{N}(\mathbf{r})} \langle n \rangle(\mathbf{r}', t). \quad (22)$$

We seek a solution of Eq. (20) in the following form:

$$p(n, m, \mathbf{r}, t) = \sum_{s=0}^{\infty} \varepsilon^s p_s(n, m, a(\mathbf{r}, t), b(\mathbf{r}, t)), \quad (23)$$

with an additional constraint: $\langle \{n, m\} \rangle_p = N \{a(\mathbf{r}, t), b(\mathbf{r}, t)\}$ for any partial sum. We expand the evolution operator in a series of powers of the ordering parameter:

$$\partial_t = \sum_{s=0}^{\infty} \varepsilon^s \mathcal{D}_s. \quad (24)$$

and rewrite the diffusion and collision operators as:

$$\begin{aligned} \widehat{\mathbf{W}}^{D_a}(\bar{c}(\mathbf{r}, t)) &= \widehat{\mathbf{W}}^{D_a}(a(\mathbf{r}, t)) + \varepsilon \left[\widehat{\mathbf{W}}^{D_a}(\bar{c}(\mathbf{r}, t)) - \widehat{\mathbf{W}}^{D_a}(a(\mathbf{r}, t)) \right] \\ \widehat{\mathbf{W}}^C &= \mathbf{1} + \varepsilon \widehat{\mathbf{C}}. \end{aligned} \quad (25)$$

Following the standard Chapman-Enskog procedure²⁸ we arrive at the following equations at the zeroth and first orders:

$$\begin{aligned} \widehat{\mathbf{W}}^{D_a} \widehat{\mathbf{W}}^{D_b} p_0(\cdot, \cdot, \mathbf{r}, t) &= p_0(\cdot, \cdot, \mathbf{r}, t) \\ \mathcal{D}_0 p_0(\cdot, \cdot, \mathbf{r}, t) &= \left[\widehat{\mathbf{D}}^b + \widehat{\mathbf{D}}^b + \widehat{\mathbf{C}} \right] p_0(\cdot, \cdot, \mathbf{r}, t) + \widehat{\mathbf{W}}^{D_a} \widehat{\mathbf{W}}^{D_b} p_1(\cdot, \cdot, \mathbf{r}, t) \end{aligned} \quad (27)$$

where we denoted the difference in Eq. (25) by $\widehat{\mathbf{D}}^{i=(a,b)}$ and used the invariance of $p_0(\mathbf{r}, t)$ under the action of the commuting operators $\widehat{\mathbf{W}}^{D_a}$ and $\widehat{\mathbf{W}}^{D_b}$.

The stationary probability distribution of (27) is binomial. To show this it is convenient to consider the generating function¹⁷ of the distribution

$$f_s(x, y, \mathbf{r}, t) = \sum_{n,m=0}^N x^n y^m p_s(n, m, a(\mathbf{r}, t), b(\mathbf{r}, t)). \quad (29)$$

In the generating function representation the diffusion operator takes the form:

$$\widehat{\mathbf{W}}^{D_a}(a) = [1 - a(1-x)] \left(1 + \frac{1-x}{N} \partial_x \right) \quad (30)$$

and the particle number expectation value is given by:

$$\langle \{n, m\} \rangle_{p_s} = \left\{ \partial_x f_s(x, y), \partial_y f(x, y) \right\} \Big|_{x, y=1}. \quad (31)$$

From the above equation we obtain the evolution equation for the particle number expectation under the action of operator $\widehat{\mathbf{W}}^{D_a}(a)$:

$$\langle n \rangle_{\widehat{\mathbf{W}}^{D_a}(a)p} = a f(1, 1) + \frac{N-1}{N} \partial_x f_s(x, y) \Big|_{x, y=1} = a + \frac{N-1}{N} \langle n \rangle_p, \quad (32)$$

and, at equilibrium:

$$\langle n \rangle_{\widehat{\mathbf{D}}^a(a)p_0(\mathbf{r}, t)} = \bar{c}(\mathbf{r}, t) - \frac{1}{N} \langle n \rangle_{p_0(\mathbf{r}, t)} = \frac{1}{2dN} \Delta \langle n \rangle_{p_0(\mathbf{r}, t)} + o(\nabla^3) \quad (33)$$

The eigenvalues of the operator (30) are $\{1 - k/N : k = 0, \dots, N\}$ and the corresponding eigenfunctions are:

$$\phi_k(x) = (1 - a(1-x))^{N-k} (1-x)^k, \quad (34)$$

so that the zeroth order approximation is given by a local binomial distribution.

The collision operator in the generating function representation has the following form:

$$\begin{aligned} \widehat{\mathbf{C}} &= \tilde{k}_1(1-x)x^{N+1} \partial_x \frac{1}{x^{N-2}} \partial_{xx} + \tilde{k}_1^*(1-x)x^N \partial_{xx} \frac{1}{x^{N-2}} \partial_x \\ &+ \tilde{k}_2(1-x)x^{N+1} y \partial_{xy} \frac{1}{x^N} - \tilde{k}_2^*(1-x)y^{N+1} \partial_{xy} \frac{1}{y^N} \\ &+ \tilde{k}_3^*(1-y)x \partial_{xy} - \tilde{k}_3(1-y)(xy)^{N+1} \partial_{xy} \frac{1}{y^N x^N}, \end{aligned} \quad (35)$$

where \tilde{k}_i and \tilde{k}_i^* are the reaction rates. Acting with the operators in (31) on equation (28) and approximating the spatial average by the Laplace operator we arrive at the FitzHugh-Nagumo system of equations (1). The reaction rates \tilde{k}_i and \tilde{k}_i^* are proportional to the probabilities of transitions p_j of Eq. (13). Computations yield the following relations between the mass action and the lattice gas kinetics coefficients:

$$k_1 = \frac{N!}{N^3(N-3)!} \tilde{k}_1 \text{ and } k_{2,3} = \tilde{k}_{2,3}. \quad (36)$$

References

1. R. Kapral and K. Showalter, editors. *Chemical Waves and Patterns*. Kluwer, Dordrecht, 1995.
2. A. Malevanets and R. Kapral. Microscopic model for FitzHugh-Nagumo dynamics. *Phys. Rev. E*, 55:5657–5670, 1997.
3. A. Malevanets and R. Kapral. Links, knots, and knotted labyrinths in bistable systems. *Phys. Rev. Lett.*, 77:767–771, 1996.
4. V. Katritch, J. Bednar, D. Michoud, R.G. Scharein, J. Dubochet, and A. Stasiak. Geometry and physics of knots. *Nature*, 384:142–145, 1996.
5. R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1:445–466, 1961.
6. J. S. Nagumo, S. Arimoto, and Y. Yoshikawa. *Proc. IRE*, 50(2061), 1962.
7. P. C. Hohenberg and B. I. Halperin. Theory of dynamical critical phenomena. *Rev. Mod. Phys.*, 49(435), 1977.
8. J. D. Gunton and M. Droz. *Introduction to the theory of metastable and unstable states*, vol. 183 of *Lecture notes in physics*. Springer-Verlag, New York, 1983.
9. T. Ohta and R. Mimura, M. Kobayashi. Higher-dimensional localized patterns in excitable media. *Physica D*, 34:115–144, 1989.
10. T. Ohta, A. Ito, and A. Tetsuka. Self-organization in an excitable reaction-diffusion systems. *Phys. Rev. A*, 42:3225–3233, 1990.
11. D. M. Petrich and R. E. Goldstein, Nonlocal contour dynamics model for chemical front motion. *Phys. Rev. Lett.* 72: 1120–1123, 1994.
12. R. E. Goldstein, D. J. Muraki and D. M. Petrich, Interface proliferation and the growth of labyrinths in a reaction-diffusion system. *Phys. Rev. E*, 53: 3933–3957, 1996.
13. A. Hagberg and E. Meron. Domain walls in nonequilibrium systems and emergence of persistent patterns. *Phys. Rev. E.*, 49:5046–5057, 1994.
14. A. Hagberg and E. Meron. From labyrinthine patterns to spiral turbulence. *Phys. Rev. Lett.*, 72:2494–2497, 1994.
15. K. A. Gorshkov, A. S. Lomov, and M. I. Rabinovich. 3-dimensional particle-like solutions of coupled nonlinear fields. *Phys. Lett. A*, 137:250–254, 1989.
16. I. S. Aranson, K. A. Gorshkov, A. S. Lomov, and M. I. Rabinovich. Stable particle-like solutions of multidimensional nonlinear fields. *Physica D*, 43:435–453, 1990.
17. C. W. Gardiner. *Handbook of Stochastic Methods*, vol. 13 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 2 edition, 1985.

18. U. Frisch, D. d'Humieres, B. Hasslacher, P. Lallemand, Y. Pomeau, and J.-P. Rivet. Lattice gas hydrodynamics in two and three dimensions. *Complex Systems*, 1:649–707, 1987.
19. N. Margolus. CAM-8: A computer architecture based on cellular automata. In R. Kapral and A. Lawniczak, editors, *Pattern Formation and Lattice-Gas Automata*, p. 165. Fields Institute Communications, AMS, 1996.
20. B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov. *Modern geometry—methods and applications*, vol. 93 of *Graduate texts in mathematics*. Springer-Verlag, New York, 2nd edition, 1992.
21. K. J. Lee and H. L. Swinney. Lamellar structures and self-replicating spots in a reaction-diffusion system. *Phys. Rev. E*, 51:1899–1915, 1995.
22. K. J. Lee, W. D. McCormick, J. E. Pearson, and H. L. Swinney. Experimental observation of self-replicating spots in a reaction-diffusion system. *Nature*, 369:215–218, 1994.
23. K. J. Lee, W. D. McCormick, Q. Ouyang, and H. L. Swinney. Pattern-formation by interacting chemical fronts. *Science*, 261:192–194, 1993.
24. G. Buck and J. Simon. Knots as dynamical-systems. *Topology and its applications*, 51:229–246, 1993.
25. H. K. Moffat. The energy spectrum of knots and links. *Nature*, 347:367–369, 1990.
26. A. T. Winfree. Stable particle-like solutions to the nonlinear wave equations of the three-dimensional excitable media. *SIAM Review*, 32:1–53, 1990.
27. M. Courtemanche, W. Skaggs, and A. T. Winfree. Stable 3-dimensional action-potential circulation in the FitzHugh-Nagumo model. *Physica D*, 41:173–182, 1990.
28. S. Chapman and T. G. Cowling. *Mathematical Theory of Non-Uniform Gases*. Cambridge University Press, Cambridge, 1952.

CHAPTER 14

NEW DEVELOPMENTS IN TOPOLOGICAL FLUID MECHANICS: FROM KELVIN'S VORTEX KNOTS TO MAGNETIC KNOTS

RENZO L. RICCA

*Department of Mathematics, University College London,
Gower Street, London WC1E 6BT, UK*

and

*ISIS, TP 723, EC Joint Research Centre,
21020 Ispra (VA), Italy*

E-mail: ricca@math.ucl.ac.uk

In this paper we review classical and new results in topological fluid mechanics based on applications of first principles of ideal fluid mechanics and knot theory to vortex and magnetic knots. After some brief historical remarks on the first original contributions to topological fluid mechanics, we review basic concepts of topological fluid mechanics and local actions of fluid flows. We review some classical, but little known, results of J.J. Thomson on vortex links, and discuss Kelvin's conjecture on vortex knots. In the context of the localized induction approximation for vortex motion, we present new results on existence and stability of vortex filaments in the shape of torus knots. We also discuss new results on inflexional magnetic knots and possible relaxation to minimal braids. These results have potentially important applications in disciplines such as astrophysics and fusion plasma physics.

1 Kelvin's vortex atoms and the origin of topological fluid mechanics

The use of topological ideas in fluid mechanics dates from the original studies of Gauss^{7,5} on linked orbits and electric circuits, from Lord Kelvin's¹³⁻¹⁴ first investigations on vortex knots, and from Maxwell's¹⁹ thoughts on magnetic flux tubes (see the table in Figure 1 below).

Gauss's work was followed by the studies of Listing on topological properties of surfaces (among which the famous one-sided band, wrongly attributed to Möbius), and by the work of Riemann on analytic properties of irrotational flows embedded in multiply connected regions (with applications to fluid flows in presence of holes).

But it was Kelvin (then W. Thompson), who gave the greatest impetus to applications of topological ideas to physics. His work was inspired by Helmholtz's⁹ influential paper on vortex motion, and was motivated by the search for a fundamental theory of matter. Kelvin's theory assumed the existence of a dynamical fluid ether permeating everything, in which natural forces were generated. Kelvin's realization that vortex filaments in inviscid fluid were

| | |
|--|---|
| Linking number formula (C.F. Gauss, 1833) | Applications to vector fields (J.C. Maxwell, 1873) |
| Classification of knots (P.G. Tait, 1867) | Applications to vortices (Lord Kelvin, 1867) |

Figure 1: First contributions to the origin of topological fluid mechanics.

permanent, dynamical entities, led him to envisage a vortex atom theory¹³, whose fundamental constituents, the atoms, were knotted vortices embedded in the ether. In absence of dissipation the topology of vortex structures, given for example by the linking of two vortices, is frozen in the fluid, so that knotted and linked structures remain knotted and linked indefinitely. For Kelvin the topological specificity of each knot and link type provided a useful paradigm to represent chemical elements and compounds. By interpreting knotted vortices as elemental building blocks, and links as compounds, it was possible to envisage chemical structures ordered in a way similar to the modern periodic table of elements. The puzzle of quantization of energy, revealed by the spectral studies of light, could thus find a simple and natural explanation in terms of the discrete specificity of the knot types. These ideas became part of a topological theory of matter³² *ante-litteram*. The mathematical study of knots and links thus became an integral part of Kelvin's programme, and this study was carried out by his friend and collaborator Tait. The results of Tait's work,³⁵ that included the first classification tables of knots, were published in a series of three remarkable papers destined to become the foundations of modern knot theory.

Other contributions followed soon. Most notably the work of J.J. Thomson³⁶ on vortex links (see Section 3 below) and the studies of fluid flows in multiply connected domains (see, for example, Lamb's *Hydrodynamics*¹⁶). These ideas survived for some time. In Lichtenstein's mathematical theory of hydrodynamics,¹⁸ for example, the importance of topology is emphatically stressed by two chapters dedicated to the subject.

While Kelvin's dream of explaining atoms as knotted vortices in a fluid ether never came to fruition, his work was seminal in the development of topological fluid mechanics. The recent revival is mainly due to the work of Moffatt,²⁰ on topological interpretation of helicity, and Arnold,¹ on asymptotic linking number of space-filling curves. Modern developments have been influ-

enced by the recent progress in the theory of knots and links, and by access to fast computer and sophisticated numerical diagnostics. Various research areas now benefit from use of topological techniques and fluid mechanics (an overview of the present state of the art is given, for example, by the article of Ricca & Berger³⁰).

We list here some current research topics relevant to topological fluid mechanics:²²

- *Knotted and linked solutions to Euler's equations:*
 - topological classification of fluid flows;
 - relationships between topology and dynamics;
 - rôle of invariants and integrability;
 - relationships between topology and stability properties.
- *Energy relaxation for topologically complex structures:*
 - magnetic knots and braids;
 - electrically charged links;
 - relationships between topology and energy;
 - energy spectra for physical knots and links.
- *Dynamical systems and measure-preserving flows:*
 - existence theorems for 3-D vector fields;
 - topologically complex closed and chaotic orbits;
 - relationships between topology and Hamiltonian flows;
 - Lie-algebras of invariants.
- *Change of topology and complexity measures:*
 - singularity formation;
 - bifurcation theory and classification of singularities;
 - physical reconnection mechanisms;
 - measures of topological complexity and diagnostics.

2 Basic concepts in topological fluid mechanics

2.1 Topological equivalence classes for frozen fields

We consider an ideal and perfectly conducting fluid in an infinite domain \mathcal{D} of \mathbb{R}^3 . Motion of fluid particles is given by a smooth velocity field $\mathbf{u} = \mathbf{u}(\mathbf{X}, t)$, where \mathbf{X} denotes the position vector and t time. The velocity field satisfies the solenoidal condition in \mathcal{D} and the condition to be at rest at infinity:

$$\nabla \cdot \mathbf{u} = 0 \quad , \quad \text{in } \mathcal{D} \quad , \quad (1)$$

$$\mathbf{u} = 0 \quad , \quad \text{as } \mathbf{X} \rightarrow \infty \quad . \quad (2)$$

Fluid particles move in \mathcal{D} from one position to another. If $\mathbf{a} = \mathbf{X}(\mathbf{a}, 0)$ denotes the initial position of a fluid particle at time $t = 0$, then we have a flow map φ_t induced by \mathbf{u} so that each particle at the initial position \mathbf{a} and time $t = 0$ is sent to the final position $\mathbf{X}(\mathbf{a}, t)$ by

$$\varphi_t : \mathbf{a} \rightarrow \mathbf{X} \quad , \quad \forall t \in I \quad , \quad (3)$$

where I denotes some finite time interval. The flow map φ is continuous, one-to-one and onto, with inverse. For an incompressible fluid the flow map is volume preserving, with Jacobian

$$J = \det \left(\frac{\partial X_i}{\partial a_j} \right) = 1 \quad . \quad (4)$$

Let $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathbf{X}, t)$ be a solenoidal ($\nabla \cdot \boldsymbol{\Omega} = 0$) vector field in the fluid domain \mathcal{D} . Then, the evolution of the vector field $\boldsymbol{\Omega}$ is governed by the following master equation:

$$\frac{\partial \boldsymbol{\Omega}}{\partial t} = \nabla \times (\mathbf{u} \times \boldsymbol{\Omega}) \quad . \quad (5)$$

If $\boldsymbol{\Omega}$ is the vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{u}$, eq. (5) is the Helmholtz equation for the transport of vorticity in ideal fluids (Euler's equations). Alternatively, if $\boldsymbol{\Omega}$ is the magnetic field \mathbf{B} , then eq. (5) governs the evolution of \mathbf{B} in ideal magnetohydrodynamics (MHD). Equation (5) admits formal integral solutions called Cauchy equations, given by

$$\boldsymbol{\Omega}(\mathbf{X}, t) = \boldsymbol{\Omega}(\mathbf{a}, 0) \cdot \frac{\partial}{\partial \mathbf{a}} \mathbf{X} = \Omega_j \frac{\partial X_i}{\partial a_j} \quad , \quad (6)$$

that conserve topology. This means that while the field geometry changes smoothly from one configuration to another by continuous actions of flow maps,

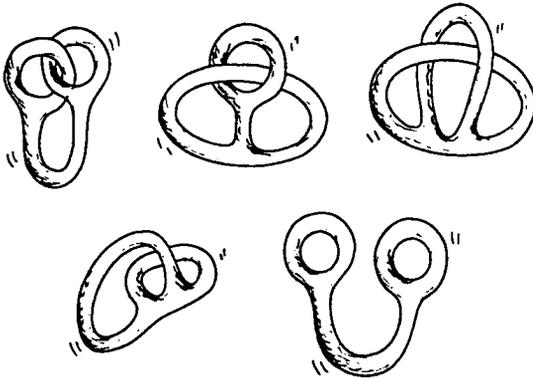


Figure 2: Topologically equivalent configurations of a fluid pretzel.

the initial field topology is conserved. Equation (6) encapsulates both the convection of the field from the initial position \mathbf{a} to the final position \mathbf{X} , and the simultaneous rotation and distortion of fluid elements by the deformation tensor $\partial X_i / \partial a_j$. Since this tensor is a time-dependent diffeomorphism of positions, it maps continuously the initial field distribution $\Omega(\mathbf{a}, 0)$ to $\Omega(\mathbf{X}, t)$ by establishing a topological equivalence between the two fields. Hence, we write

$$\Omega(\mathbf{a}, 0) \sim \Omega(\mathbf{X}, t). \quad (7)$$

Under these conditions the field Ω is said to be 'frozen' in the fluid and initial and final configurations are said to be isotopic to each other.

Continuous deformations of fluid structures are often complicated by twisting and folding actions of fluid flows. The five configurations of a fluid pretzel shown in Figure 2 provide a striking example of equivalent isotopies of a fluid structure by (non-trivial) flow maps.

2.2 Action of local flows and Reidemeister's moves

Ideal topological fluid mechanics deals essentially with the study of fluid structures that are continuously deformed from one configuration to another by ambient isotopies. Since the fluid flow map φ is both continuous and invertible, then $\varphi_{t_1}(\mathcal{K})$ and $\varphi_{t_2}(\mathcal{K})$ generate isotopies of a fluid structure \mathcal{K} (for example a vortex filament) for any $\{t_1, t_2\} \in I$. Isotopic flows generate equivalence classes of (linked and knotted) fluid structures. In the case of (vortex

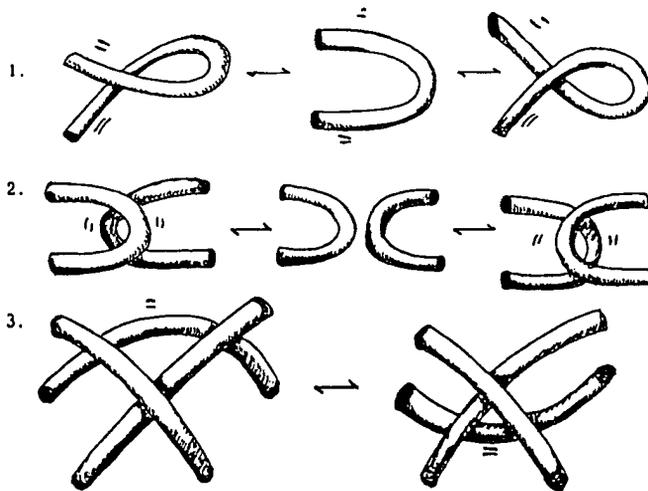


Figure 3: The three types of Reidemeister's moves can be performed by natural actions of local fluid flows on fluid flux tube strands.

or magnetic) fluid flux tubes, fluid actions induce continuous deformations in \mathcal{D} . One of the simplest deformations is local stretching of the tube. From a mathematical viewpoint this deformation corresponds to a time-dependent, continuous re-parametrization of the tube centreline. This re-parametrization (via homotopy classes) generates ambient isotopies of the flux tube, with a continuous deformation of the integral curves.

It is well known (see, for example, Kauffman¹¹) that knot topology is conserved under the action of the Reidemeister moves (see Figure 3). In the context of the Euler equations these moves are performed quite naturally by the action of local flows on flux tube strands. If the fluid in $(\mathcal{D} - \mathcal{K})$ is irrotational, then these fluid flows (with velocity \mathbf{u}) must satisfy the Dirichlet problem for the Laplacian of the stream function ψ , that is

$$\left. \begin{aligned} \mathbf{u} &= \nabla\psi \\ \nabla^2\psi &= 0 \end{aligned} \right\} \text{ in } (\mathcal{D} - \mathcal{K}), \quad (8)$$

with normal component of the velocity to the tube boundary \mathbf{u}_\perp given. Equations (8) admit a unique solution in terms of local flows,² and these flows are interpretable in terms of Reidemeister's moves performed on the tube strands. Note that boundary conditions prescribe only \mathbf{u}_\perp , whereas no condition is im-

posed on the tangential component of the velocity. This is consistent with the fact that tangential effects do not alter the topology of the physical knot (or link). The three types of Reidemeister's moves are therefore performed by local fluid flows, which are solutions to (8), up to arbitrary tangential actions.

2.3 *Ideal versus 'real' topological fluid mechanics*

Relationships between topology and dynamics of fluid structures are very little explored. Questions about topology and dynamics or stability of linked vortices, topology and energy levels of magnetic loops, or energy relaxation of knotted magnetic fields, are still very little studied.

In ideal conditions (i.e. in absence of dissipative effects) all topological properties and physical quantities are conserved. These form a set of scalar and vector invariants that guide the evolution of the system towards (homotopic) solutions, whose existence is guaranteed by the diffeomorphisms associated with the flow maps. Changes in the topology of the system occur only if singularities, bifurcations and dissipative effects are present. Clearly, if we want to model 'real' flow maps, then we cannot neglect the presence of wakes, boundary layers and other fluid regions, where dissipative effects are indeed relevant. Similarly, we cannot neglect the presence of physical regions, where particle trajectories have wild behaviours, with bifurcations, multiple points and singularities. In neglecting the presence of these physical regions we are in fact limiting the validity of the models. Results obtained by techniques of ideal topological fluid mechanics (where dissipative effects are ignored) should therefore be preliminary to the study of real flows, but then complemented or adjusted by 'real' fluid mechanics.

3 Links of thin core vortex rings

The first mathematical study of dynamical aspects of linked vortex rings was done by J.J. Thomson³⁶ (see also the paper by Ricca & Weber³²). His work was inspired by Kelvin's vortex atom theory, and provides one of the most remarkable examples of combination of topological ideas and fluid mechanics. Thomson's idea was to study vortex structures, linked and knotted together, by using thin core models of vortex rings and basic notions of linking, based on Gauss's formula of linking number.⁷ Thomson tackled the problem by considering a particular geometry given by two linked vortex rings lying on the mathematical surface of a torus Π of radius R and small diameter d . The simplest example of this kind of link was given by two inter-linked rings, C_1 and C_2 , embedded on Π as shown in Figure 4.

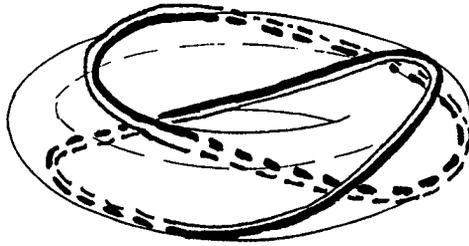


Figure 4: Simple link of two rings C_1 and C_2 , lying on the mathematical torus Π : the number of components is $n = 2$, and the linking number of the system is $Lk = 1$.

In Thomson's view the vortex link geometry is the result of the uniform, rigid rotation of two point vortices (representing the cross-sections of two vortex filaments) around their center of mass (in the meridian plane of Π), making one complete rotation around the great circle (of radius R) in the longitudinal direction. The vortex filaments were given by the collection of the point vortex positions. The resulting vortex system, made of these thin, closed vortex filaments embedded in ideal fluid, is frozen. Hence, the dynamics of the system is expected to be influenced by the type of linking.

Let $\lambda = \max |X_i - X_i^*|$, for points $\{X_i, X_i^*\} \in C_i$, $i = 1, 2$, and $\delta = \min |X_1 - X_2|$ for points $X_1 \in C_1$ and $X_2 \in C_2$. If we assume that $\lambda \gg \delta$, where $\lambda = O(2R)$ and $\delta = O(d)$, then we can show that:

Theorem (Thomson, 1883). *Consider the link formed by two vortex rings of equal circulation Φ and relative linking number Lk , embedded and equally spaced on a torus Π in \mathcal{D} . The vortex system is steady and stable iff*

$$\frac{M(2\pi\rho\Phi)^{1/2}}{LkP^{3/2}} < 1, \quad (9)$$

where ρ is the fluid density (constant) and $M = |M|$ and $P = |P|$ are the intensities of the angular momentum M , and the linear momentum P of the system.

The simplest link system (with $Lk = 1$) rotates and translates as a rigid body, with angular velocity Ω and translational velocity V given by

$$\Omega = \frac{\Phi}{\pi a^2}, \quad V = \frac{\Phi}{4\pi R} \log \frac{64R^2}{a^2}. \quad (10)$$

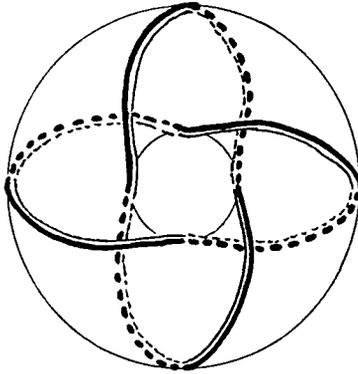


Figure 5: Example of higher-order link made of two rings: in this case $n=2$, $Lk = 2$.

Higher-order link systems (i.e. two-component link with $Lk > 1$) are realized by a higher number of full rotations of the point vortex system around the great circle (see Figure 5).

Consider now n vortex rings linked together. Following a similar construction, the n -component link is now given by n inter-linked vortices on Π (see Figure 6). The system is generated by the rigid rotation of n point vortices equally spaced on the torus small circumference. After a long and laborious analysis Thomson finds the following result:

Theorem (Thomson, 1883). *Consider the link of n vortex rings of equal circulation Φ and relative linking number Lk , embedded and equally spaced on a torus Π in \mathcal{D} . The vortex system is steady and stable iff $n \leq 6$, with period of vibration*

$$T = \frac{2\pi}{\Phi \left(\frac{2}{d^2} - \frac{(2Lk^2 - 1)}{4a^2} \log \frac{d}{a} \right)}. \quad (11)$$

This result has been confirmed by later works in the theory of point vortex motion in the plane.³³

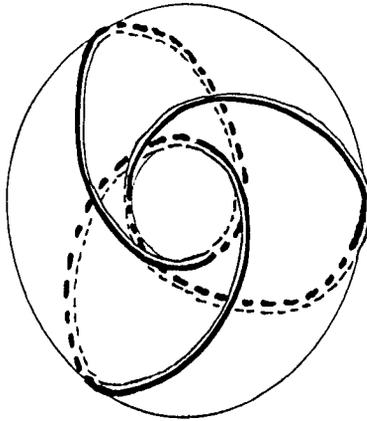


Figure 6: Example of higher-order link given by three rings on the torus Π : $n=3$ and relative linking $Lk = 1$.

4 Evolution and stability of thin vortex knots

4.1 Kelvin's conjecture and vortex knot dynamics

In his vortex atom theory, Kelvin¹⁴ conjectured (see *Vortex Statics*, p. 123, ¶ 16) that thin vortices in the shape of torus knots could exist as steady and stable fluid structures. Vortex knot solutions (to Euler's equations), if existed, could move with constant speed in the fluid and, if disturbed, vibrate about their equilibrium configuration. From a mathematical viewpoint the search for the existence of vortex knots remained open, and only in recent years there has been a real progress in this study. Here we want to present and discuss some new results.

Thin vortex knots have been found as solutions to the so-called 'localized induction approximation'²⁷ (LIA for short). This is an approximation of the Biot-Savart law for Euler's equations. Under LIA, vortex motion is governed by a law, that after appropriate re-scaling of the time variable, takes the simple form

$$\text{LIA:} \quad \mathbf{u} = \dot{\mathbf{X}} = \mathbf{X}' \times \mathbf{X}'' = c\hat{\mathbf{b}}, \quad (12)$$

where \mathbf{u} is the vortex velocity, the dot denotes time-derivative and the prime denotes derivative with respect to arc-length along the tube axis. c and $\hat{\mathbf{b}}$ are local curvature and unit binormal to the axis. It is interesting to note that since eq. (12) is equivalent to the non-linear Schrödinger equation,⁸ we

have a countably infinite number of polynomial conserved quantities (soliton invariants) that can be written as global geometric functionals.^{17,23}

Existence and steadiness of knotted solutions to LIA have been studied by Kida¹⁵ and Keener.¹² Kida's solutions are torus knots in the physical space and represent the first vortex knot solutions found by analytical methods. We have:

Theorem (Kida, 1981). *Let \mathcal{K}_v denote the embedding of a knotted vortex filament in an ideal fluid in \mathcal{D} . If \mathcal{K}_v evolves under LIA, then there exists a class of steady solutions in the shape of torus knots $\mathcal{K}_v \equiv \mathcal{T}_{p,q}$.*

In geometric terms Kida's solutions are closed curves embedded on the mathematical torus, wrapping the torus $p > 1$ times in the longitudinal direction and $q > 1$ times in the meridian direction (p, q co-prime integers). The winding number is given by $w = q/p$, and self-linking given by $Lk = pq$, two topological invariants of the knot type. Kida gives the solutions in terms of fully non-linear relationships that involve elliptic functions of traveling waves. A more direct and simpler approach has been proposed by Ricca,²⁵ and is based on linear perturbation techniques and cylindrical polar coordinates (r, α, z) . By this approach we find 'small-amplitude' torus knot solutions (asymptotically equivalent to Kida's solutions) given by

$$\left. \begin{aligned} r &= r_0 + \epsilon_0 k_r \sin(\omega\phi + \phi_0) \\ \alpha &= \frac{s}{r_0} + \epsilon_0 \frac{k_r}{\omega r_0} \cos(\omega\phi + \phi_0) \\ z &= \frac{\hat{t}}{r_0} + \epsilon_0 k_r \left(1 - \frac{1}{w^2}\right)^{1/2} \cos(\omega\phi + \phi_0) . \end{aligned} \right\} \quad (13)$$

r_0 is the radius of the torus circular axis and $\epsilon_0 = a/r_0 \ll 1$ is the inverse of the aspect ratio of the vortex, with a the radius of the vortex cross-section and $k_r = O(r_0)$ a scale factor. Moreover $\phi = (s - \kappa t)/r_0$, with t time, \hat{t} a time re-scaled with the vortex circulation, and ϕ_0 a constant.

4.2 New results on stability of vortex knots

Since torus knots have two isotopes $\mathcal{T}_{p,q}$ and $\mathcal{T}_{q,p}$ (for given p and q), that are topologically equivalent but geometrically different, the question of their evolution and stability is particularly interesting. A linear stability analysis^{25,26} based on equations (13) leads to the following result:

Theorem (Ricca, 1993; 1995). *Let $\mathcal{T}_{p,q}$ denote the embedding of a 'small-amplitude' vortex torus knot \mathcal{K}_v evolving under LIA. $\mathcal{T}_{p,q}$ is steady and stable*

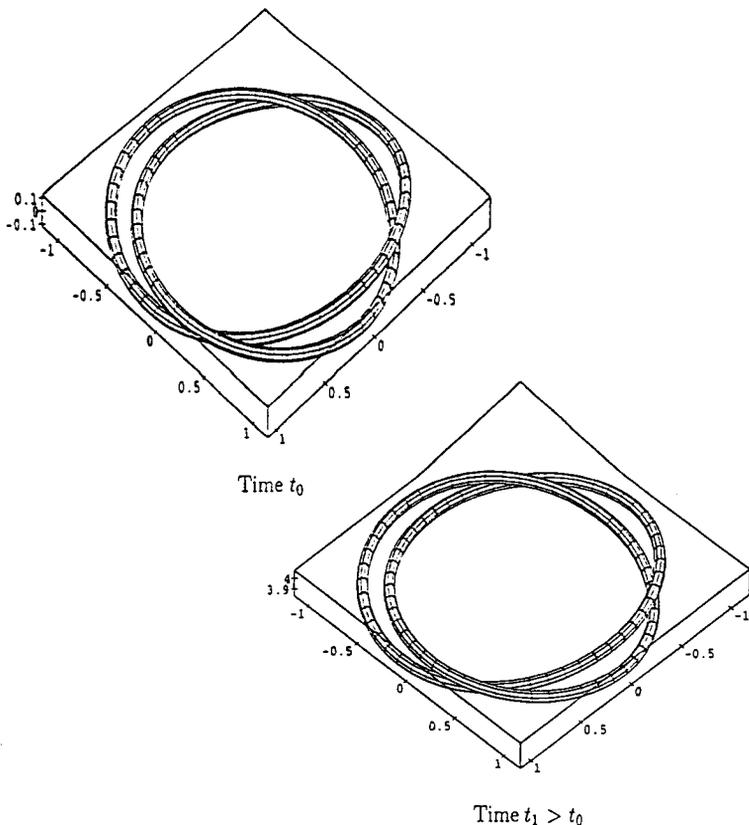


Figure 7: Evolution of torus knot $\mathcal{T}_{2,3}$ under LIA. The knot is found to be stable as predicted by the LIA analysis of Ricca. The knot is visualized by centering a thin tube on the knot axis. The tube shown is therefore a virtual object and its thickness is not measured by a_0 .

under linear perturbations iff $q > p$ ($w > 1$).

Kelvin's conjecture can be therefore tested using this criterium. Numerical calculations^{34,31} have been performed to check and investigate the validity of the above result, and confirm that knots, with winding number $w > 1$ are indeed stable under LIA evolution.

Figure 7 shows two snapshots of the stable knot $\mathcal{T}_{2,3}$ and Figure 8 shows the knot $\mathcal{T}_{3,2}$ as it becomes unstable and unfolds. Another interesting result³¹ is the discovery of a strong stabilizing effect due to the full Biot-Savart law. Take for example the knot $\mathcal{T}_{3,2}$: this knot becomes immediately unstable under LIA, whereas it remains stable under Biot-Savart, travelling a considerable distance. Although we find that these knots eventually de-stabilize (remember

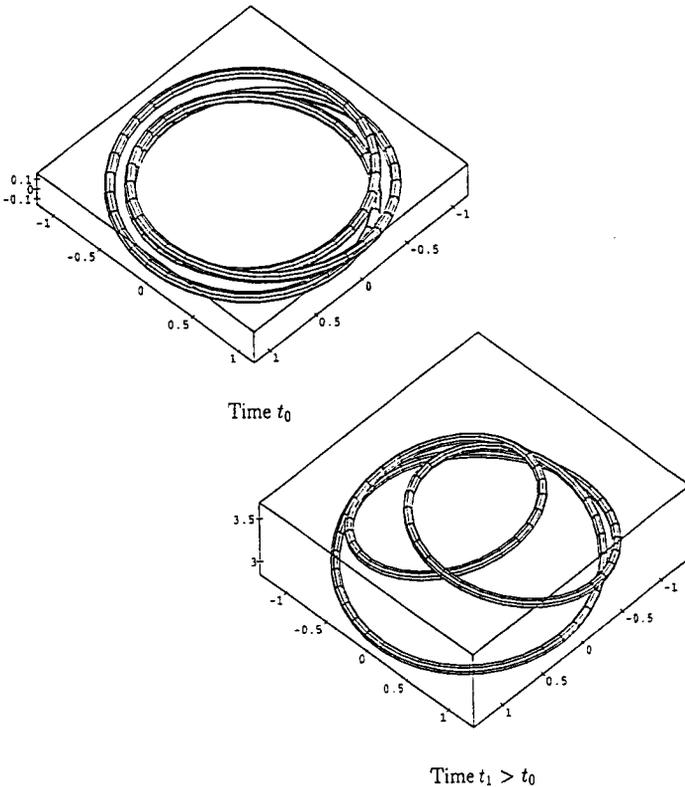


Figure 8: Evolution of torus knot $T_{3,2}$ under LIA. The knot is found to be unstable as predicted by the LIA analysis of Ricca. The knot is however stabilized when its evolution is governed by the Biot-Savart law.

that some numerical noise is always present), the time which elapses and the distance over which the knot travels before breaking-up is very large and has physical significance.

Finally, let us point out that unstable vortex knots evolve under LIA towards a reconnection event. This is another interesting feature of vortex knot evolution, especially in view of the great interest for the study of singularity formation. No doubt that these results will stimulate more numerical work and will certainly give new impetus to the mathematical search for the existence of steady and stable vortex knot solutions under the Euler equations.

5 Magnetic knots, minimal braids and energy estimates

5.1 Evolution of inflexional magnetic knots

Magnetic knots are the physical analog of vortex knots, when we replace vorticity with the magnetic field (so that the physical knot becomes a magnetic flux tube). Magnetic knots evolve according to a dynamics given by the Lorentz force. For a solenoidal magnetic field $\mathbf{B} = \mathbf{B}_m + \mathbf{B}_a$ given by a meridian (i.e. poloidal) component \mathbf{B}_m and an axial (i.e. toroidal) component \mathbf{B}_a , where

$$\mathbf{B}_m = [0, B_\theta(r, \vartheta(s)), 0], \quad \mathbf{B}_a = [0, 0, B_s(r)], \quad (14)$$

with B_θ and B_s smooth functions of radius of the tube cross-section r ($0 \leq r \leq a$) and azimuth angle $\vartheta = \vartheta(s)$ ($0 \leq \vartheta \leq 2\pi$; s arc-length on the tube axis), we have²⁸

$$\mathbf{F}_\perp = B_s^2 \frac{c}{K} \hat{\mathbf{n}} - \left[\frac{B_\theta^2}{r} + \frac{1}{2} \frac{\partial}{\partial r} (B_\theta^2 + B_s^2) \right] \hat{\mathbf{e}}_r, \quad (15)$$

where \mathbf{F}_\perp denotes the component of the force perpendicular to the tube axis. Here K is a scale factor (function of the geometry), and $\hat{\mathbf{n}}$ and $\hat{\mathbf{e}}_r$ are two unit vectors in the principal normal direction and in the radial direction to the tube axis. Since \mathbf{F}_\perp is the only component of the force responsible for the motion of the knot in the fluid, in a first approximation we can write $\mathbf{F} \approx c\hat{\mathbf{n}}$, with force proportional to curvature, along the principal normal of the knot axis. This force induces a tension in the physical knot and a shortening of the lines of force.

In general magnetic knots exhibit inflexional configurations. The geometry of these configurations is characterized by a change in concavity in the tube axis, at a point where curvature vanishes (inflexion point). Inflexional states are easily identified in plane curves: in this case the inflexional geometry is simply given by an S -shaped curve with the inflexion point at the change of concavity. Inflexional configurations in magnetic field structures, however, are ubiquitous, especially in rich topologies. Moffatt & Ricca²⁴ showed that the appearance of inflexional states is invariably associated with the continuous exchange of writhe and twist (see Figure 9), a natural mechanism in the evolution of magnetic structures.

The dynamics of magnetic flux tubes in inflexional configuration has been studied by applying the Lorentz force equations to a generic deformation through inflexion. It can be shown²⁸ that inflexional states represents disequilibria for magnetic configurations. In particular we can state the following result.²⁹

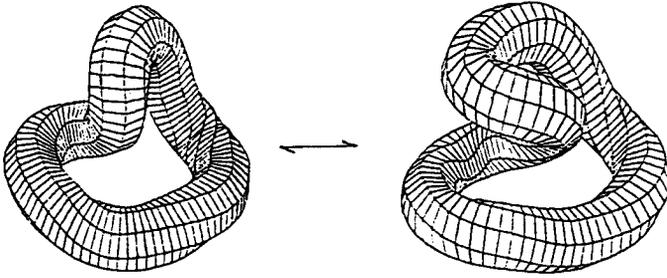


Figure 9: Competition between twist and writhe in a magnetic knot.

Theorem (Ricca, 1997). *Let \mathcal{K}_m be the embedding of a magnetic knot in \mathcal{D} . If \mathcal{K}_m has a finite number of inflexion points in isolation, then \mathcal{K}_m evolves to an inflexion-free configuration, possibly in braid form.*

As we shall see below this result has important consequences for the first stages of the relaxation of magnetic knots.

5.2 Relaxation of inflexional knots to minimal braids

Since inflexional magnetic knots are in disequilibrium, they remove inflexions by re-arranging the geometry to form topologically equivalent configurations free from inflexions. In general the Lorentz force induce a natural tendency to minimize the magnetic tension present in the tube by reducing the surplus of internal magnetic twist (through an increase of writhing), and by removing inflexion points. This favours a deformation to topologically equivalent inflexion-free configurations. In absence of other forces, the evolution is then dominated by curvature forces that induce a continuous, progressive shortening of the field lines (hence of the knot) toward a minimum energy state. In the ideal process, virtual crossings and inflexional states are naturally removed and the knot relaxes isotopically to an inflexion-free configuration, with least possible number of (real) crossings ('minimal closed braid form'; see Figure 10). In general this number of crossings is equal to, or higher than, the topological crossing number.

Minimal braids are particular geometric representations of knot types. From a purely topological viewpoint, any knot can be isotoped to a closed braid by a sequence of Reidemeister's moves (in braid theory this result is

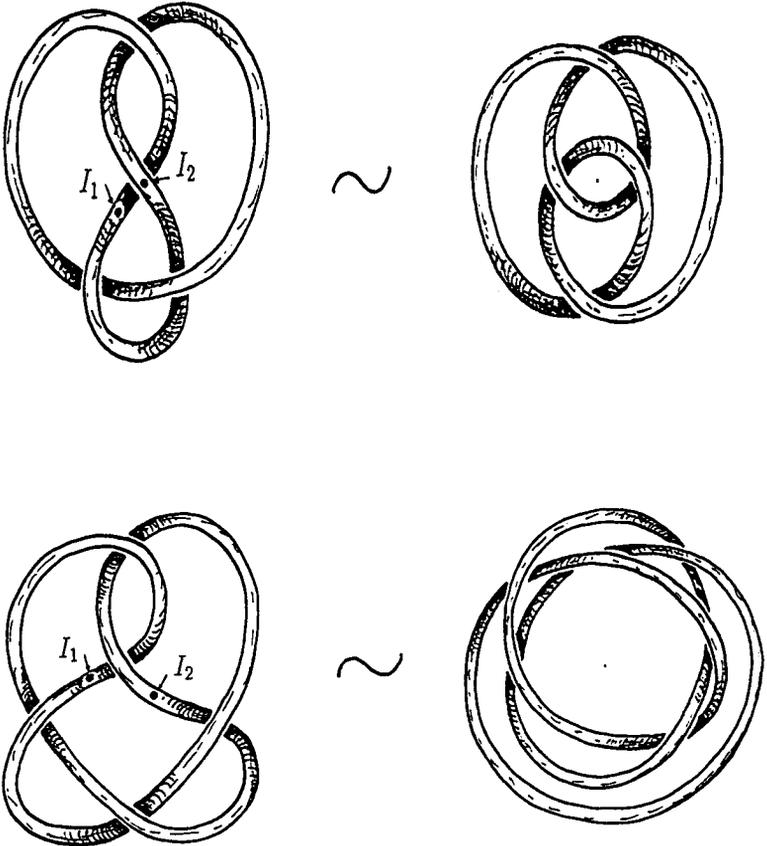


Figure 10: Two different knot types are shown in standard, minimal projection (on the left hand side) and in their topologically equivalent minimal, closed braid form (on the right hand side). Note that the knots in their standard representation exhibit at least two points of inflexion, denoted by I_1 and I_2 in the diagram. These knots can be both isotoped to their minimal closed braid representation, which is given by an inflexion-free configuration with the least possible number of crossings.

known as Markov's theorem⁴). The deformation of an inflexional knot into a closed braid representation (as is given in the 'conjugacy class' of the knot type) is likely to introduce new crossings. To see this, let us consider the two examples of isotopic transformation of knotted loops shown in Figure 10. The diagrams on the left show two different knot types in minimal projection, i.e. in the plane projection for which the number of crossings is the topological crossing number C_{\min} . In this configuration the knots have at least two points of inflexions (denoted by I_1 and I_2 in the diagrams; note that inflexional states are intrinsic geometric features independent of projection angle and viewing direction). The corresponding diagrams on the right, however, show no points of inflexions, and represent the isotopic configurations of the knots in the minimal braid form, with curvature vector pointing always inward the braid region. Note that in passing from the minimal standard to the minimal braid configuration the four crossing knot conserves the minimum number of crossings ($C_{\min} = C_0 = 4$), whereas the five crossing knot (with $C_{\min} = 5$) has the least possible number of crossings $C_0 = 6 > C_{\min}$. In general we have $C_0 \geq C_{\min}$. Our analysis,²⁹ based on standard results of knot theory,^{10,37} shows that indeed there are infinitely many knots (whose simplest representative is the five-crossing knot) that cannot be transformed to minimal braids by 'equi-minimal' isotopies (i.e. by conserving the minimum possible number of crossings). The existence of a family of knots that have 'non-equi-minimal' braid representatives seem to have important consequences for the estimates of energy minima of magnetic knots.

5.3 Possible consequences for energy estimates

Mathematical estimates of minimum energy levels based on topological information have shown^{21,63,29} that minimum energy states of physical knots can be related to topological quantities such as linking number and crossing number. Lower bounds for energy levels of magnetic knots are given by relations of the kind

$$E_{\min} \geq f(\Phi, V, n, C_{\min}), \quad (16)$$

where E_{\min} denotes the ground state energy and $f(\cdot)$ gives the relationship between physical conserved quantities — such as total flux Φ , magnetic volume V , number of tubes n (in the case of an n -component braid or link) — and topology (here given by C_{\min}). Typically, in these relations energy increases with knot complexity. The inequality sign allows ample margins for errors, so that these estimates are still rather qualitative. For inflexional magnetic knots, far from their minimum energy state, our results indicate that ground energy levels may be strongly influenced by the presence of inflexions. Since inflexions

represents disequilibria of thin magnetic knots, inflexional knots will tend to remove inflexions and relax to minimal braids first, before relaxing further to their ground energy state. But minimal braids are likely to have least possible number of crossings $C_0 > C_{\min}$ (as for the five-crossing knot of Figure 10). For this sub-family of minimal braids ('non-equi-minimal' braids) we expect an E_{\min} higher than the theoretical bound given by the equality sign in (16).

This studies find useful applications in applied disciplines, such as astrophysics, solar physics and fusion plasma physics. For solar coronal loops, for example, a difference in the crossing number of the relaxing magnetic braid has important physical consequences for energy estimates, especially when these estimates are based on theoretical models that are very sensitive to variations in geometric and topological information. The accuracy of these estimates is crucial to give precise evaluations of the amount of energy that can be released into heat during flares and microflares. Future progress in this direction will be very important for energy studies.

Acknowledgments

The author would like to acknowledge financial support from The Leverhulme Trust (UK) and from ISIS-JRC, Ispra (EC).

References

1. Arnold, V.I. 1974 In *Proc. Summer School in Differential Equations*. Acad. Sci. Armenian S.S.R., Erevan (in Russian). [(1986) *Sel. Math. Sov.* 5, 4, 327–345].
2. Batchelor, G.K. 1967 *An Introduction to Fluid Dynamics*. Cambridge University Press.
3. Berger, M.A. 1993 *Phys. Rev. Lett.* 70, 705–708.
4. Birman, J.S. 1976 *Braids, links and mapping class group*. Annals of Math. Studies 82, Princeton University Press.
5. Epple, M. 1998 *Math. Intelligencer* 20, 45–52.
6. Freedman, M.H. & He, Z.-X. 1991 *Ann. Math.* 134, 189–229.
7. Gauss, C.F. 1833 In *Works*. Königlichlen Gesellschaft der Wissenschaften zu Göttingen, (in German).
8. Hasimoto, H. 1972 *J. Fluid Mech.* 51, 477–485.
9. Helmholtz, H. 1858 *J. Reine Angew. Math.* 55, 25–55 (in German).
10. Jones, V.F.R. 1987 *Ann. Math.* 126, 335–388.
11. Kauffman, L.H. 1987 *On Knots*. Annals Study 115, Princeton University Press.

12. Keener, J.P. 1990 *J. Fluid Mech.* **211**, 629–651.
13. Kelvin, Lord (Thomson, W.) 1867 *Phil. Mag.* **33**, 511–512.
14. Kelvin, Lord (Thomson, W.) 1875 *Proc. Roy. Soc. Edinb.* ss. 1875–76, 115–128.
15. Kida, S. 1981 *J. Fluid Mech.* **112**, 397–409.
16. Lamb, Sir H. 1879 *Hydrodynamics*. Cambridge University Press.
17. Langer, J. & Perline, R. 1991 *J. Nonlin. Sci.* **1**, 71–93.
18. Lichtenstein, L. 1929 *Grundlagen der Hydromechanik*. Verlag, Berlin.
19. Maxwell, J.C. 1873 *A Treatise on Electricity and Magnetism*. MacMillan & Co., Oxford.
20. Moffatt, H.K. 1969 *J. Fluid Mech.* **35**, 117–129.
21. Moffatt, H.K. 1990 *Nature* **347**, 367–369.
22. Moffatt, H.K. 1992 In *Topological Aspects of the Dynamics of Fluids and Plasmas* (ed. H.K. Moffatt *et al.*). Kluwer, Dordrecht, The Netherlands.
23. Moffatt, H.K. & Ricca, R.L. 1991 In *The Global Geometry of Turbulence* (ed. J. Jiménez). NATO ASI B **268**, Plenum Press, New York.
24. Moffatt, H.K. & Ricca, R.L. 1992 *Proc. R. Soc. Lond.* A **439**, 411–429.
25. Ricca, R.L. 1993 *Chaos* **3**, 83–91 [1995 Erratum. *Chaos* **5**, 346].
26. Ricca, R.L. 1995 In *Small-Scale Structures in Three-Dimensional Hydro and Magnetohydrodynamics Turbulence* (ed. M. Meneguzzi *et al.*). Lecture notes in Physics **462**. Springer-Verlag.
27. Ricca, R.L. 1996 *Fluid Dyn. Res.* **18**, 245–268.
28. Ricca, R.L. 1997 *Solar Physics* **172**, 241–248.
29. Ricca, R.L. 1998 In *Knot Theory* (ed. V.F.R. Jones *et al.*). Institute of Mathematics **42**, Polish Academy of Sciences, Warszawa.
30. Ricca, R.L. & Berger, M.A. 1996 *Phys. Today* **49**, 24–30.
31. Ricca, R.L., Samuels D.C. & Barenghi C.F. 1998 In *Proc. VII European Turbulence Conference* (ed. U. Frisch). Kluwer, Dordrecht, The Netherlands.
32. Ricca, R.L. & Weber, C. 1998 In preparation.
33. Saffman, P.G. 1992 *Vortex Dynamics*. Cambridge University Press.
34. Samuels, D.C., Barenghi, C.F. & Ricca, R.L. 1998 *J. Low Temp. Physics* **100**, 509–514.
35. Tait, P.G. 1898 In *Scientific Papers*. Cambridge University Press.
36. Thomson, J.J. 1883 *A Treatise on the Motion of Vortex Rings*. Macmillan & Co., London.
37. Vogel, P. 1990 *Comment. Math. Helvetici* **65**, 104–113.

CHAPTER 15

HAMILTONIAN APPROACH TO KNOTTED SOLITONS

Antti J. NIEMI

Department of Theoretical Physics, Uppsala University, P.O. Box 803, S-75108, Uppsala, Sweden

We describe a first principles Hamiltonian approach, where knots appear as solitonic solutions to the pertinent nonlinear equations of motion. This makes it possible to predict all properties of knotted configurations in terms of fundamental data that are only characteristics of the underlying physical environment. In particular, no knot-specific parameters appear in this approach.

1 Introduction

The study of knots - mathematically defined as embeddings of a circle S^1 in three dimensional Euclidean space R^3 - is rapidly becoming highly important in various physical, chemical and biological problems, ranging from early universe cosmology and the structure of elementary particles to He^3 superfluids, polymers and DNA. By studying the properties of knots and the mechanisms which lead to their formation, we learn how highly nonlinear, complex systems organize themselves into regular global patterns.

Ideally, we would like to predict the properties of a knotted configuration directly from those of the underlying physical environment. For this we need a mathematical description of the environment, where the knots appear *dynamically* as solutions to the pertinent equations of motion¹. In such a first principles approach we can then make definite predictions. In particular, all physical characteristics of any knot such as its mass (energy), length, thickness, overall geometrical shape, and also the interactions between different knots, become quantities that at least in principle can be computed directly from the underlying theory. A first principles approach to knots can not contain any intrinsically knotlike quantities, all parameters that appear in the theory should either describe the generic, statistic properties of the physical environment such as its temperature, viscosity *etc*, or then be *a priori* measurable quantities of the microscopic pointlike constituents such as the mass and electric charge of elementary particles, atoms *etc*.

The first who proposed that such a first principles approach to knots should be possible was Lord Kelvin. In² he suggested that (thin) vortex filaments in the shape of torus knots - a general family of knots that can be described as a loop wrapping around the surface of a doughnut a number of times³ - should be stable. However, at the time the theoretical understanding of nonlinear

phenomena was insufficient for realizing his conjecture. Indeed, it could not even be formulated in a proper mathematical framework.

Today Kelvin's conjecture on the stability of torus knots still remains unproven, but now we do have a mathematically consistent formulation. Moreover, tentative numerical simulations do indicate that Kelvin's conjecture could be realized in a definite theoretical model¹.

In order to formulate Kelvin's conjecture in a mathematically consistent manner¹, we need a three dimensional Hamiltonian field theory, with knots appearing as finite energy *solitonic*⁴ solutions to the pertinent Euler-Lagrange equations of motion. The underlying physical principles that leads us to adopt a Hamiltonian framework with knots as solitons, are today universally accepted as fundamental. Indeed, a physical theory which fails to admit a Hamiltonian interpretation should be interpreted as an effective theory of some underlying Hamiltonian structure. A Hamiltonian approach allows us to describe all properties of knots in a predictive manner, as functions of the fundamental physical parameters which appear in the Hamiltonian. In particular, in this approach there are *no* knot specific *ad hoc* parameters, but *all* properties of knots are derived directly from those of the physical environment where knots are formed, by solving the Hamiltonian equations of motion. At least in principle, we can then compute all physical characteristics of all knots such as their stability and overall shape including thickness, length, energy (mass), string tension *etc.*, and we can study their vibrational and rotational excitations and effects of quantum mechanical and thermal fluctuations, and investigate interactions between different knotlike configuration. The results are presented as functions of a relatively small number of *a priori* known parameters that characterize only the physical environment where the knots are formed, no adjustable or measurable knot specific parameters are present.

Unfortunately, the equations of motion that describe knotted solitons are highly complex nonlinear partial differential equations. Even the very construction of a Hamiltonian for knots is quite nontrivial, and until very recently no examples were known. Fortunately, it appears that this complexity also strongly limits the possible form of Hamiltonians that describe knots, maybe even to a point where we can eventually have a full *classification*. Indeed, since a knot is a highly complicated topological object, the dynamical field *i.e.* order parameter that appears in the Hamiltonian, is strongly restricted by the requirement that it must allow for the description of the various topological properties such as the linking number of a knot. The Hamiltonian must also be consistent with various obvious symmetries such as rotation and translation invariance, and the functional form of the nonlinear terms are restricted by the requirement that the equations of motion indeed support stable knotted

solitons. The concept of *universality* can also be introduced. In the present context it states that two Hamiltonians with the same order parameter are in the same universality class provided they differ from each other only by such higher order derivative terms which do not influence the stability of knots. These higher order terms in a derivative expansion become generically relevant only when we are interested in higher order corrections to explicit profiles such as the energy density distribution. But at the level where we are only interested in global, large scale topological or stability aspects these higher order terms are not relevant. By implementing these and possibly additional natural principles like demanding locality of the Hamiltonian, we expect that we may eventually succeed in obtaining a relatively simple classification of all Hamiltonian field theory models that can describe knotted solitons.

Until now, we have found only one viable candidate of a three dimensional Hamiltonian field theory model where knots appear as finite energy solitons¹. This model is indeed highly nonlinear, to the extent that any analytic approach to study the properties of its knots is essentially hopeless. Even a numerical investigation of its properties is very demanding¹, and is only now becoming realistic with the development of a new generation of supercomputers and novel computational techniques. Even though for the moment only very tentative numerical investigations of this model have been performed, the results are however quite encouraging and strongly suggest that it indeed describes Kelvin's torus knots as stable solitons. Since the theoretical principles that lead to the construction of the model are quite restrictive, it may be essentially unique for a general class of knotlike solitons, strongly supporting the importance of the concept of universality in fundamental knot theory.

In the following we shall describe the Hamiltonian approach to knots as solitons, introduced in¹. We concentrate on the general topological properties of this approach, but we shall also address some of the complications that appear in numerical simulations.

2 A Knot Hamiltonian

Knots are highly complicated topological objects, and in order to analyze their properties various invariants have been introduced. Such invariants are useful *e.g.* when we wish to determine, whether two *a priori* different knots are actually topologically equivalent in the sense that they can be continuously transformed onto each other by Reidemeister moves³. Such moves do not allow for transformations that break the knot, and in particular do not allow the knot to cross through itself. The most complete of such knot invariants are the polynomial invariants introduced by Jones³. However, at the moment

Jones polynomials have not yet been cultivated into a effective computational tool that could be used in connection of a Hamiltonian field theory approach. For the present purpose it is more convenient to use the Hopf invariant, a much more rudimentary topological invariant that computes the (self)linking number of knots.

Besides a (self)linking number of a knot, the Hopf invariant can also be interpreted as a winding number for maps from a unit sphere S^3 in four dimensional Euclidean space to a unit sphere S^2 in three dimensions. This enables us to introduce a dynamical order parameter for knots: The maps $S^3 \rightarrow S^2$ fall into disjoint homotopy classes, labelled by integers using the homotopy group $\pi_3(S^2) \sim \mathbb{Z}$. Since S^3 is topologically identical with the three dimensional Euclidean space when we identify all points at infinity, $S^3 \sim \mathbb{R}^3 \cup \{\infty\}$, the Hopf invariant can be described using a three component unit vector $\mathbf{n}(\mathbf{x}) : \mathbb{R}^3 \rightarrow S^2$

$$\mathbf{n}(\mathbf{x}) = \begin{pmatrix} n_1(\mathbf{x}) \\ n_2(\mathbf{x}) \\ n_3(\mathbf{x}) \end{pmatrix}$$

that approaches a constant vector at spatial infinity,

$$\lim_{|\mathbf{x}| \rightarrow \infty} \mathbf{n}(\mathbf{x}) = \mathbf{n}_0$$

Since knots are highly localized configurations in \mathbb{R}^3 , this vector field $\mathbf{n}(\mathbf{x})$ with its asymptotic boundary condition is then also an appropriate dynamical variable *i.e.* an order parameter to describe the dynamics of knots. Furthermore, since it is natural to expect that the relevant Hamiltonian for $\mathbf{n}(\mathbf{x})$ should be rotation invariant, we can always select the asymptotic vector \mathbf{n}_0 as

$$\mathbf{n}_0 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \quad (1)$$

The Hopf invariant of $\pi_3(S^2)$ can be represented as an integral invariant of the field $\mathbf{n}(\mathbf{x})$. For this, we define the antisymmetric tensor

$$F_{ij} = \epsilon_{abc} n^a \partial_i n^b \partial_j n^c \quad (2)$$

where ϵ_{abc} is antisymmetric in its indices with $\epsilon_{123} = 1$. We recognize F_{ij} as the pull-back of the volume element two-form on the target S^2 . Since it is closed,

$$dF = \epsilon_{ijk} \partial_i F_{jk} = 0$$

we can represent it as a curl of a vector field A_i ,

$$F = dA \Leftrightarrow F_{ij} = \partial_i A_j - \partial_j A_i \quad (3)$$

Modulo a U(1) gauge transformation $A_i \rightarrow A_i + \partial_j \gamma$ we can then use (2) to solve for A_i as a functional of $\mathbf{n}(\mathbf{x})$, but unfortunately the result fails to be local in $\mathbf{n}(\mathbf{x})$.

The Hopf invariant is the integral

$$Q_H = \frac{1}{4\pi} \int d^3x \epsilon_{ijk} F_{ij} A_k \quad (4)$$

and one can verify that it is indeed a topological invariant of \mathbf{n} , *i.e.* it remains invariant under a local variation $\mathbf{n} \rightarrow \mathbf{n} + \delta\mathbf{n}$,

$$\delta Q_H = \frac{1}{4\pi} \delta \int d^3x \epsilon_{ijk} F_{ij} A_k = 0$$

This can be verified by explicitly varying Q_H with respect to $\mathbf{n}(\mathbf{x})$. It can also be verified by introducing a parametrization of $\mathbf{n}(\mathbf{x})$ in terms of a unit four vector $\Phi_\mu(\mathbf{x})$ ($\mu = 1, 2, 3, 4$), an approach that has the advantage that it makes the connection between the Hopf invariant and the linking number of a knot manifest. For this, we arrange Φ_μ into a two component complex vector $Z(\mathbf{x})$

$$\begin{aligned} Z_1(\mathbf{x}) &= \Phi_1(\mathbf{x}) + i\Phi_2(\mathbf{x}) \\ Z_2(\mathbf{x}) &= \Phi_3(\mathbf{x}) + i\Phi_4(\mathbf{x}) \end{aligned}$$

so that

$$|Z_1|^2 + |Z_2|^2 = 1$$

and use it to parametrize the unit vector $\mathbf{n}(\mathbf{x})$ by

$$n^a = Z^\dagger \sigma^a Z \quad (5)$$

with σ^a the standard Pauli matrices. The two-form F_{ij} in (2) then becomes

$$F_{ij} = i(\partial_i Z^\dagger \partial_j Z - \partial_j Z^\dagger \partial_i Z) \quad (6)$$

and modulo a U(1) gauge transformation we have

$$A_i = \frac{i}{2} (Z^\dagger \partial_i Z - \partial_i Z^\dagger Z) \quad (7)$$

Indeed, under a U(1) gauge transformation determined by a function $\gamma(\mathbf{x})$ such that $Z \rightarrow e^{i\gamma} Z$, both F_{ij} and \mathbf{n} remain invariant while A_i is shifted by

$A_i \rightarrow A_i + \partial_i \gamma$. This implies that Q_H is indeed a functional of \mathbf{n} only, and substituting in (4) we get

$$Q_H = \frac{1}{12\pi^2} \int d^3x \epsilon_{\mu\nu\rho\sigma} \epsilon_{ijk} \Phi_\mu \partial_i \Phi_\nu \partial_j \Phi_\rho \partial_k \Phi_\sigma \quad (8)$$

which is the standard integral representation of the winding number for maps $S^3 \rightarrow S^3$. In particular, by antisymmetry one directly finds that Q_H is manifestly invariant under arbitrary local variations $\Phi_\mu \rightarrow \Phi_\mu + \delta\Phi_\mu$ so that (4) is indeed an integer valued topological invariant of the vector field $\mathbf{n}(\mathbf{x})$.

If we parametrize the unit vector $\mathbf{n}(\mathbf{x})$ by

$$\mathbf{n}(\mathbf{x}) = \begin{pmatrix} \sin \varphi \cdot \sin \theta \\ \cos \varphi \cdot \sin \theta \\ \cos \theta \end{pmatrix} \quad (9)$$

our choice (1) implies that

$$\theta(\mathbf{x}) \xrightarrow{|\mathbf{x}| \rightarrow \infty} 0 \pmod{2\pi} \quad (10)$$

so that at large distances from the knot we approach the north pole of the target S^2 with $\theta(\vec{\mathbf{x}}) \approx 0$. For a nontrivial Hopf invariant we then need $\theta(\vec{\mathbf{x}}_c) = \pi$ at some points $\vec{\mathbf{x}}_c$. This corresponds to the south pole of the target S^2 . We argue that the preimage of $\vec{\mathbf{x}}_c$ in R^3 coincides with the center *i.e.* core of the knot, described by an embedding of a circle S^1 in R^3 . For this we observe that the internal structure of a knot can be investigated by cutting the knot once at its generic point with a plane that makes a right angle to its core. This cross sectional plane is topologically identical to a sphere S^2 : At the core we have $\theta(\vec{\mathbf{x}}_c) = \pi$ corresponding to the south pole and outside of the knot on the cross sectional plane we have $\theta(\vec{\mathbf{x}}) \approx 0$ corresponding to the north pole. The physical extent (or thickness) of the knot is then determined by the region where $\theta(\mathbf{x})$ varies from $\theta = \pi$ to $\theta = 0$. Furthermore, $\varphi(\vec{\mathbf{x}})$ increases (or decreases, depending on orientation) by 2π when we go around the core once on the cross sectional plane. (More generally, φ is defined modulo $2\pi n$ where n is an integer.) Since we have selected the cross sectional plane at a generic point along the knot, this implies that instead of a mapping from S^3 into S^2 , for a knot we actually have a mapping $S^2 \times S^1 \rightarrow S^2$, where the S^2 in the pre-image corresponds to the cross-sectional plane at each point along the core S^1 of the knot. This local, cross sectional picture of a knot described by $\mathbf{n}(\mathbf{x})$ becomes transparent when we introduce the following parametrization of the

four vector $\Phi_\mu(\mathbf{x})$,

$$\Phi = \begin{pmatrix} \cos \phi_{12} \sin \vartheta \\ \sin \phi_{12} \sin \vartheta \\ \cos \phi_{34} \cos \vartheta \\ \sin \phi_{34} \cos \vartheta \end{pmatrix} \quad (11)$$

Comparing (5), (9) we find

$$\begin{aligned} \varphi &= \phi_{34} - \phi_{12} \\ \theta &= 2\vartheta \end{aligned} \quad (12)$$

which implies in particular, that the pre-image $\vec{\mathbf{x}}_c$ of the south pole $\theta = \pi$ is indeed a circle $S^1 \in R^3$, parametrized by the angle ϕ_{12} - the parameter for the core of the knot.

For the vector field (7) we find

$$A_i = \cos^2 \vartheta \partial_i \phi_{34} + \sin^2 \vartheta \partial_i \phi_{12} \quad (13)$$

and (2), (6) yields

$$F = dA = \sin 2\vartheta d\vartheta \wedge (d\phi_{34} - d\phi_{12}) \quad (14)$$

so that the Hopf invariant (4) becomes

$$Q_H = \frac{1}{\pi^2} \int \sin 2\vartheta d\vartheta \wedge d\phi_{34} \wedge d\phi_{12} \quad (15)$$

Here the asserted local $S^1 \times S^2$ structure of a knot is manifest.

We shall now proceed to the construction of the energy density (Hamiltonian) of knots described by the vector field $\mathbf{n}(\mathbf{x})$. Due to a global rotation symmetry of the Hamiltonian, we conclude that the vector field $\mathbf{n}(\mathbf{x})$ can appear in the energy density only thru its derivatives. This allows us to construct the Hamiltonian systematically, by expanding it in the powers of derivatives of $\mathbf{n}(\mathbf{x})$. The simplest nontrivial, rotationally invariant term involves two derivatives,

$$E_2 = g^2 \int d^3x \partial_i \mathbf{n} \cdot \partial_i \mathbf{n} \quad (16)$$

where g is a coupling constant; in natural units it is inversely proportional to a length scale when we select $\mathbf{n}(\mathbf{x})$ to be dimensionless. The Hamiltonian (16) is the standard action of the $O(3)$ nonlinear σ -model, known to admit stable solitons in two dimensions⁴. However, for stable finite energy solitons in three

dimensions, we need to introduce additional terms that are of higher order in derivatives of $\mathbf{n}(\mathbf{x})$. The first such term appears at fourth order in derivatives, and by demanding that the Hamiltonian admits a relativistic interpretation (relevant in applications to high energy physics) this term is uniquely given by

$$E_4 = \int d^3x F_{ij}^2 = \int d^3x (\partial_i \mathbf{n} \cdot \partial_j \mathbf{n})^2$$

(modulo an overall dimensionless parameter, that we have normalized to unity). The following Hamiltonian^{5, 1}

$$E = E_2 + E_4 \tag{17}$$

is then our proposal for a Hamiltonian that describes knotted solitons. Indeed, using Sobolev inequalities one can show⁶ that the energy density (17) is bounded from below by the Hopf invariant

$$E \geq c \cdot |Q_H|^{\frac{3}{4}}$$

where the constant c is known to be nonvanishing. This is a strong evidence that solitons with a nontrivial Hopf invariant *i.e.* with nontrivial self-linking, should exist as solutions to the Euler-Lagrange equations of (17).

The existence of finite energy solitons in (17) is also consistent with the Derrick scaling argument: If we assume that $\mathbf{n}(\mathbf{x})$ is a critical point of the Hamiltonian and consider a scaling $\mathbf{x} \rightarrow \lambda \cdot \mathbf{x}$ so that $\mathbf{n}(\mathbf{x}) \rightarrow \mathbf{n}(\lambda \cdot \mathbf{x})$, then the energy density of the scaled configuration

$$E(\lambda) = \lambda E_2 + \lambda^{-1} E_4$$

must have an extremum at $\lambda = 1$. This yields the following virial theorem

$$E_2 = E_4 \tag{18}$$

which is necessarily obeyed by a knotlike soliton. In a numerical construction of knots this virial theorem appears as a most useful tool.

Besides E_2 and E_4 , we may also consider higher order derivative terms in (17). But such terms are not needed for the stability of the knot, they will only have an effect on the local details of the energy density distribution on the cross sectional planes. As such they may be relevant in some applications, but should not have any effect on the large distance, overall topological properties of the knot. In this sense the minimal Hamiltonian (17) determines a *universality class*.

In constructing an actual knotted soliton, instead of the angular variables $\theta(\mathbf{x})$ and $\varphi(\mathbf{x})$ in (9) we find it more convenient to introduce real variables $U(\mathbf{x})$ and $V(\mathbf{x})$ obtained by

$$\varphi = -\arctan\left(\frac{V}{U}\right)$$

$$\theta = 2\arctan\sqrt{U^2 + V^2}$$

Geometrically $U(\mathbf{x})$ and $V(\mathbf{x})$ are coordinates on a Riemann sphere, defined by projecting S^2 with respect to the north pole $\theta = 0$. In these variables the Hamiltonian (17) becomes

$$H = \int d^3x \frac{4g^2}{(1 + U^2 + V^2)^2} (\partial_\mu U^2 + \partial_\mu V^2) + \frac{16}{(1 + U^2 + V^2)^4} (\partial_\mu U \partial_\nu V - \partial_\nu U \partial_\mu V)^2 \quad (19)$$

3 Aspects of Numerical Solution

The Euler-Lagrange equations obtained by varying (19) with respect to $U(\mathbf{x})$ and $V(\mathbf{x})$ are highly nonlinear, to the extent that an analytic solution is essentially impossible. Consequently it appears that the only tools available for studying the knots in (19) are numerical. However, it turns out that even a numerical integration of the pertinent equations is highly nontrivial, there are several theoretical complications that need to be resolved¹.

One of the issues that we need to address concerns the (expected) chaotic nature of the highly nonlinear equations of motion. It is a reason, why we do not expect a direct Newton's iteration to converge towards a stable configuration unless we somehow succeed in constructing an initial configuration which is very close to an actual solution. Instead, we expect Newton's iteration to exhibit chaotic behavior.

Since it is very difficult to construct an initial configuration which is known to be close to an actual solution, we find it convenient to first formulate the problem at an abstract level. For this we consider a generic static energy functional $E(q)$ with some variables q_a . We introduce an auxiliary variable τ , and instead of the original Euler-Lagrange equations obtained by varying $E(q)$ with respect to its dynamical variables q_a , we extend the (stationary) Euler-Lagrange equations of $E(q)$ to the following parabolic gradient flow equation

$$\frac{dq_a}{d\tau} = -\frac{\delta E}{\delta q_a} \quad (20)$$

Since

$$\frac{\partial E}{\partial \tau} = -\left(\frac{\delta E}{\delta q_a}\right)^2 \quad (21)$$

we conclude that the energy decreases along the trajectories of (20). Furthermore, if we square (20) and integrate from some initial value $\tau = T$ to $\tau \rightarrow \infty$ we get

$$\int_T^\infty d\tau \left(\frac{dq_a}{d\tau}\right)^2 = \int_T^\infty d\tau \left(\frac{\delta E}{\delta q_a}\right)^2 \quad (22)$$

Hence a τ -bounded trajectory of (20) must flow towards a stable critical point of $E(q)$. In particular, by starting at an initial time $\tau = T$ from an initial configuration $q = q_0$ and by following a bounded trajectory of (20), in the $\tau \gg T$ limit we flow towards a stable critical point of $E(q)$.

In (19) the angles $U(\mathbf{x})$ and $V(\mathbf{x})$ correspond to the generic variables q_a , and by denoting $W_1 = U$, $W_2 = V$ we find for the flow equation

$$\frac{\partial W_a(\mathbf{x})}{\partial \tau} = -g^2 \cdot \frac{\delta E_2}{\delta W_a(\mathbf{x})} - \frac{\delta E_4}{\delta W_a(\mathbf{x})} \quad (23)$$

which is the equations that we solve in an actual numerical simulation¹.

Notice that if we introduce the scaling $\mathbf{x} \rightarrow \lambda \cdot \mathbf{x}$ we find that the scaled Hamiltonian obeys the same flow equation but with scaled variables,

$$\frac{\partial W_a(\mathbf{x})}{\partial(\frac{1}{\lambda}\tau)} = -(\lambda^2 g) \cdot \frac{\delta E_2}{\delta W_a(\mathbf{x})} - \frac{\delta E_4}{\delta W_a(\mathbf{x})} \quad (24)$$

This implies that a flow towards a knot with coupling constant g coincides with the flow towards a topologically identical knot with coupling constant $\lambda^2 g$, provided we rescale the flow variable τ into $\frac{1}{\lambda}\tau$. Thus a knotlike soliton is essentially unique; it is sufficient to consider the flow towards a soliton with a definite value for the coupling constant g , as solitons with other values of g are obtained from this configuration by scaling with the proper parameter λ .

In a numerical simulation we need in addition to decide an optimal overall size for the lattice that we use in formulating our simulation. If the lattice is too small in comparison to the scale of the knot, the knot may not fit in the lattice. On the other hand, if the lattice is too large in comparison to the knot, we may either miss the knot entirely or use an unnecessarily large amount of computer time in constructing it.

Since the coupling constant g in (20) is the only dimensionfull quantity that appears in our equations, it determines the physical extent of the knot such as

the location $\vec{\mathbf{x}}_c$ of its core and the overall thickness around this core. Besides g , there are only dimensionless numerical factors that affect the overall size of the knot. Since such numerical factors can only be obtained by actually solving the equations of motion, we conclude that in general the problem of selecting an optimal size of the numerical integration lattice can be quite involved.

In order to approach this problem, we have developed a simple renormalization procedure that allows us to select the initial configuration so that the location $\vec{\mathbf{x}}_c$ of its core approximatively coincides with the location of the core for an actual soliton¹. For this we first observe that a knotted soliton obeys the virial theorem (18), in our present variables

$$g^2 E_2 = E_4$$

By demanding that this virial theorem is also obeyed during the flow (23), we promote the coupling constant g into a τ -dependent variable $g \rightarrow g(\tau)$,

$$g(\tau) = \sqrt{\frac{E_4(\tau)}{E_2(\tau)}}$$

When we approach an actual knotlike soliton as $\tau \rightarrow \infty$, the variable $g(\tau)$ must then flow towards an asymptotic value g^* which is the value of the coupling for the actual soliton,

$$g(\tau) \xrightarrow{\tau \rightarrow \infty} g^* \tag{25}$$

This renormalization procedure then fixes the overall scale for the location of the core at $\vec{\mathbf{x}}_c$, and allows us to choose the size of our lattice appropriately. Notice that the (local) thickness of the final knot is also determined by the asymptotic value g^* of the coupling constant, through the equations of motion. However, the renormalization (25) does not help us very much in selecting the thickness of the initial condition so that numerical convergence is secured. This poses a problem for which we at the moment lack a firm solution, besides experimenting with various different types of initial configurations.

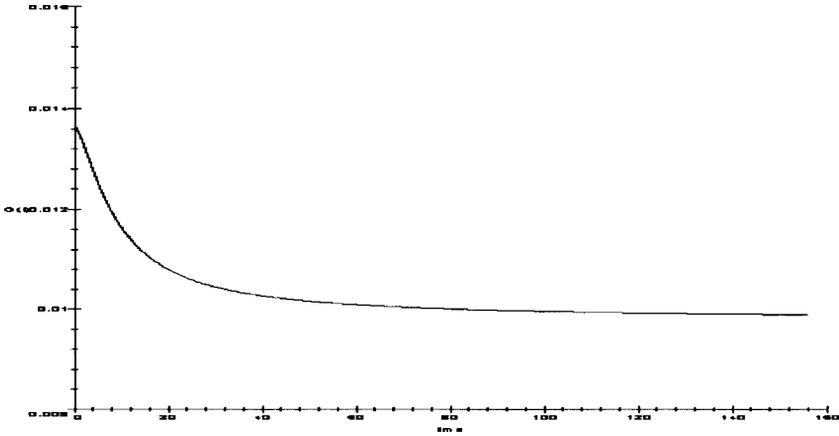


Figure 1: When we approach an actual knotlike soliton, the coupling constant $g(\tau)$ flows towards an asymptotic value g^*

Finally, on a finite size lattice we also need to determine boundary conditions at the boundary of the lattice. The proper boundary condition for $U(\mathbf{x})$ and $V(\mathbf{x})$ is that both vanish when we are far away from the knot. However, we expect that on a finite lattice a simulation with such trivial boundary conditions can only lead to a flow towards the trivial configuration $U(\mathbf{x}) \equiv V(\mathbf{x}) \equiv 0$. In order to impose the boundary conditions properly in our numerical simulation, we can adapt an iterative process where we first specify the boundary conditions using the initial configuration $U_0(\mathbf{x}), V_0(\mathbf{x})$. At later values of τ we then update these boundary conditions successively, by interpolating the iterated configurations from the interior of the lattice to its boundary. In this manner we expect that we eventually obtain boundary conditions that correspond to those of an actual knotlike soliton. An alternative method that can be used to determine the boundary conditions, is to start the iteration of (23) from a sufficiently large initial lattice with boundary conditions determined by the initial configuration. By successively shrinking the size of the lattice and determining boundary conditions in the shrunked lattice using the appropriate restriction of the pertinent iterated configuration from the larger lattice, we expect to converge towards boundary conditions that coincide with those of an actual soliton.

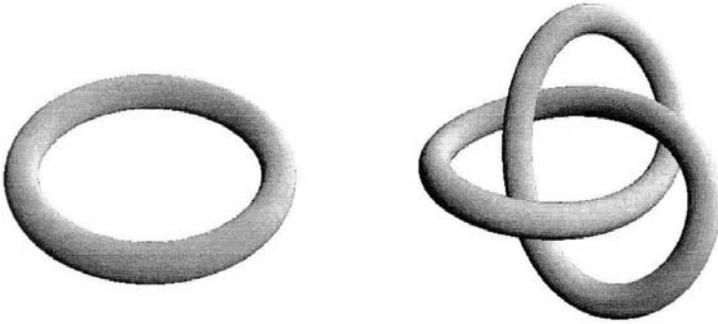


Figure 2: The unknot (left) and the trefoil (right) are the simplest examples of knots that have been constructed numerically.

In a generic simulation of the flow equation (23) we determine the length of each time step $\Delta\tau_n = \tau_n - \tau_{n-1}$ adaptively, either by demanding that the relative variation of total energy remains bounded, *e.g.*

$$\left| \frac{\Delta E(\tau_n)}{E(\tau_n)} \right| \leq 10^{-4} \quad (26)$$

or then that the relative error at each lattice node remains bounded, *e.g.*

$$|W_a(\tau_n) - W_a(\tau_{n-1})| \leq 10^{-4} \quad (27)$$

We have performed¹ extensive numerical simulations both for the toroidal unknot soliton and the simplest nontrivial torus knot, the trefoil. Our simulations are consistently converging towards definite fixed points, suggesting that the present Hamiltonian indeed describes stable knotlike solitons. In particular, our results indicate that the present model should realize Kelvin's conjecture on the existence of stable torus knots.

In our simulations we have found the PDE2D finite element algorithm⁷ quite convenient. We have chosen a finite element approach, since it computes a continuous piecewise polynomial approximation to the solution. In a problem of topological nature this should be a definite advantage *e.g.* over a finite difference approach, where the solution is approximated only at discrete lattice nodes.

The simulations that we have performed, while very extensive should at this time be still considered as tentative: These simulations, and related but independent simulations performed by other groups, have clearly revealed the complexity of the problem. The computers available today are simply not sufficiently powerful to allow for an extensive investigation of complicated knots.

However, it appears that with the present rate of development in computer CPU and processor speed, we are quite soon able to make a serious, detailed numerical investigation of knotted solitons in the present model. Indeed, we hope that by the time this article is published, detailed realistic simulations have already appeared in the literature. Thus the numerical study of knots, and the realization of Kelvin's dream appears to finally become a reality.

Acknowledgments

I wish to thank Ludvig Faddeev for collaboration, and Sami Virtanen for helping with the figures.

References

1. L.D. Faddeev and A.J. Niemi, *Nature* **387**, 58 (1997), and **hep-th/9705176**.
2. Lord Kelvin (W.Thomson) *Mathematical and Physical Papers* vol. 4 (Cambridge University Press, Cambridge 1910).
3. L.H. Kauffman, *Knots and Physics* (World Scientific, Singapore, 1993).
4. C. Rebbi and G. Soliani, *Solitons and Particles* (World Scientific, Singapore, 1984).
5. L.D. Faddeev, *Quantization of Solitons* (Princeton preprint IAS-75-QS70, Institute for Advanced Study, 1970)
6. A.F. Vakulenko and L.V. Kapitanski, *Dokl. Akad. Nauk USSR* **248**, 810 (1979)
7. G. Sewell, *Adv. Eng. Software* **17**, 105 (1993)

CHAPTER 16

ENERGY OF KNOTS

J. O'HARA

*Department of Mathematics, Tokyo Metropolitan University,
1-1 Minami-Ohsawa, Hachioji-Shi, Tokyo 192-03, JAPAN*

A *knot energy functional* is a real valued functional on the space of knots that blows up if a knot has self-intersections. We consider the regularization of modified electrostatic energy of charged knots. We study several kinds of knot energy functionals and consider whether there exists an *energy minimizer* in each knot type, which is characterized as an embedding that attains the minimum value of the knot energy functional within its ambient isotopy class.

1 Introduction

A *knot* f is an embedding from a circle S^1 into a 3 dimensional manifold M . Usually we take the Euclidian space \mathbf{R}^3 or the 3-sphere S^3 as M . A *knot type* $[f]$ is an ambient isotopy class of a knot f . When M is \mathbf{R}^3 or S^3 it is same as the equivalence class of f with respect to the orientation preserving homeomorphisms of \mathbf{R}^3 or S^3 .

Let us consider the problem of identifying the knot type of a given knot diagram. Unfortunately there are no knot invariants that have been proved to detect all the knot types completely. Our approach is to define a suitable functional on the space of knots and define the **ideal configuration** for each knot type as the embedding that attains the minimum value of the functional within its ambient isotopy class. Suppose we can deform a knot along the gradient flow so as to decrease the value of the functional until it comes to a critical point. Assume that the knot type is kept unchanged during this deformation process. Assume also that the number of critical points is finite for each knot type. Then we can detect the knot type. In order to guarantee the invariance of the knot type during the deformation process, crossing changes should not be allowed. Thus we require that our functional should satisfy the following property:

(*) If a knot degenerates to an immersion with double points then the functional blows up.

We call our functional a **knot energy functional** if it satisfies the above property. We remark that this property is not sufficient to guarantee the existence of the ideal configuration for each knot type because a knot might degenerates to a map that does not belong to the original knot type as is explained in Conjecture 5 and Conjecture 12.

Let \mathcal{E} be a knot energy functional. Define the *minimum energy of a knot type* $[f]$, $\mathcal{E}([f])$, by

$$\mathcal{E}([f]) = \inf_{\bar{f} \in [f]} \mathcal{E}(\bar{f}).$$

If there is a knot $f_0 \in [f]$ that satisfies $\mathcal{E}(f_0) = \mathcal{E}([f])$ we call f_0 an **energy minimizer** of $[f]$ with respect to \mathcal{E} or the **\mathcal{E} -minimizer** for short. Our motivation is to find a good knot energy functional which provides ideal configurations of knots as its minimizers.

This article is arranged as follows. In §2 we define the α -energy $E^{(\alpha)}$ as the *regularization* of the $r^{-(\alpha+1)}$ -modified electrostatic energy ($\alpha \geq 2$) of charged knots and give some formulae. We call α the *power index* of $E^{(\alpha)}$. In §3 we study $E = E^{(2)}$. Freedman, He and Wang showed that E is *Möbius invariant* and therefore that there are infinitely many E -minimizers for each prime knot type. On the other hand Kusner and J. Sullivan conjectured that there are no E -minimizers in any composite knot type. We give two ways to produce energy minimizers for all the knot types later in §4 and §5. Then we give a criterion for a knot to be E -critical. We introduce Kim and Kusner's conjecture on the existence of unstable E -critical (p, q) -torus knots ($p > 2$ or $q > 2$). Finally we introduce Freedman, He and Wang's results that E bounds the *average crossing number* and hence that only finitely many knot types can occur under any given threshold on E . In §4 we define the (α, p) -energy functional for knots $e^{\alpha, p}$ with a higher power index αp with $E^{(\alpha)} = \alpha e^{\alpha, 1}$. We show that only finitely many solid tori appear as tubular neighborhoods of knots under any given $e^{\alpha, p}$ ($\alpha p > 2$) threshold, and therefore that there are $e^{\alpha, p}$ -minimizers ($\alpha p > 2$) for any knot type. We also define an α -energy polynomial for knots. In §5 we define the (α, p) -energy functional $e_M^{\alpha, p}$ for knots in a Riemannian 3-manifold M . We show that there are $e_M^{\alpha, p}$ -minimizers ($\alpha p > 2$) for any knot type if M is compact or if M is the hyperbolic 3 space \mathbf{H}^3 . When $M = S^3$ we conjecture that there are (hopefully finitely many) E_{S^3} -minimizers for each knot type. Thus the existence of energy minimizers depends on the power index of energy, the primeness of knots, and the metric of the ambient space. In §6 we consider several kinds of quantities such as the *thickness* and the *self distance*.^a In §7 we make some remarks on related topics. In §8 we give a short summary.

Throughout this article we assume suitable differentiability for knots, mostly C^2 .

This article is mostly a translation of [63]. The updated version will be available at

^aSome of them are studied in other chapters in this book, though their definitions might be slightly different.

<http://www.comp.metro-u.ac.jp/~ohara/texfiles/>.

The reader is referred to [56-62] for the details unless otherwise mentioned.

2 Definition of α -energy $E^{(\alpha)}$

One of the most natural and naive candidates for a knot energy functional would be the electrostatic energy of charged knots. Another candidate would be the bending and twisting energy of elastic knots which will be considered later in §7 (4). The first attempt to consider the electrostatic energy of charged knots was made in the finite dimensional category by Fukuhara [22]. He considered the space of the polygonal knots whose vertices are charged, and studied their *modified* electrostatic energy under the assumption that Coulomb's repelling force between a pair of point charges of distance r is proportional to r^{-m} ($m = 3, 4, 5, \dots$). The reason for this modification is given in what follows. He wrote a computer program to evolve a knot to decrease this energy.

Let $h : S^1 = \mathbf{R}/\mathbf{Z} \rightarrow \mathbf{R}^3$ be a knot. To avoid complications we assume that h is parametrized by the arc-length, namely, $|h'(t)| = 1$ ($\forall t \in S^1$). Hereafter we always assume this hypothesis whenever we denote a knot by h .

Suppose that the knot is uniformly electrically charged. (We need this assumption of uniformity for the following reason. Suppose electrons can move freely along a knot. Then the energy of a knot which is almost same as the standard circle except for a small tangle can be close to that of the standard circle itself if the tangle part carries a small charge. Thus every knot will degenerate to the circle in order to decrease its energy.)

Then the electrostatic energy of a charged knot $h(S^1)$ is given by

$$\iint_{S^1 \times S^1} \frac{dx dy}{|h(x) - h(y)|}$$

which turns out to be infinite for any knot h , because this integral blows up around the diagonal set $\Delta = \{(x, y) \in S^1 \times S^1 | x = y\}$. There are two ways to cancel the blow-up to obtain a finite valued functional. The simplest way is to subtract a term which blows up to the same order as $|h(x) - h(y)|^{-1}$ at Δ . An alternative way is multiplication by a term which goes to zero to the same order as $|h(x) - h(y)|$ at Δ , as we will see in (6) and §7 (2). (Note that this can also be viewed as subtraction after we take the logarithm.) Let us consider the subtraction method first. In this case we can take $\delta(x, y)^{-1}$ as the subtraction term, where $\delta(x, y)$ is the shortest arc-length between $h(x)$ and $h(y)$ along the knot, which is the same as $\min\{|x - y|, 1 - |x - y|\}$ in this case. Then we have

$$E^{(1)}(h) = \iint_{S^1 \times S^1} \left(\frac{1}{|h(x) - h(y)|} - \frac{1}{\delta(x, y)} \right) dx dy,$$

but $E^{(1)}$ is not a knot energy functional since $E^{(1)}$ takes a finite value on a "singular knot" with a self-intersection, as we shall see in (4). Thus we are obliged to consider the modified electrostatic energy defined under the non-physical assumption that Coulomb's repelling force between a pair of point charges of distance r is proportional to $r^{-(\alpha+1)}$ where $\alpha > 1$. Thus modified electrostatic energy of a charged knot $h(S^1)$ is given by

$$\iint_{S^1 \times S^1} \frac{dx dy}{|h(x) - h(y)|^\alpha},$$

which is again infinite for any knot h . In order to determine the subtraction term let us expand $|h(x) - h(y)|^{-\alpha}$ by $|x - y|$. Suppose h is smooth. Since $(h', h') \equiv 1$ we have $(h', h'') = 0$, $(h', h^{(3)}) + |h''|^2 = 0$, and so on. Therefore Taylor's formula implies

$$\begin{aligned} |h(x+t) - h(x)|^{-\alpha} &= t^{-\alpha} + \frac{\alpha|h''(x)|^2}{24}t^{2-\alpha} + \frac{\alpha(h''(x), h^{(3)}(x))}{24}t^{3-\alpha} \\ &+ \left\{ \frac{\alpha}{2} \left(\frac{(h''(x), h^{(4)}(x))}{40} + \frac{|h^{(3)}(x)|^2}{45} \right) + \frac{\alpha(\alpha+2)|h''(x)|^4}{1152} \right\} t^{4-\alpha} + \dots \quad (1) \end{aligned}$$

for $0 < t \ll 1$. Since

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} t^p dt < \infty \quad (p > -1)$$

we can take the first n terms of

$$\begin{aligned} \delta(x, y)^{-\alpha} &+ \frac{\alpha|h''(x)|^2}{24}\delta(x, y)^{2-\alpha} + \frac{\alpha(h''(x), h^{(3)}(x))}{24}\delta(x, y)^{3-\alpha} \\ &+ \left\{ \frac{\alpha}{2} \left(\frac{(h''(x), h^{(4)}(x))}{40} + \frac{|h^{(3)}(x)|^2}{45} \right) + \frac{\alpha(\alpha+2)|h''(x)|^4}{1152} \right\} \delta(x, y)^{4-\alpha} \\ &+ \dots \end{aligned}$$

as the subtraction term, where

$$n = \begin{cases} 1 & (1 < \alpha < 3), \\ [\alpha - 1] & (\alpha \geq 3). \end{cases}$$

Therefore we can define a finite valued functional $E^{(\alpha)}$ according to the value of α as follows. When $1 < \alpha < 3$ define $E^{(\alpha)}$ by

$$E^{(\alpha)}(h) = \iint_{S^1 \times S^1} \left(\frac{1}{|h(x) - h(y)|^\alpha} - \frac{1}{\delta(x, y)^\alpha} \right) dx dy, \quad (2)$$

when $3 \leq \alpha < 4$ define $E^{(\alpha)}$ by

$$E^{(\alpha)}(h) = \iint_{S^1 \times S^1} \left(\frac{1}{|h(x) - h(y)|^\alpha} - \frac{1}{\delta(x, y)^\alpha} - \frac{\alpha|h''(x)|^2}{24\delta(x, y)^{\alpha-2}} \right) dx dy,$$

when $4 \leq \alpha < 5$ define $E^{(\alpha)}$ by

$$E^{(\alpha)}(h) = \iint_{S^1 \times S^1} \left(\frac{1}{|h(x) - h(y)|^\alpha} - \frac{1}{\delta(x, y)^\alpha} - \frac{\alpha|h''(x)|^2}{24\delta(x, y)^{\alpha-2}} - \frac{\alpha(h''(x), h^{(3)}(x))}{24\delta(x, y)^{\alpha-3}} \right) dx dy,$$

and so on. We call α the **power index**. Thus $E^{(\alpha)}$ can be interpreted as the *regularization* of the $r^{-(\alpha+1)}$ -modified electrostatic energy of charged knots. We remark that we have assumed higher differentiability of h here for the sake of simplicity. In fact $E^{(\alpha)}(h)$ ($1 \leq \alpha < 3$) can be defined if h is of class C^2 .

Numerical calculations of the values of $E^{(\alpha)}$ have been carried out. In connection with this, we give formulae for the asymptotic behavior of the energy of charged n -gons as n goes to infinity. Take n points equally scattered along a knot $h(S^1)$, and put a point charge $\frac{1}{n}$ upon each point. Then the $r^{-(\alpha+1)}$ -modified electrostatic energy of these n points is given by

$$\hat{E}_n^{(\alpha)}(h) = \frac{1}{n^2} \sum_{i \neq j} \frac{1}{|h(\frac{i}{n}) - h(\frac{j}{n})|^\alpha},$$

which blows up as n goes to infinity. Since $\hat{E}_n^{(\alpha)}(h)$ corresponds to the first term of the integrand in (2), the asymptotic behavior of $\hat{E}_n^{(\alpha)}(h)$ as n goes to infinity is given according to the value of α by the following.

When $\alpha = 1$

$$\hat{E}_n^{(1)}(h) \sim E^{(1)}(h) + 2(\log \frac{n}{2} + C),$$

where C is Euler's constant

$$C = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{2} + \dots + \frac{1}{n} - \log n \right),$$

which appears in these formulae only when α is an integer.

When $1 < \alpha < 3$

$$\hat{E}_n^{(\alpha)}(h) \sim E^{(\alpha)}(h) + 2 \left(\zeta(\alpha)n^{\alpha-1} - \frac{2^{\alpha-1}}{\alpha-1} \right), \tag{3}$$

when $\alpha = 3$

$$\hat{E}_n^{(3)}(h) \sim E^{(3)}(h) + 2(\zeta(3)n^2 - 2) + \frac{1}{4} \left(\int_0^1 |h''(x)|^2 dx \right) \left(\log \frac{n}{2} + C \right),$$

when $3 < \alpha < 4$

$$\begin{aligned} \hat{E}_n^{(\alpha)}(h) \sim E^{(\alpha)}(h) + 2 \left(\zeta(\alpha)n^{\alpha-1} - \frac{2^{\alpha-1}}{\alpha-1} \right) \\ + \frac{\alpha}{12} \left(\int_{S^1} |h''(x)|^2 dx \right) \left(\zeta(\alpha-2)n^{\alpha-3} - \frac{2^{\alpha-3}}{\alpha-3} \right), \end{aligned}$$

when $\alpha = 4$

$$\begin{aligned} \hat{E}_n^{(4)}(h) \sim E^{(4)}(h) + 2\left(\zeta(4)n^3 - \frac{8}{3}\right) + \frac{1}{3} \left(\int_{S^1} |h''(x)|^2 dx \right) (\zeta(2)n - 2) \\ + \frac{1}{3} \left(\int_{S^1} (h''(x), h^{(3)}(x)) dx \right) \left(\log \frac{n}{2} + C \right), \end{aligned}$$

and so on. Here \sim 's in the above formulae mean that the differences between the left hand sides and the right hand sides go to 0 as n goes to infinity.

Hereafter we consider $E^{(\alpha)}$ only when $\alpha < 3$ for the sake of simplicity. Instead of $E^{(\alpha)}$ ($\alpha \geq 3$) we will study $e^{\alpha,p}$ in §4. When $\alpha < 3$ $E^{(\alpha)}(h) > 0$ for any knot h since the integrand is non-negative and not identically zero. In this case the regularization is done by taking the difference of the extrinsic energy based on the distance measured in the ambient space and the intrinsic self-energy based on the distance measured in the knot itself.

Let us investigate when $E^{(\alpha)}$ becomes a knot energy functional. Suppose \bar{h} is a planar figure eight with a perpendicular double point P . Then the contribution of the neighborhood of P to the integral (2) is given by

$$\int_{-c}^c \int_{-c}^c \frac{dsdt}{(s^2 + t^2)^{\frac{\alpha}{2}}} \sim 4 \int_0^{c'} \int_0^{\frac{\pi}{2}} \frac{d\theta dr}{r^{\alpha-1}} \quad (4)$$

$$\begin{cases} < \infty & (\alpha < 2) \\ = \infty & (\alpha \geq 2). \end{cases}$$

Therefore $E^{(\alpha)}$ takes a finite value for a "singular knot" with a self-intersection and hence $E^{(\alpha)}$ is not a knot energy functional if $\alpha < 2$. On the other hand if $2 \leq \alpha < 3$ then for any $b \in \mathbf{R}$ there is a positive constant $C = C_\alpha(b)$ such that if $E^{(\alpha)}(h) \leq b$ then $|h(x) - h(y)| \geq C\delta(x, y)$ for any $x, y \in S^1$. This means

that $E^{(\alpha)}$ ($2 \leq \alpha < 3$) bounds Gromov's **distortion** ([26]) from above, where Gromov's distortion is given by

$$\text{Distor}(h) = \sup_{x \neq y} \frac{\delta(x, y)}{|h(x) - h(y)|}. \tag{5}$$

This estimate implies that h^{-1} satisfies Lipschitz condition. Since h is parametrized by arc-length it means that h is a bi-Lipschitz map.

Thus we have

Theorem 1 ([56,60]) $E^{(\alpha)}$ ($1 \leq \alpha < 3$) is a knot energy functional if and only if $\alpha \geq 2$.

We call $E^{(\alpha)}$ ($2 \leq \alpha < 3$) the α -**energy**. Let $E^{(2)}$, which is the simplest example, be denoted by E .

Example 2 Let $h_0(t) = (\frac{1}{2\pi} \cos 2\pi t, \frac{1}{2\pi} \sin 2\pi t, 0)$ be the standard planar circle. Then its energy is given by

$$\begin{aligned} E(h_0) &= \int_0^1 dx \int_0^{\frac{1}{2}} \left(\frac{\pi^2}{\sin^2 \pi y} - \frac{1}{y^2} \right) dy \\ &= 2 \left[-\pi \cot \pi y + \frac{1}{y} \right]_0^{\frac{1}{2}} \\ &= 4. \end{aligned}$$

This knot energy functional E was defined in [57] and is called the **Möbius energy** in [21] *et al.* In fact we have

$$\begin{aligned} E_{[57]}(h) &= \lim_{\epsilon \downarrow 0} \left\{ \frac{1}{2} \iint_{\epsilon \leq |x-y| \leq 1-\epsilon} \frac{dx dy}{|h(x) - h(y)|^2} - \frac{1}{\epsilon} \right\} \\ &= \frac{1}{2} \iint_{S^1 \times S^1} \left\{ \frac{1}{|h(x) - h(y)|^2} - \frac{1}{|h_0(x) - h_0(y)|^2} \right\} dx dy \\ &= \frac{1}{2} E(h) - 2. \end{aligned}$$

We end this section with the *cosine energy formula* of Doyle and Schramm ([2]). Kusner and J. Sullivan remark that this is useful in numerical experiments. Let $S_x(y)$ be the circle tangent to the knot $h(S^1)$ at $h(x)$ that passes through $h(y)$ and alternatively let $S_y(x)$ be the circle tangent to $h(S^1)$ at $h(y)$ that passes through $h(x)$. Let $\theta_h(x, y)$ be the angle between $S_x(y)$ and $S_y(x)$ at $h(x)$ or $h(y)$. Then

$$E(h) = \iint_{S^1 \times S^1} \frac{(1 - \cos \theta_h(x, y)) dx dy}{|h(x) - h(y)|^2} + 4 \tag{6}$$

3 Properties of $E = E^{(2)}$

3.1 Continuity

The knot energy functional E is continuous with respect to the C^2 -topology. Namely, for a given knot h and a positive constant ϵ , there are positive constants $\delta_0, \delta_1, \delta_2$ such that if a knot h_1 satisfies

$$\begin{aligned} |h(x) - h_1(x)| &< \delta_0, \\ |h'(x) - h_1'(x)| &< \delta_1, \\ |h''(x) - h_1''(x)| &< \delta_2 \end{aligned}$$

for all $x \in S^1$ then

$$E(h) - E(h_1) < \epsilon.$$

On the other hand E is not continuous with respect to the C^1 -topology. If a knot degenerates to an immersion which is piecewise of class C^1 , but not of class C^1 , then its energy blows up.

3.2 Möbius invariance

In order to state the theorem of the Möbius invariance of E , let us extend the domain of the knot energy functional. Let X be S^1, \mathbf{R} , or an interval in \mathbf{R} . Let $f : X \rightarrow \mathbf{R}^3$ be an embedding of class C^2 which is not necessarily parametrized by arc-length. By reparametrization and rescaling E can be written as

$$E(f) = \iint_{X \times X} \left(\frac{1}{|f(x) - f(y)|^2} - \frac{1}{\delta_f(f(x), f(y))^2} \right) |f'(x)| |f'(y)| dx dy, \quad (7)$$

where $\delta_f(f(x), f(y))$ is the shortest arc-length between $f(x)$ and $f(y)$ along the knot.

Freedman, He and Wang showed that E is Möbius invariant.

Theorem 3 ([21]) *Let $f : S^1 \rightarrow \mathbf{R}^3$ be a knot and let T be a Möbius transformation of $\mathbf{R}^3 \cup \{\infty\}$, i.e. a composition of reflections with respect to spheres in \mathbf{R}^3 .*

- (1) *If $T \circ f(S^1) \subset \mathbf{R}^3$ then $E(T \circ f) = E(f)$.*
- (2) *If $T \circ f(S^1)$ passes through ∞ then $E((T \circ f) \cap \mathbf{R}^3) = E(f) - 4$.*

Proof.^b (a) Put

$$|T'(P)| = |\det DT(P)|^{\frac{1}{6}}$$

^bAfter a suggestion of S. Nayatani.

for $P \in \mathbf{R}^3$. (When T is a Möbius transformation of $\mathbf{R}^n \cup \{\infty\}$ the argument is parallel if we put $|T'(P)| = |\det DT(P)|^{\frac{1}{2n}}$.) In particular, when T is a reflection with respect to a sphere of radius r with center 0, $|T'(P)| = \frac{r}{|P|}$. Then

$$|T(P) - T(Q)| = |T'(P)||T'(Q)||P - Q|$$

for $P, Q \in \mathbf{R}^3$ and

$$|(T \circ f)'(x)| = |T'(f(x))|^2 |f'(x)|$$

for $f: S^1$ or $\mathbf{R} \rightarrow \mathbf{R}^3$. Therefore

$$\frac{|(T \circ f)'(x)|| (T \circ f)'(y)|}{|T \circ f(x) - T \circ f(y)|^2} = \frac{|f'(x)||f'(y)|}{|f(x) - f(y)|^2}. \quad (8)$$

(b) Let h_0 be the standard planar circle. Then $E(h_0) = 4$ (Example 2). Therefore for a closed knot $f: S^1 \rightarrow \mathbf{R}^3$ we have

$$E(f) = \iint_{S^1 \times S^1} \left(\frac{|f'(x)||f'(y)|}{|f(x) - f(y)|^2} - \frac{1}{|h_0(x) - h_0(y)|^2} \right) dx dy + 4 \quad (9)$$

and for an open knot $g: \mathbf{R} \rightarrow \mathbf{R}^3$ we have

$$E(g) = \iint_{\mathbf{R} \times \mathbf{R}} \left(\frac{|g'(x)||g'(y)|}{|g(x) - g(y)|^2} - \frac{1}{|x - y|^2} \right) dx dy.$$

Put $\tau^{-1} = T_0 \circ h_0: (0, 1) \rightarrow \mathbf{R}$, where T_0 is a Möbius transformation that maps $h_0((0, 1))$ onto \mathbf{R} . Then τ is a diffeomorphism. For an open knot $k: (0, 1) \rightarrow \mathbf{R}^3$

$$\begin{aligned} E(k) &= \iint_{\mathbf{R} \times \mathbf{R}} \left(\frac{|(k \circ \tau)'(t)|| (k \circ \tau)'(s)|}{|k \circ \tau(t) - k \circ \tau(s)|^2} - \frac{1}{|t - s|^2} \right) dt ds \\ &= \iint_{(0,1) \times (0,1)} \left(\frac{|k'(x)||k'(y)|}{|k(x) - k(y)|^2} - \frac{|(T_0 \circ h_0)'(x)|| (T_0 \circ h_0)'(y)|}{|T_0 \circ h_0(x) - T_0 \circ h_0(y)|^2} \right) dx dy \\ &= \iint_{(0,1) \times (0,1)} \left(\frac{|k'(x)||k'(y)|}{|k(x) - k(y)|^2} - \frac{1}{|h_0(x) - h_0(y)|^2} \right) dx dy. \quad (10) \end{aligned}$$

(8) and (9) imply (1) of the theorem and (8), (9) and (10) imply (2) of the theorem if we put $T \circ f = k$. \square

There are two important corollaries of the Möbius invariance property.

Among all the open knots only the straight lines give the absolute minimum value of E , namely 0. Theorem 3 (2) implies that only the standard planar circle gives the absolute minimum value of E of closed knots, namely 4.

3.3 Existence of E -minimizers

In order to state the theorem about the existence of the E -minimizers let us recall the following definitions. A *composite knot* is a knot which is a connected sum of two non-trivial knots. A *prime knot* is a knot which is not a composite knot.

Theorem 4 ([21]) *There is an E -minimizer for any knot type of a prime knot.*

Sketch of Proof. Let $[f]$ be a knot type of a prime knot. Take a sequence of knots $\{f_1, f_2, \dots\}$ in $[f]$ such that

$$\lim_{n \rightarrow \infty} E(f_n) = E([f]) = \inf_{\bar{f} \in [f]} E(\bar{f}).$$

We can always choose a converging subsequence by suitable reparametrization, rescaling, and congruent transformation of \mathbf{R}^3 . Take the limit f_∞ of a converging subsequence. There is a possibility that f_∞ is the trivial knot because a tangle in f_n might shrink to a point as n goes to infinity. This is called a “**pull-tight**”. Make a new sequence $\{\bar{f}_1, \bar{f}_2, \dots\}$ in $[f]$ by applying Möbius transformations if necessary so that each \bar{f}_n is in a “relaxed position” and so that we do not have a pull-tight any more.

Take the limit \bar{f}_∞ of a converging subsequence. Then \bar{f}_∞ belongs to the same knot type $[f]$ and $E(f_\infty) = E([f])$, which means that \bar{f}_∞ is an E -minimizer. \square

Here are some remarks.

The number of E -minimizers for each non-trivial prime knot type is not finite. Let f_0 be an E -minimizer and T be any Möbius transformation. Then $T \circ f_0$, or its mirror image if necessary, belongs to the same knot type and takes the same value of E as f_0 , and therefore is again an E -minimizer of the same knot type.

It is not known whether there exist any critical points in the trivial knot type except for the standard planar circle. Up to now numerical experiments suggest that there are none ([46]). If this is correct, it would give an alternative proof of Hatcher’s results that the space of the trivial knots in S^3 deformation retracts onto the space of great circles in S^3 ([28]).

On the other hand, as is seen later, Kusner and J. Sullivan conjecture that there are unstable critical points in the (p, q) torus knot type if p or q is greater than 2.

Since E does not blow up even if we have a pull-tight, E is unable to evolve a small tangle. This is why a higher power index is used or the total squared curvature functional is added in numerical experiments as in [23], [27] and [72].

We can not apply the proof of Theorem 3 to composite knot types. Let $[f_1], [f_2]$ be knot types of non-trivial knots and let $[f_1 \# f_2]$ be the knot type of their connected sum. Take a sequence of knots $\{g_1, g_2, \dots\}$ in $[f_1 \# f_2]$ with $\lim_{n \rightarrow \infty} E(g_n) = E([f_1 \# f_2])$. In this case we can not prevent pull-tights of both components corresponding to $[f_1]$ and $[f_2]$ at the same time. When we evolve one component by a Möbius transformation, it might make the other component smaller. In fact Kusner and J. Sullivan observed that the both components shrink to points in numerical experiments ([41]). Consider the corresponding open knot by a reflection with respect to a sphere with center on the knot. Then the above phenomenon means that both components corresponding to $[f_1]$ and $[f_2]$ are getting far from each other. If there is enough distance the interaction of both components in the integral would converge to 0. Therefore the energy of the open knot corresponding to $[f_1 \# f_2]$ would converge to the sum of the energies of the open knots corresponding to $[f_1]$ and $[f_2]$. Applying a reflection with respect to a sphere again to make knots closed, one is lead to the following conjecture.

Conjecture 5 ([41]) (1) *There are no E -minimizers in composite knot types.*
 (2) *Let $[f_1], [f_2]$ be knot types. Then*

$$E([f_1 \# f_2]) = E([f_1]) + E([f_2]) - 4.$$

3.4 Gradient

The first variation or the gradient of E can not be expressed by local data of knots because E is defined as an integral of global quantities.

Let f be a knot of class C^4 . Let $P_{f(x)} : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ be the orthogonal projection onto the plane which is perpendicular to $f'(x)$:

$$P_{f(x)}(U) = U - \frac{(U, f'(x))}{|f'(x)|^2} f'(x).$$

Let $G_f : S^1 \rightarrow \mathbf{R}^3$ be given by

$$G_f(x) = 2 \int_{S^1} \left\{ 2 \frac{P_{f(x)}(f(y) - f(x))}{|f(y) - f(x)|^2} - \frac{1}{|f'(x)|} \frac{d}{dx} \left(\frac{f'(x)}{|f'(x)|} \right) \right\} \frac{|f'(y)|}{|f(y) - f(x)|^2} dy.$$

Then G_f is the L^2 -gradient of E ([21]), namely we have

$$\frac{d}{d\epsilon} E(f + \epsilon p) \Big|_{\epsilon=0} = \int_{S^1} (G_f(x), p(x)) |f'(x)| dx,$$

for a map $p : S^1 \rightarrow \mathbf{R}^3$ of class C^2 , where $(,)$ is the standard inner product of \mathbf{R}^3 .

3.5 Criterion for criticality

A necessary condition for a knot to be critical with respect to E is given in [61] as follows. Let h be a knot of class C^4 that is parametrized by arc-length. Put

$$P_h(t) = \int_0^t \left\{ \int_0^1 \left(2 \frac{h(y) - h(x)}{|h(y) - h(x)|^4} - \frac{h''(x)}{\delta(x, y)^2} \right) dy \right\} dx.$$

Then h is critical with respect to E if the following two conditions are satisfied:

- (1) There holds $\det(h'(0), h'(t), P_h(t)) = 0$ for any $t \in S^1$.
- (2) If $h'(t) = \pm h'(0)$ then $P_h(t) = Ch'(0)$ for some $C \in \mathbf{R}$.

3.6 Unstable E -critical torus knots

In this subsection we introduce Kim and Kusner's results on the energy of torus knots on Clifford tori in S^3 ([35]).

We can define the energy $E_{R^n}(f)$ of an embedding $f : S^1$ or $\mathbf{R} \rightarrow \mathbf{R}^n$ by the same formula (2). As is mentioned in the proof of Theorem 3, the Möbius invariance property also holds for E_{R^n} . Let σ be stereographic projection from $S^3 \subset \mathbf{R}^4$ onto $\mathbf{R}^3 \cup \{\infty\}$. Then σ can be extended to a Möbius transformation of $\mathbf{R}^4 \cup \{\infty\}$. Therefore if $\sigma \circ f(S^1) \subset \mathbf{R}^3$ for a knot f in S^3 then $E_{R^4}(f) = E_{R^4}(\sigma \circ f) = E(\sigma \circ f)$. Thus it is equivalent to consider $E_{R^4}|_{S^3}$ instead of E .

Let T_r be a Clifford torus of radius r ($0 < r < 1$) in $S^3 \subset \mathbf{C}^2$:

$$T_r = \{(z, w) \in S^3 \subset \mathbf{C}^2; |z| = r, |w| = \sqrt{1 - r^2}\},$$

and let $f_{r;p,q}$ be a (p, q) torus knot on T_r :

$$f_{r;p,q} : [0, 2\pi] \ni \theta \mapsto (re^{p\theta i}, \sqrt{1 - r^2} e^{q\theta i}) \in S^3 \subset \mathbf{C}^2. \tag{11}$$

Kusner and Stengle showed

$$E_{R^4}(f_{r;p,q}) = 4 + 4\pi^2 \sum_{|z| < 1} \text{Res} \frac{-\{r^2(1 - r^2)p^2 + q^2\}}{r^2(1 - r^2)z^{1-p}(z^p - 1)^2 + z^{1-q}(z^q - 1)^2}, \tag{12}$$

where the summation is taken over the residues inside the unit disk.

On the other hand S^1 acts isometrically on S^3 by

$$a : S^1 \ni e^{it} \mapsto \left(a_t : S^3 \ni \begin{pmatrix} z \\ w \end{pmatrix} \mapsto \begin{pmatrix} e^{ipt}z \\ e^{iqt}w \end{pmatrix} \in S^3 \right).$$

With regard to this action, $E_{R^4}|_{S^3}$ is invariant and $f_{r,p,q}$ is an orbit. Therefore Palais' principle of symmetric criticality ([64]) implies that if

$$\frac{\partial}{\partial r} E_{R^4}|_{S^3}(f_{r,p,q}) \Big|_{r=r_0} = 0 \tag{13}$$

then $f_{r_0,p,q}$ is critical for $E_{R^4}|_{S^3}$.

Since a (p, q) torus knot is prime Theorem 4 guarantees that there is an E -minimizer. Using numerical experiments Kim and Kusner conjectured that $\sigma \circ f_{r_0,p,q}$ where r_0 is obtained by (13) is a stable local minimum for E if and only if $p = 2$ or $q = 2$. On the other hand, when $(p, q) = (2, 3)$, i.e. in the trefoil case, formula (12) implies that $E_{R^4}|_{S^3}(f_{r;2,3})$ takes the minimum value approximately 74.41204 at $r_0 \approx 0.880$, which is close to the minimum energy of the trefoil obtained by numerical experiments.

3.7 Finiteness of knot types

Let $f : X \rightarrow \mathbf{R}^3$ be an embedding, where X is S^1 , \mathbf{R} or an interval in \mathbf{R} . Put

$$\phi_f : X \times X \setminus \Delta \ni (x, y) \mapsto \frac{f(x) - f(y)}{|f(x) - f(y)|} \in S^2.$$

Then the **average crossing number** ([FH]) of f is given by

$$\begin{aligned} \text{ac}(f) &= \frac{1}{\text{Vol}(S^2)} \iint_{X \times X} |\det(D\phi_f)| dx dy & (14) \\ &= \frac{1}{4\pi} \iint_{X \times X} \frac{|\langle f'(x), f'(y), f(x) - f(y) \rangle|}{|f(x) - f(y)|^3} dx dy, \end{aligned}$$

where $\langle f'(x), f'(y), f(x) - f(y) \rangle$ is the scalar triple product. Freedman, He and Wang showed that E bounds the average crossing number from above. Namely:

(1) For any $f : \mathbf{R} \rightarrow \mathbf{R}^3$ $\text{ac}(f) \leq \frac{1}{2\pi} E(f)$. Therefore if we denote the minimum crossing number of a knot type $[f]$ by $c([f])$ Theorem 3 implies

$$c([f]) \leq \frac{1}{2\pi} (E([f]) - 4).$$

Hence if $E(f) < 6\pi + 4 \approx 22.8$ then $f(S^1)$ is the unknot. It does not seem that this quantity $6\pi + 4$ is sharp. Numerical experiments implies that the minimum value of E of non-trivial knots is about 74 which is attained by a trefoil.

(2) For any $f : S^1 \rightarrow \mathbf{R}^3$ $\text{ac}(f) \leq \frac{11}{12\pi} E(f) + \frac{1}{\pi}$.

Since there are only finitely many knot types whose minimum crossing numbers are less than a given number n , only finitely many knot types can occur under any given energy threshold.

We would like to remark that if we change the integrand of (14) to be the Jacobian of ϕ_f , $\det(D\phi_f)$, instead of its absolute value, then the integral gives a summand of the **self linking number** ([9-11], [65]). It is defined on knots with non-vanishing curvature as

$$\text{sl}(f) = \frac{1}{4\pi} \iint_{S^1 \times S^1} \frac{\langle f'(x), f'(y), f(x) - f(y) \rangle}{|f(x) - f(y)|^3} dx dy + \frac{1}{2\pi} \int_{S^1} \tau(t) dt,$$

where $\tau(t)$ is the torsion

$$\tau(t) = \frac{\langle f'(t), f''(t), f'''(t) \rangle}{|f''(t)|^2}.$$

The self linking number is \mathbf{Z} -valued and is not a knot invariant.

4 Higher power index

As we saw in the last section there are E -minimizers in prime knot types (Theorem 4) but it is conjectured that there are no E -minimizers in composite knot types (Conjecture 5). There are potentially two ways to produce energy minimizers for each knot type. One is to make the power index of the energy greater than 2, and the other is to change the ambient space. In this section we consider the first approach. Since it does not seem easy to study $E^{(\alpha)}$ when $\alpha \geq 3$ we consider the following family of knot energy functionals instead. In this section we assume that a knot h is parametrized by arc-length.

Let $\alpha \geq 0$. Define $\psi_h^{(\alpha)} : S^1 \times S^1 \setminus \Delta \rightarrow \mathbf{R}$ by

$$\begin{aligned} \psi_h^{(\alpha)}(x, y) &= \int_{\delta(x,y)^{-1}}^{|h(x)-h(y)|^{-1}} \xi^{\alpha-1} d\xi \\ &= \begin{cases} \frac{1}{\alpha} \left(\frac{1}{|h(x) - h(y)|^\alpha} - \frac{1}{\delta(x, y)^\alpha} \right) & (\alpha \neq 0), \\ \log \frac{\delta(x, y)}{|h(x) - h(y)|} & (\alpha = 0). \end{cases} \end{aligned}$$

For $0 < p \leq \infty$ define $e^{\alpha,p}(h)$ by the L^p -norm of $\psi_h^{(\alpha)}$ ([58]):

$$e^{\alpha,p}(h) = \|\psi_h^{(\alpha)}\|_{L^p(S^1 \times S^1)}.$$

Hölder's inequality implies that if $1 \leq p_1 < p_2$ then $e^{\alpha,p_1}(h) \leq e^{\alpha,p_2}(h)$ for any α and h .

When $0 < \alpha, p < +\infty$ $e^{\alpha,p}(h)$ is given by

$$e^{\alpha,p}(h) = \frac{1}{\alpha} \left[\iint_{S^1 \times S^1} \left\{ \frac{1}{|h(x) - h(y)|^\alpha} - \frac{1}{\delta(x,y)^\alpha} \right\}^p dx dy \right]^{\frac{1}{p}}. \tag{15}$$

Therefore $E^{(\alpha)} = \alpha e^{\alpha,1}$ ($1 \leq \alpha < 3$) and in particular $E = 2e^{2,1}$.

When $(\alpha, p) = (0, \infty)$ we have

$$e^{0,\infty}(h) = \log(\text{Distor}(h)), \tag{16}$$

where $\text{Distor}(h)$ is Gromov's distortion given by (5).

Since (1) implies $\lim_{y \rightarrow x} \psi_h^{(2)}(x, y) = \frac{1}{24} |h''(x)|^2$, we have

$$e^{2,\infty}(h) \geq \frac{1}{24} \max_{x \in S^1} |h''(x)|^2.$$

We call the product αp the **power index** of $e^{\alpha,p}$. When $(\alpha, p) = (0, \infty)$ we define the power index αp to be 2. Since the integrand is non-negative and not identically zero, $e^{\alpha,p}(h) > 0$ for any knot h . We can define $e^{\alpha,p}$ similarly for an embedding from \mathbf{R} or an interval into \mathbf{R}^3 .

The formula (1) implies that the integrand of (15) is $O(|x - y|^{(2-\alpha)p})$ near the diagonal set Δ , hence $e^{\alpha,p}$ is well-defined if and only if $(2 - \alpha)p > -1$. The estimate (4) shows that $e^{\alpha,p}$ does not blow up even if a knot has a double point if $\alpha p < 2$. Just like in the case of $E^{(\alpha)}$, $e^{\alpha,p}$ ($\alpha p \geq 2$) bounds the distortion from above. Therefore we have

Theorem 6 ([60]) *$e^{\alpha,p}$ is well-defined if and only if α, p satisfies*

$$\alpha \leq 2 \quad \text{or} \quad p < \frac{1}{\alpha - 2} \quad (2 < \alpha < 4),$$

and is a knot energy functional if and only if the power index $\alpha p \geq 2$.

We call $e^{\alpha,p}$ the **(α, p) -energy functional for knots** if α and p satisfy the above two conditions.

The properties of $e^{\alpha,p}$ depend on whether the power index $\alpha p = 2$ or $\alpha p > 2$. The quantity $e^{\alpha,p}$ ($\alpha p = 2$) bounds a Lipschitz constant of h^{-1} whereas $e^{\alpha,p}$ ($\alpha p > 2$) controls the behavior of h' . There exists a positive constant $A = A(\alpha, p)$ such that for any $b \in \mathbf{R}$ if $e^{\alpha,p}(h) \leq b$ then h' satisfies

$$|h'(x) - h'(y)| \leq \angle h'(x) \cdot h'(y) \leq Ab^{\frac{2}{2(p+2)}} \delta(x, y)^{\frac{\alpha p - 2}{2(p+2)}} \tag{17}$$

for all $x, y \in S^1$, where $\angle h'(x) \cdot h'(y)$ is the angle between $h'(x)$ and $h'(y)$ ([60]). In other words if $e^{\alpha p}(h)$ with $\alpha p > 2$ is finite then h can not have a sudden turn.

This inequality shows that $e^{\alpha p}$ ($\alpha p > 2$) blows up if a knot has a pull-tight, which implies that $e^{\alpha p}$ is not Möbius invariant if $\alpha p > 2$.

The behavior of $e^{\alpha p}$ when a knot has a pull-tight is given as follows.

Proposition 7 ([58]) *Consider a line segment. Put a small 1-tangle T in the center of the line segment to make an open knot. Shrink the tangle T to a point while keeping its shape similar. Under this pull-tight process the value of $e^{\alpha p}$ of the open knot behaves according to the power index αp as follows:*

- (1) *If $\alpha p > 2$ then the value of $e^{\alpha p}$ blows up.*
- (2) *If $\alpha p = 2$ then the value of $e^{\alpha p}$ converges to a positive constant which depends on the shape of the tangle T .*
- (3) *If $\alpha p < 2$ then the value of $e^{\alpha p}$ converges to 0 which is equal to the value of $e^{\alpha p}$ of the straight line segment. Note that in this case $e^{\alpha p}$ is not a knot energy functional.*

Let us give a brief explanation. Consider the behavior of the contribution of the tangle T to the integral (15) under the pull-tight process. Suppose the tangle T shrinks to half its size. Then the integrand becomes $2^{\alpha p}$ times larger on $T \times T$, whereas the area of $T \times T$ becomes 2^{-2} times smaller. Thus the effect of pull-tight vanishes when $\alpha p = 2$.

The inequality (17) implies

Theorem 8 ([60]) *Suppose $\alpha p > 2$. Then for any $b \in \mathbf{R}$ there is a set of finitely many solid tori $T_i \subset \mathbf{R}^3$ which satisfies the following two conditions:*

- (1) *If $e^{\alpha p}(h) \leq b$ then $h(S^1)$ is included in some T_i after an orientation preserving congruent translation of \mathbf{R}^3 .*
- (2) *If $e^{\alpha p}(h) \leq b$ and $h(S^1) \subset T_i$ then $h(S^1)$ intersects each meridian disk of T_i transversely at one point.*

As a corollary only finitely many knot types can occur under any given threshold of $e^{\alpha p}$ ($\alpha p > 2$). On the other hand, when $(\alpha, p) = (0, \infty)$ there is a $b \in \mathbf{R}$ such that there are infinitely many knot types that satisfy $e^{0, \infty} = \log(\text{Distor}) \leq b$ ([25]).

Theorem 8 implies the existence of energy minimizers. Let $[h]$ be any knot type. Take a sequence $\{h_1, h_2, \dots\}$ in $[h]$ such that

$$\lim_{n \rightarrow \infty} e^{\alpha p}(h_n) = e^{\alpha p}([h]) = \inf_{\bar{h} \in [h]} e^{\alpha p}(\bar{h}).$$

Take a converging subsequence $\{h_{n_i}\}$. The above theorem ensures that there is a solid torus T_j such that infinitely many h_{n_i} 's are included in T_j in a good manner. Let h_∞ be their limit. Then h_∞ is also contained in T_j in a good

manner and hence belongs to the same knot type $[h]$. Therefore h_∞ is an $e^{\alpha,p}$ -minimizer for $[h]$. Thus we have

Corollary 9 ([60]) *There is an $e^{\alpha,p}$ -minimizer for any knot type if $\alpha p > 2$. We conjecture that the number of $e^{\alpha,p}$ -minimizers ($\alpha p > 2$) for each knot type is finite. We remark that the number of the solid tori associated to $e^{\alpha,p}$ -minimizers ($\alpha p > 2$) in the sense of Theorem 8 is finite.*

As an another corollary of Theorem 8, only finitely many knot types can occur under any given threshold of $e^{\alpha,p}$ ($\alpha p > 2$). On the other hand, when $(\alpha, p) = (0, \infty)$ there is a $b \in \mathbf{R}$ such that there are infinitely many knot types that satisfy $e^{0,\infty} = \log(\text{Distor}) \leq b$ ([25]).

For $0 < \alpha \leq 2$ define the α -energy polynomial $E_h^{(\alpha)}(t)$ of a knot h by an exponential generating function of $(e^{\alpha,n}(h))^n$:

$$\begin{aligned} E_h^{(\alpha)}(t) &= \iint_{S^1 \times S^1} \exp(t\psi_h^{(\alpha)}(x, y)) dx dy \\ &= \iint_{S^1 \times S^1} \exp\left\{ \frac{t}{\alpha} \left(\frac{1}{|h(x) - h(y)|^\alpha} - \frac{1}{\delta(x, y)^\alpha} \right) \right\} dx dy \\ &= 1 + e^{\alpha,1}(h)t + \frac{(e^{\alpha,2}(h))^2}{2!}t^2 + \frac{(e^{\alpha,3}(h))^3}{3!}t^3 + \dots \end{aligned}$$

Then $E_h^{(\alpha)}(t)$ is a well-defined knot energy functional for any $t > 0$.

5 Energy of knots in a Riemannian manifold

An alternative approach to produce energy minimizers for all the knot types is to change the ambient space.

5.1 General cases

Let $f : S^1 \rightarrow M$ be a smooth embedding into a Riemannian manifold $M = (M, g)$. Let $d_M(f(x), f(y))$ be the infimum of the lengths of the paths which join $f(x)$ and $f(y)$, and let $\delta_f(f(x), f(y))$ be the shortest arc-length between the two points. Define

$$e_M^{\alpha,p}(f) = \frac{1}{\alpha} \left[\iint_{S^1 \times S^1} \left\{ \frac{1}{d_M(f(x), f(y))^\alpha} - \frac{1}{\delta_f(f(x), f(y))^\alpha} \right\}^p |f'(x)||f'(y)| dx dy \right]^{\frac{1}{p}} \tag{18}$$

for

$$p \geq \frac{2}{\alpha} \quad (0 < \alpha \leq 2) \quad \text{or} \quad \frac{1}{\alpha - 2} > p \geq \frac{2}{\alpha} \quad (2 < \alpha < 4).$$

We call αp the power index of $e_M^{\alpha,p}$. If $p = 1$ we write $E_M^{(\alpha)} = \alpha e_M^{\alpha,1}$ and if $\alpha = 2$ in addition we write $E_M = E_M^{(2)}$.

The well-definedness of $e_M^{\alpha,p}$ is a consequence of Nash's theorem which guarantees that any manifold M can be embedded isometrically into the Euclidian space \mathbf{R}^N for sufficiently large N ([54]).

Note that $e_M^{\alpha,p}(f) \geq 0$ for any knot f since the integrand is non-negative. The Jacobian term $|f'(x)||f'(y)|$ makes $e_M^{\alpha,p}$ invariant under reparametrization.

Just like in the Euclidian case $e_M^{\alpha,p}$ is a knot energy functional. More precisely, for any $b \in \mathbf{R}$ there is a positive constant $C = C(\alpha, b, l)$ such that if a knot f of length l satisfies $e_M^{\alpha,p}(f) \leq b$ then $d_M(f(x), f(y)) \geq C\delta_f(f(x), f(y))$ for all $x, y \in S^1$.

We call $e_M^{\alpha,p}$ the **unit density (α, p) -energy functional for knots in M** . When $p = 1$ $E_M^{(\alpha)} = \alpha e_M^{\alpha,1}$ can be interpreted as the regularization of modified electrostatic energy of charged knots under the assumption that the electric density is constantly equal to 1. In this case the total quantity of electric charge is equal to the length l_f of a knot $f(S^1)$.

Since there are no standard homotheties in M we can not normalize a knot so that its length is 1. This allows us two other ways of generalizing (15) as is explained in [62]. One is scale invariant when $M = \mathbf{R}^3$ and the other is defined under the assumption that the total quantity of electric charge is equal to 1 when $p = 1$. These three definitions differ from each other only by a power of l_f . We consider the unit density energy here because it seems most suitable to produce energy minimizers.

Let M be a compact manifold and let $b \in \mathbf{R}$. Then there is a positive constant $l_1 = l_1(M, \alpha, p; b)$ such that if $e_M^{\alpha,p}(f) \leq b$ then the length l_f of the knot $f(S^1)$ satisfies $l_f \leq l_1$. If $\alpha p > 2$ there is a positive constant $l_2 = l_2(M, \alpha, p; b)$ such that if $e_M^{\alpha,p}(f) \leq b$ then $l_f \geq l_2$. Therefore neither a tangle nor a knot itself shrinks to a point in this case, which allows us to use the same argument as in the Euclidian case (Corollary 9) to obtain

Theorem 10 *Let M be a compact manifold. Then there is an $e_M^{\alpha,p}$ -minimizer for any knot type if the power index $\alpha p > 2$.*

5.2 The spherical case

On the other hand when the power index $\alpha p = 2$ we conjecture that it depends on the metric of the ambient space whether there are energy minimizers or not. To see this, let us consider E_M when the ambient space M is the 3-dimensional sphere $S^3 \subset \mathbf{R}^4$ or the 3-dimensional hyperbolic space \mathbf{H}^3 , which are manifolds with constant curvature ± 1 . Note that E_{S^3} and $E_{\mathbf{R}^4}|_{S^3}$ are different. Since $d_{S^3} > d_{\mathbf{R}^4}$ we always have $E_{S^3} < E_{\mathbf{R}^4}|_{S^3}$. Recall that in the Euclidian case

there are infinitely many E_{R^3} -minimizers in prime knot types whereas it is conjectured that there are no E_{R^3} -minimizers in composite knot types.

Let S_r be a circle in $S^3 \subset \mathbf{R}^4$ with radius r ($0 < r \leq 1$). Then $E_{S^3}(S_r)$ is a decreasing function of r with

$$\begin{aligned} E_{S^3}(S_1) &= 0, \\ \lim_{r \downarrow 0} E_{S^3}(S_r) &= E_{R^3}(h_0) = 4. \end{aligned}$$

This implies that E_{S^3} is not Möbius invariant.

Let τ_r denote the trefoil on the Clifford torus $T_r \subset S^3 \subset C^2$ of radius r ($0 < r < 1$), which was given as $f_{r;2,3}$ in (11). Then

$$E_{S^3}(\tau_r) = 2\pi \int_{-\pi}^{\pi} \left\{ \frac{9 - 5r^2}{4 \arcsin^2 \sqrt{r^2 \sin^2 x + (1 - r^2) \sin^2 \frac{3x}{2}}} - \frac{1}{x^2} \right\} dx.$$

Numerical experiment by Ligocki [47] implies that $E_{S^3}(\tau_r)$ is concave with respect to r and takes the minimum value approximately 54.3263 at $r \approx 0.861388$. On the other hand the spherical energy of a trefoil with a small “pulled-tight” tangle is greater than $E_{R^3}([\text{Trefoil}]) - 4 \approx 70$.

From these two examples we conjecture that the value of E_{S^3} increases if a tangle or a knot itself shrinks to a point. This can be explained roughly as follows. Suppose a tangle or a knot itself shrinks to a point by a Möbius transformation. Then $E_{R^4}|_{S^3}$ is invariant. Since $d_{S^3} > d_{R^4}$ we always have $E_{S^3} < E_{R^4}|_{S^3}$. Since d_{S^3}/d_{R^4} converges to 1 as d_{R^4} goes down to 0, the contribution of the shrinking part to E_{S^3} converges to that to $E_{R^4}|_{S^3}$, which makes E_{S^3} greater. Thus we are lead to

Conjecture 11 *There are (hopefully finitely many) E_{S^3} -minimizers for any knot type.*

5.3 The hyperbolic case

On the contrary we have an opposite conjecture in the hyperbolic case. Let $\{S_r\}$ be a family of circles of radius r in hyperbolic space. Then $E_{H^3}(S_r)$ is an increasing function of r with

$$\begin{aligned} \lim_{r \downarrow 0} E_{H^3}(S_r) &= E_{R^3}(h_0) = 4, \\ \lim_{r \rightarrow \infty} E_{H^3}(S_r) &= \infty. \end{aligned}$$

Thus we are lead to

Conjecture 12 *There are no E_{H^3} -minimizers in any knot type. Any knot will shrink to a point if it evolves itself so as to decrease its value of E_{H^3} .*

On the other hand when the power index $\alpha p > 2$ we have

$$\lim_{r \downarrow 0} e_{H^3}^{\alpha,p}(S_r) = \lim_{r \rightarrow \infty} e_{H^3}^{\alpha,p}(S_r) = \infty.$$

In fact if $\alpha p > 2$ then for any $b \in \mathbf{R}$ there is a positive constant $l_3 = l_3(\alpha, p; b)$ such that if $e_{H^3}^{\alpha,p}(f) \leq b$ then $l_f \leq l_3$. Therefore there is a compact subset $K_b \subset \mathbf{H}^3$ such that if $e_{H^3}^{\alpha,p}(f) \leq b$ then $f(S^1)$ is included in K_b after an isometric transformation of \mathbf{H}^3 . Then Theorem 5.1 implies

Theorem 13 *There is an $e_{H^3}^{\alpha,p}$ -minimizer for any knot type if the power index $\alpha p > 2$.*

6 Thickness and self distance

There are two more kinds of knot energy functionals, which are not derived from electrostatic energy. One is the “thickness” studied in [75], [30], and [31], which indicates how much a knot can be fattened, and the other is “self distance” which indicates “how close a knot approaches itself”. In this section we assume that a knot $h : S^1 = \mathbf{R}/\mathbf{Z} \rightarrow \mathbf{R}^3$ is parametrized by arc-length.

The **normal embedding radius** or Simon’s **thickness** ([75]) of a knot h , $\text{ner}(h)$, is given by

$$\text{ner}(h) = \sup\{r > 0 \mid D_{h(x)}(r) \cap D_{h(y)}(r) = \emptyset \ (x \neq y)\},$$

where $D_{h(x)}(r)$ is the normal disk of $h(S^1)$ of radius r at $h(x)$. The normal embedding radius is locally controlled by the curvature and globally by the *doubly critical self distance* which will be given below ([75]).

The **inflation radius** $\text{ir}(h)$ of a knot h is defined by

$$\text{ir}(h) = \sup\{\epsilon_0 \mid N_{\epsilon_0}(h(S^1)) \text{ is homeomorphic to } D^2 \times S^1 \ (0 < \forall \epsilon < \epsilon_0)\},$$

where $N_{\epsilon}(h(S^1))$ is the ϵ -neighborhood of $h(S^1)$.

Clearly $\text{ner}(h) \leq \text{ir}(h)$.

Put

$$\Lambda_h : S^1 \times S^1 \setminus \Delta \ni (x, y) \mapsto |h(x) - h(y)| \in \mathbf{R},$$

where Δ is the diagonal set.

Define the **minimum distance** $\text{md}(h)$ of a knot h to be the smallest (possibly degenerate) local minimum value of Λ_h if one exists, otherwise put $\text{md}(h) = \max_{x \neq y} \Lambda_h(x, y)$.

Define the **doubly critical self distance** $\text{dcsd}(h)$ of a knot h ([75]) to be the smallest critical value of Λ_h , namely

$$\text{dcsd}(h) = \min_{x \neq y} \{ |h(x) - h(y)| : h'(x) \perp (h(x) - h(y)), h'(y) \perp (h(x) - h(y)) \}$$

Put

$$\Lambda_{h,x} = \Lambda_h(x, \cdot) : S^1 \setminus \{x\} \ni y \mapsto |h(x) - h(y)| \in \mathbf{R}.$$

Let $\chi_h(x) > 0$ be the smallest critical value of $\Lambda_{h,x}$. Then Kuiper's **self distance** $\text{sd}(h)$ of a knot h ([39]) is defined by

$$\begin{aligned} \text{sd}(h) &= \inf_{x \in S^1} \chi_h(x) \\ &= \inf_{x \neq y} \{ |h(x) - h(y)| : h'(x) \perp (h(x) - h(y)) \}. \end{aligned}$$

The self distance and the doubly critical self distance are not stable under slight perturbations. For example $\text{sd}(h)$ might be attained by an inessential critical pair of points which will disappear by a slight perturbation. Let

$$N_\epsilon^{(1)}(h) = \{ g : S^1 \rightarrow \mathbf{R}^3 : |h(x) - g(x)| < \epsilon, |h'(x) - g'(x)| < \epsilon \quad (\forall x \in S^1) \}$$

be the ϵ -neighborhood of h with respect to the C^1 -topology. Define the **essential self distance** $\text{ess.sd}(h)$ and **essential doubly critical self distance** $\text{ess.dcsd}(h)$ of a knot h by

$$\begin{aligned} \text{ess.sd}(h) &= \inf_{\epsilon > 0} \sup_{g \in N_\epsilon^{(1)}(h)} \text{sd}(g), \\ \text{ess.dcsd}(h) &= \inf_{\epsilon > 0} \sup_{g \in N_\epsilon^{(1)}(h)} \text{dcsd}(g). \end{aligned}$$

Clearly $\text{ess.sd}(h) \geq \text{sd}(h)$ and $\text{ess.dcsd}(h) \geq \text{dcsd}(h)$.

Let $\beta_h(x)$ be the smallest (possibly degenerate) local minimum value of $\Lambda_{h,x}$ if one exists, otherwise put $\beta_h(x) = \max_{y \in S^1} \Lambda_{h,x}(y)$. Then the **radius of monotonicity** or the **beads radius** of a knot h , $\text{rm}(h)$, is defined by

$$\begin{aligned} \text{rm}(h) &= \inf_{x \in S^1} \beta_h(x) \\ &= \sup \left\{ r_0 > 0 \mid \begin{array}{l} B_{h(x)}(r) \cap h(S^1) \text{ is connected } (0 < \forall r < r_0, \forall x) \\ h(S^1) \text{ is not included in } B_{h(x)}(r_0) \quad (\forall x) \end{array} \right\}, \end{aligned}$$

where $B_P(r) \subset \mathbf{R}^3$ denotes the 3-ball of center P and radius $r > 0$.

If h is a non-trivial knot then $\Lambda_{h,x}$ has a (possibly degenerate) local minimum for any $x \in S^1$ because otherwise $h(S^1)$ bounds a disk. In this case the

last condition “ $h(S^1)$ is not included in $B_{h(x)}(r_0)$ ($\forall x$)” in the right hand side is not necessary.

By definition we have

$$\begin{aligned} \text{sd}(h) &\leq \text{dcsd}(h) \leq \text{md}(h), \\ \text{sd}(h) &\leq \text{rm}(h) \leq \text{md}(h), \\ \text{rm}(h) &\leq 2 \text{ir}(h). \end{aligned}$$

The relation between these quantities and $e^{\alpha,p}$ is as follows. If $\alpha p > 2$ then there is a positive constant $A' = A'(\alpha, p)$ such that

$$\text{sd}(h) \geq A' e^{\alpha,p}(h)^{-\frac{p}{\alpha p - 2}} \quad (19)$$

for any knot h ([60]). Since

$$\text{sd}(h) \leq \min\{\text{ess.sd}(h), \text{dcsd}(h), \text{ess.dcsd}(h), \text{md}(h), \text{rm}(h), 2\text{ir}(h)\}$$

this means that $e^{\alpha,p}$ ($\alpha p > 2$) bounds these quantities except for the normal embedding radius. In general $e^{\alpha,p}$ ($\alpha p > 2$) does not bound the normal embedding radius because $e^{\alpha,p}$ ($\alpha p > 2$) does not bound the curvature $|h''|$ unless $(\alpha, p) = (2, \infty)$. On the other hand Proposition 7 implies that $e^{\alpha,p}$ ($\alpha p = 2$) does not bound these quantities. (19) means that $e^{\alpha,p}$ ($\alpha p > 2$) bounds sd^{-1} from above. As we saw in Theorem 6, $e^{\alpha,p}$ ($\alpha p = 2$) bounds the distortion from above. We conjecture that sd^{-1} bounds the distortion from above.

7 Remarks

7.1 Computer experiments

Computer algorithms aimed at finding ideal configurations of knots will be found elsewhere in this book. For numerical experiments the reader is referred to such articles as [1], [6], [8], [22,23], [27], [41], [46], [73,74], [77], and home pages such as [24], [29], [72], and [81]. See also the remarks after Theorem 4.

7.2 Other kinds of knot energy functionals

As we mentioned in section 2, we can define other kinds of knot energy functional by multiplying by a term which goes to zero to the same order as $|h(x) - h(y)|^2$ at the diagonal set. Let $\theta_x(y)$ be the angle between $h(y) - h(x)$ and $h'(x)$. The **normal projection energy** of Buck and Orloff ([7]) is given by

$$E_{np}(h) = \iint_{S^1 \times S^1} \frac{\sin^2 \theta_x(y)}{|h(x) - h(y)|^2} dx dy.$$

The symmetric normal projection energy of Buck is given by

$$E_{snp}(h) = \iint_{S^1 \times S^1} \frac{|\sin \theta_x(y) \sin \theta_y(x)|}{|h(x) - h(y)|^2} dx dy.$$

7.3 Generalizations of E

Several kinds of generalizations of E which keep the Möbius invariance property have been carried out. An energy functional for links is studied in [41] and [42]. Lin defined "X-energy" for knots ([48]). Generalizations to higher dimensions are studied in [2] and [41].

7.4 Kirchhoff elastica

In order to define the ideal configuration of a knot one might ask what actually happens if one forms a knot in a piece of elastic wire. According to the Bernoulli-Euler theory the bending energy of an elastic rod is proportional to the total squared curvature $\kappa^{(2)}$. A Critical point of $\kappa^{(2)}$ is called an **elastica**. Langer and Singer showed that ([43-45])

(1) If f is a closed planar elastica then f is equal to the circle or a figure eight which is unique up to similarity or a multiple cover of one of these two.

(2) The knot types of non-planar closed elasticae are (p, q) -torus knots with $p > 2q$.

(3) The only stable closed elastica in \mathbf{R}^3 is the singly-covered circle.

T. Kawakubo takes the twisting energy into consideration ([33]). He studies a functional on the space of *framed knots*, where the framing is not necessarily \mathbf{Z} -valued.

7.5 Random knotting

A *random knotting probability* $P_K(N)$ is the probability that a Gaussian random polygon of N steps belongs to a knot type K . Deguchi and Tsurusaki proposed an asymptotic formula

$$P_K(N) \sim C(K)N^{\nu(K)} \exp(-N/N(K)) \quad (N \gg 1).$$

They evaluated $P_K(N)$ and explicitly obtained optimum values of the parameters for curves fitting the numerical data ([13,14]).

8 Summary

Here is a table of the answers and conjectures to our motivational problems:

Table 1: Are there energy minimizers, and if so, how many?

| Energy | $M = S^3$ | $M = \mathbf{R}^3$ | $M = \mathbf{H}^3$ |
|----------------------|----------------------------|--|--------------------|
| $E_M = 2e_M^{2,1}$ | Conj. YES, $\# < \infty$ | Prime knots : YES, $\# = \infty$ | Conj. NO |
| | | Composite knots : Conj. NO | |
| $e_M^{j,p} (jp > 2)$ | YES Conj. $\# < \infty$ | YES after rescaling Conj. $\# < \infty$ | YES |

Conj. means Conjecture.

$\#$ denotes the number of the minimizers for each knot type.

Q: Can we define ideal configurations for knots as energy minimizers?

Q: If so, is the number of energy minimizers in each knot type finite?

Acknowledgments

The author is deeply grateful to his colleague Prof. M. Guest for helpful suggestions.

References

1. K. Ahara, computer program and video (1990)
2. D. Auckly and L. Sadun, Geometric Topology (Proceedings of the 1993 Georgia International Topology Conference, AMS/IP Studies in Adv. Math., W. H. Kazez ed., Amer. Math. Soc. and International Press, Cambridge, MA. (1996)
3. R. Bott and C. Taubes, J. Math. Phys. **35**, 5247 (1994)
4. K. Brakke, Exper. Math. **1**, 141 (1992)

5. G. Buck, *J. Knot Theory Ramifications* **3**, 355 (1994)
6. G. Buck and J. Orloff, *Topology Appl.* **247**, 51 (1993)
7. G. Buck and J. Orloff, *Topology Appl.* **205**, 61 (1995)
8. G. Buck and J. Simon, *Topology Appl.* **51**, 229 (1993)
9. G. Călugăreanu, *Rev. Math. Pures Appl.* **4**, 5 (1959)
10. G. Călugăreanu, *Czechoslovak Math. J.* **11**, 588 (1961)
11. G. Călugăreanu, *Comm. Acad. R. P. Romine*, 829 (1961)
12. A.Y.K. Chui and H.K. Moffatt, *Proc. R. Soc. Lond. A* **451**, 609 (1995)
13. T. Deguchi and K. Tsurusaki, *Geometry and Physics, Lect. Notes in Pure and Applied Math. Ser. 184*, J.E. Andersen, J. Dupont, H. Pedersen, and A. Swann eds., Marcel Dekker, Basel Switzerland, 557 (1997)
14. T. Deguchi and K. Tsurusaki, *Lectures at Knots '96*, S. Suzuki ed., World Scientific, Singapore, 95 (1997)
15. J. des Cloizeaux and R. Ball, *Comm. Math. Phys.* **80**, 543 (1981)
16. Y. Diao, C. Ernst, and E. Janse van Rensburg, *J. Knot Theory Ramifications* **6**, 633 (1997)
17. B. Duplantier, *C. R. Acad. Sci. Math. Paris t. 293 Série I*, 693 (1981)
18. B. Duplantier, *Comm. Math. Phys.* **82**, 41 (1981)
19. B. Duplantier, *Comm. Math. Phys.* **85**, 221 (1982)
20. M. H. Freedman and Z.-X. He, *Ann. of Math.* **134**, 189 (1991)
21. M. H. Freedman, Z.-X. He, and Z. Wang, *Ann. of Math.* **139**, 1 (1994)
22. S. Fukuhara, *A Fête of Topology*, Y. Matsumoto, T. Mizutani, and S. Morita, eds., Academic Press, 443 (1987)
23. S. Fukuhara, computer program
24. <http://www.gang.umass.edu/docs/archive/knots.html>
25. M. Gromov, *J. Differential Geom.* **13**, 303 (1978)
26. M. Gromov, *J. Differential Geom.* **18**, 1 (1983)
27. C. Gunn, computer program : Linkmover, Geometry Supercomputer Project (1988)
28. A. Hatcher, *Ann. of Math.* **117**, 553 (1983)
29. M. Huang, <http://www.eecs.uic.edu/mhuang/>
30. V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet, and A. Stasiak, *Nature* **384**, 142 (1996)
31. V. Katritch, W. K. Olson, P. Pieranski, J. Dubochet, and A. Stasiak, *Nature* **388**, 148 (1997)
32. L. Kauffman, *J. Math. Phys.* **36**, 2402 (1995)
33. T. Kawakubo, a talk at the Japan Math. Soc. Meeting (1996)
34. N. Kawazumi, *J. Fac. Sci. Univ. Tokyo IA* **37**, 263 (1990)
35. D. Kim and R. Kusner, *Exper. Math.* **2**, 1 (1993)
36. N. Koiso, *Osaka J. Math.* **29**, 539 (1992)

37. N. Koiso, *Osaka J. Math.* **30**, 559 (1993)
38. N. Koiso, *Geometry and Global Analysis (Sendai)* (1993)
39. N. H. Kuiper, private communication
40. R. Kusner, P. Norman, and J. Sullivan, *GANG/UMass Amherst preprint III.18* (1993)
41. R. B. Kusner and J. M. Sullivan, *Geometric Topology, Proceedings of the 1993 Georgia International Topology Conference, AMS/IP Studies in Adv. Math.*, W. H. Kazez ed., Amer. Math. Soc. and International Press, Cambridge, MA., 570 (1996)
42. R. B. Kusner, J. M. Sullivan, J. Lawrence, and N. Schmitt, a talk in the *AMS Meeting in Iowa* (1996)
43. J. Langer and D. A. Singer, *J. London Math. Soc.* **30**, 512 (1984)
44. J. Langer and D. A. Singer, *J. Differential Geom.* **20**, 1 (1984)
45. J. Langer and D. A. Singer, *Topology* **24**, 75 (1985)
46. T. J. Ligocki and J. A. Sethian, *J. Knot Theory Ramifications* **3**, 477 (1994)
47. T. J. Ligocki, private communication
48. X.-S. Lin, preprint
49. X.-S. Lin and Z. Wang, *J. Diff. Geom.* **44**, 74 (1996)
50. S. Lomonaco, *The Interface of Knots and Physics*, L. Kauffman, ed., *Proc. Symp. Appl. Math.* vol. 51, Amer. Math. Soc., Providence (1996)
51. K. Millett, *J. Knot Theory Ramifications* **3**, 263 (1994)
52. H. K. Moffatt, *Nature* **347**, 367 (1990)
53. N. Nakauchi, *Proc. Amer. Math. Soc.* **118**, 293 (1993)
54. J. Nash, *Ann. of Math.* **63**, 20 (1956)
55. A. Nobutovsky, *Comm. Pure Appl. Math.* **48**, 147 (1995) 147-161
56. J. O'Hara, Master Thesis, Univ. of Tokyo (1988)
57. J. O'Hara, *Topology* **30**, 241 (1991)
58. J. O'Hara, *Topology Appl.* **48**, 147 (1992)
59. J. O'Hara, *Topology-Hawaii*, K. H. Dovermann ed., World Scientific, Singapore, 201 (1992)
60. J. O'Hara, *Topology Appl.* **56**, 45 (1994)
61. J. O'Hara, *MSRI Preprint 042* (1994)
62. J. O'Hara, *Proceedings of Knots '96*, S. Suzuki ed., World Scientific, Singapore, 449 (1997)
63. J. O'Hara, *Suugaku* **49-4**, 131 (1997)49-4
64. R. Palais, *Comm. Math. Physics* **69**, 19 (1979)
65. W. F. Pohl, *J. of Math. Mech.* **17**, 975 (1968)
66. W. F. Pohl, *Amer. J. Math.* **90**, 1321 (1968)

67. W. F. Pohl, *Lecture Notes in Math.* 894, International Symposium in honour of N. H. Kuiper, Utrecht 1980, Springer, 113
68. R. Randell, *J Knot Theory Ramifications* **3**, 279 (1994)
69. D. Rolfsen, *Knots and Link*, Publish or Perish (1976)
70. M. Sakuma, preprint
71. M. Sakuma, *Problems in low dimensional topology and related topics*, S. Kojima and S. Negami eds., in Japanese, 7 (1987)
72. R. Scharein, <http://www.cs.ubc.ca/spider/scharein/>
73. J. K. Simon, *J. Knot Theory Ramifications* **3**, 299 (1994)
74. J. K. Simon, *Mathematical approaches to biomolecular structure and dynamics*, IMA Vol. Math. Appl. Vol. 82, J.P. Mesirov, K. Schulten, and D.W. Sumners Eds., Springer-Verlag, NY (1996)
75. J. K. Simon, a talk at KNOTS'96, Waseda University, Tokyo (1996)
76. A. Stasiak, V. Katritch, J. Bednar, D. Michoud, and J. Dubochet, *Nature* **384**, 122 (1996)
77. J. M. Sullivan, VHS video tape, available from the Geometry Center, Minneapolis (1993)
78. J. H. White, *Amer. J. Math.* **91**, 693 (1969)
79. J. H. White, *J. Differential Geom.* **4**, 207 (1970)
80. J. H. White, *J. Differential Geom.* **5**, 357 (1971)
81. Y.-Q. Wu, <http://www.math.uiowa.edu/~wu/>

CHAPTER 17

MÖBIUS-INVARIANT KNOT ENERGIES

R.B. KUSNER

*Department of Mathematics, University of Massachusetts,
Amherst, MA, USA 01003-4515*

J.M. SULLIVAN

*Department of Mathematics, University of Illinois,
Urbana, IL, USA 61801-2975*

There has been recent interest in knot energies among mathematicians and natural scientists. When discretized, such energies can lead to effective algorithms for recognizing when two curves represent the same knot. These energies may also help model physical systems, such as long protein chains or DNA knots, subject to van der Waals interactions. Knot energies often are normalized to be scale-invariant; some important energies are also invariant under Möbius transformations of space. We describe computer experiments with such Möbius-invariant knot energies. We also discuss ways of extending these to energies for higher-dimensional submanifolds. The Appendix gives a table of computed Möbius-energy-minimizing knots and links through eight crossings. (This article is an updated version of our report¹ in *Geometric Topology*.)

1 Introduction

Is there an optimal way to tie a knot in space, or to embed a more general submanifold? And is there a natural way to evolve any embedding isotopically to an optimal one, so that we could detect whether two embeddings are isotopic?

One approach to such questions is to associate to any submanifold an energy, and look for minimizers or critical points of this energy. If the energy is infinite for immersions which are not embeddings, then presumably its gradient flow will prevent self-crossings and preserve isotopy type. One way to get an energy with an infinite barrier against self-crossings is to think of spreading charge along the submanifold and then consider the electrostatic potential. Such an energy for knots was introduced by Ohara² and studied by Freedman, He and Wang.³ (A new regularization of this energy has recently been found by Brylinski.⁴) We define an analogous knot energy for k -dimensional submanifolds in n -dimensional euclidean space \mathbb{R}^n (or the sphere S^n). Our knot energy is again a repulsive potential between points on the submanifold, depending only on first-order data. It is given by a regularized inverse power law, with the power chosen to make the energy scale-invariant and the regularization to make it invariant under conformal (Möbius) transformations of the ambient space.

The gradient flow of our knot energy appears to lead to optimal embeddings, both theoretically and computationally. In particular, for classical knots and links, we have used our knot energy to create an algorithm, implemented in Brakke's *evolver*,⁵ to untangle complicated curves to a simple representative for their knot type by gradient descent. In most cases, we reach the energy minimum. For instance, all unknots we have tried evolve to the round circle, and both curves in the famous Perko pair evolve to the same configuration, proving they are the same knot. Thus in most cases, this is an effective algorithm for classifying knots. However, we have also found certain links with several distinct local minima at different energy values; for these rare cases, gradient descent methods will not always reach the same final configuration.

Knot energies for curves were introduced into mathematics motivated by physical considerations; they are closely related to classically defined energies for divergence-free vector fields which arise in modeling incompressible fluid flow.^{6,7} These new knot energies may help to model certain natural phenomena. For example, the inverse power laws in knot energies seem related to some of the energies involved in arising in protein folding problems. And recent experiments suggest that the speed of DNA knots in electrophoresis gels is correlated to other notions of knot energy.^{8,9,10,11}

For surfaces, we have previously modeled¹² another Möbius-invariant energy, the elastic bending energy popularized by Willmore,¹³ in the *evolver*. It is known that this energy describes the behavior of lipid vesicles, and in fact such vesicles have been observed undergoing Möbius transformations in laboratory experiments.¹⁴ To model these vesicles in more detail, one might like to include a van der Waals interaction between different surface molecules; perhaps our Möbius-invariant knot energy would be an interesting choice for modeling such a nonlocal interaction. Our knot energies in higher dimensions or codimensions do not have obvious physical interpretation or application, although they have been useful, for example, in the topological study of knotted spheres in four-space.¹⁵

Our paper is organized as follows: in section 2, we define our family of knot energies for submanifolds of arbitrary dimensions. The next section explores the particular case of energies for knots and links, while section 4 discusses alternative regularizations of the knot energy. Section 5 shows why we should not expect minimizers for composite knots. We discuss the discretizations we have implemented in the *evolver*, and their success in untangling complicated unknots, in the following two sections. Section 8 relates knot energy to other measures of geometric complexity, like crossing number and ropelength. The next two sections discuss critical points for the energy which are guaranteed by symmetry, and the construction in this way of Hopf links with distinct local

minima for the energy. Section 11 considers the energy for higher-dimensional submanifolds in a bit more detail. Finally, we have computed energy minimizers for all knots and links up through eight crossings, and present in the last section the results of this computation and a table of their energies; our appendix shows stereoscopic pictures of the Möbius-energy minimizers.

2 Defining Möbius Energies

Recall Coulomb's Law which asserts that the potential energy between a pair of unit point charges at points x and y in \mathbb{R}^3 is given by the reciprocal $1/|x - y|$ of their distance in space. If we imagine charge uniformly spread over a k -dimensional oriented submanifold M of \mathbb{R}^n , the total energy would be given by a double integral over all pairs of points on M of some inverse power of distance. Although for physical charges in \mathbb{R}^n we might think of using the power $n - 2$, we prefer to choose the power $2k$, which makes the integrand scale-invariant. (Without scale invariance, we would need a constraint on the size of M to get nontrivial energy minima.)

Of course, the description we have given so far ignores the fact that such an integrand will blow up as x approaches y , in such a way that the integral will be infinite for any M . So we include a regularizing factor f and define

$$E_f(M) = \iint_{M \times M} \frac{f(x, y)}{|x - y|^{2k}} d \text{vol}_M(x) d \text{vol}_M(y).$$

If we did not have the factor of f , this integrand clearly would be scale-invariant. In fact, it is also easy to show (see Figure 1) that it would be invariant under inversion ($x \mapsto x/|x|^2 =: \tilde{x}$), and hence under the full conformal group of Möbius transformations of $\mathbb{R}^n \cup \{\infty\}$. Note that when computing the energy, we can view M as a submanifold of $S^n \subset \mathbb{R}^{n+1}$ via stereographic projection, instead of \mathbb{R}^n . This follows because stereographic projection from S^n to $\mathbb{R}^n \cup \{\infty\}$ extends to a Möbius transformation of $\mathbb{R}^{n+1} \cup \{\infty\}$, and the formula for energy is independent of the ambient dimension.

Thus we would like to choose our regularizer f (which is supposed to vanish as x approaches y) to be independent of scale and also to be Möbius-invariant. We will allow this function f to depend on first-order information—the tangent planes to M at the points x and y —although this was suppressed in the notation used above.

Given a point $x \in M$ and any other point p in space, there is a unique round k -sphere $S_x(p)$ tangent to M at x and passing through p . Thus given two points x and y of M , we have two oriented k -spheres $S_x(y)$ and $S_y(x)$

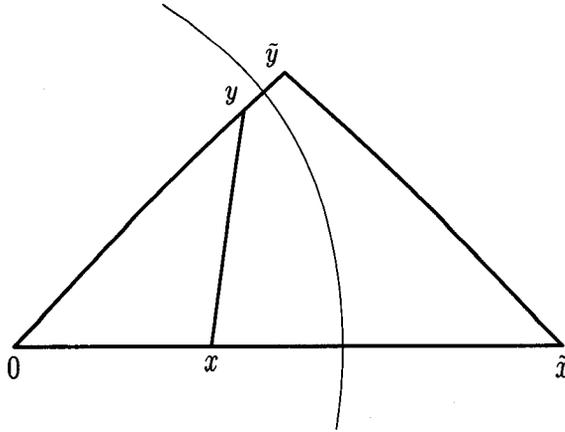


Figure 1: Here \tilde{x} and \tilde{y} are the inversions of x and y in the sphere shown. The similar triangles prove $|\tilde{x}||\tilde{y}| / |\tilde{x} - \tilde{y}|^2 = |x||y| / |x - y|^2$. Since the conformal expansion factor for inversion from M to \tilde{M} at x is $|\tilde{x}|/|x|$, the volume element changes by the k^{th} power $d \text{vol}_{\tilde{M}}(\tilde{x}) = (|\tilde{x}|/|x|)^k d \text{vol}_M(x)$. Combining these facts shows the integrand of our energy is Möbius-invariant.

which meet at equal angles at x and y . These spheres, and in particular the angle at which they meet, are defined in a Möbius-invariant manner.

By the angle between these k -spheres, we mean the angle between their tangent k -planes at points of intersection. In fact, a configuration of two oriented k -planes in \mathbb{R}^n is described by k principal angles $\alpha_1, \dots, \alpha_k$, but perhaps most useful is the combined angle α whose cosine is the inner product of two simple unit k -vectors $u_1 \wedge \dots \wedge u_k$ and $v_1 \wedge \dots \wedge v_k$ representing the two planes:

$$\cos \alpha = \prod \cos \alpha_i = \det[u_i \cdot v_j] = (u_1 \wedge \dots \wedge u_k) \cdot (v_1 \wedge \dots \wedge v_k).$$

We propose taking $f(x, y)$ to be some function of these angles α_i between the spheres $S_x(y)$ and $S_y(x)$. It should be nonnegative, to keep the energy well-behaved, and should vanish when the angles are zero, in order to cancel the singularity in the integrand. We would like the energy E_f to have the following basic properties, which qualify it as a “knot energy”:

- $E_f(M)$ is nonnegative, and zero only for $M = S^k$, the round k -sphere;
- $E_f(M)$ is infinite for immersions which are not embeddings, creating a barrier against M “crossing itself”;
- $E_f(M)$ is finite for all compact k -dimensional embedded smooth sub-manifolds $M \subset \mathbb{R}^n$.

The first two properties will be true for essentially any f which is a nonnegative function of the angle α , vanishing only at 0. The third property follows if f vanishes sufficiently fast at 0 to regularize the integral.

Note that our Möbius energies $E_f(M)$ are somewhat like “quadratic forms” on the space of oriented submanifolds. We can also examine the associated “bilinear form” $E_f(M, N)$, given by the double integral over $M \times N$. If we interpret M and N as chains or integral currents, then indeed this cross-energy will be linear in each argument, but only for positive multiples. We can think of the first submanifold M defining a potential P_M^f at all points of space. Then $E_f(M, N)$ is just the integral of this potential over the points of N . Since f depends on the tangent planes of M and N , the potential P_M^f is a function not merely of points in space, but of k -vectors at those points. In the language of geometric measure theory, P_M^f is a parametric integrand, and in fact we might think of E_f as given by a bilinear parametric integrand.

3 The Excess-Length Picture and a Standard Choice of Regularization

One good choice for the regularization in the energy E_f is $f_0 := (1 - \cos \alpha)^k$. For the remainder of this paper, we will study mostly this particular energy; we will write simply E for the energy E_{f_0} with this choice of f . This energy E generalizes the energy \tilde{E} for knots $K \subset \mathbb{R}^3$ studied by O’Hara² and by Freedman, He and Wang.³

$$\tilde{E}(K) = \iint_{K \times K} \left(\frac{1}{|x - y|^2} - \frac{1}{d_K(x, y)^2} \right) ds_K(x) ds_K(y),$$

where $d_K(x, y)$ is the shorter arclength distance within K from x to y . In fact, Peter Doyle and Oded Schramm¹⁶ introduced, in the one dimensional case, the idea of a regularization by a multiplicative factor depending on angle, and observed that $E(K) = \tilde{E}(K) - 4$ when the factor used is $f_0 = (1 - \cos \alpha)$. (Recently, Brylinski⁴ has proposed another regularization: if we define

$$B_s(K) = \iint_{K \times K} |x - y|^s ds_K(x) ds_K(y),$$

then this function of a complex number s can be meromorphically continued from the right halfplane to the entire plane, and Brylinski shows that it has poles only at the negative odd integers, and thus in particular not at $s = -2$. It turns out that $B_{-2}(K) = \tilde{E}(K) - 4 = E(K)$.)

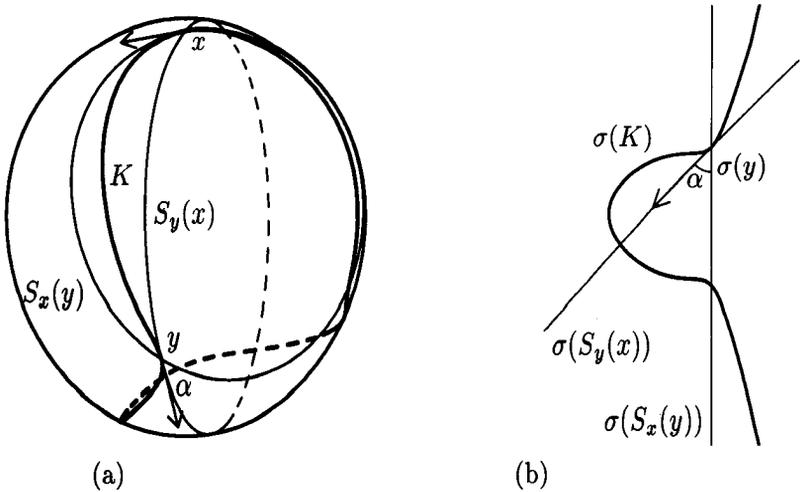


Figure 2: The excess-length picture shows how to calculate the potential $P_K(x)$ for a curve K in the sphere. When we apply the stereographic projection σ (from x as north pole) the spheres $S_x(y)$ and $S_y(x)$ become straight lines at angle α . The integral of $1 - \cos \alpha$ is the excess length of $\sigma(K)$ over that of the straight line.

One way Doyle and Schramm explained the equivalence of E and \tilde{E} is through a picture which interprets the potential $P_K(x) := P_K^{f_0}(x)$ as an excess length. Given a curve K in S^n , we want to evaluate P_K on a tangent direction at some point $x \in S^n$. To compute this, rotate the sphere so that x is at the north pole, and then stereographically project to \mathbb{R}^n (sending x to infinity). Rotate this euclidean space so that the given tangent direction at x becomes the vertical direction. (Figure 2 shows this in the case $n = 2$, when we are projecting a curve from the two-sphere to the plane; although here all embedded curves are unknotted, the energy still makes sense.) If $\sigma(K)$ is the stereographic projection of the curve K , then we can check that $P_K(x)$ equals the integral $\int_{\sigma(K)} (1 - \cos \alpha) ds_{\sigma(K)}$, where in this picture the angle α is simply the angle that the tangent to $\sigma(K)$ makes with the vertical. (Our stereographic projection is scaled nonstandardly, to take the equatorial sphere to a sphere of half the size; the extra factor of $1/2$ is needed because we want to use an inversion in a unit sphere in \mathbb{R}^{n+1} .)

If K is a closed curve, and x is not on K , this integral is simply the length of $\sigma(K)$, because the $\cos \alpha$ term integrates to zero. Thus the potential at points not on K does not depend on tangent directions, and in fact is the same as it would be if we used no regularization, setting $f = 1$. Integrating along a second, disjoint curve L , we see that the cross energy $E(K, L)$ equals the unregularized energy $E_1(K, L)$. (This is what Dave Auckly and Lorenzo

Sadun¹⁷ refer to as the additive link property for the energy E .)

But of course for x on K , we need the regularization, as $\sigma(K)$ is an infinite curve in \mathbb{R}^n , asymptotic at either end to a vertical line. Notice that the difference between the length of the curve $\sigma(K)$ and its vertical progression is given by the integral of $1 - \cos \alpha$, which gave the potential P_K . Each term is infinite, but their difference is the excess length of $\sigma(K)$ compared with the vertical straight line, which is finite if K is smooth enough at x .

One might hope to use the same picture to define energies for k -dimensional submanifolds. If a submanifold M is stereographically projected from a point $x \in M$, the image is asymptotic at infinity to a flat k -plane. However, this image submanifold has infinite excess area when compared to this k -plane, since its distance from the k -plane varies as we approach infinity in different directions (unless x is an umbilic point of M). Thus a naïve definition of energy in terms of excess area will not work; in other words the energy $E_{(1-\cos \alpha)}$ is not finite for dimensions $k > 1$. Our choice of $f_0 = (1 - \cos \alpha)^k$ overcomes this difficulty, although we do not have an equally nice geometric picture for the resulting energy E .

4 Other Möbius-Invariant Knot Energies

The Möbius energy E has been useful—theoretically, and from the perspective of computer experiments—but other Möbius-invariant knot energies are possible. First we might consider another regularizer f in our general energy E_f . Discussions with Doyle, Schramm, and Bill Thurston have focused our attention on $f = |\sin \alpha|$ as giving an interesting energy for curves. One problem here is that f is not differentiable, so E_f does not have a well-behaved gradient flow, and it is hard to model numerically. Higher powers of this function give regularizers for k -submanifolds; perhaps in this case there are further good choices for f .

Our E_f is defined in terms of first-order information at the two points x and y in the double integral. If we allow the use of higher-order information, there are other possibilities for the regularization. For two-dimensional surfaces, Auckly and Sadun¹⁷ have suggested a regularization using the squared-mean-curvature integral (which is second-order data), but this is difficult to bound below.

We have also proposed a “holomorphic” energy for embedded Riemann surfaces by considering a relative energy within conformal classes. We choose a reference embedding of a surface in space; then the energy of any conformally equivalent embedding is given by comparing the straight-line distance between a pair of points on the surface with the corresponding distance on the reference

surface. We have yet to find a proper regularization for this energy, but for spheres there is of course a unique conformal class, and the round sphere serves as a natural reference surface. This idea of relative energy should extend to rather general subsets of \mathbb{R}^n , and in fact we have already succeeded in regularizing it for embedded 1-complexes or “knotted graphs”.¹⁸

5 Prime Decomposition

We are interested in minimizing E_f within isotopy classes. Given a submanifold $M \subset \mathbb{R}^n$, we write $[M] = \{N : N \sim M\}$ for its isotopy class, and

$$E_f([M]) = \inf_{N \sim M} E_f(N)$$

for the infimum energy. One basic result³ is that prime knot types have E -minimizing representatives—this infimum is achieved. On the other hand, it seems that under energy minimization, composite knots decompose into their summands in a natural way. More generally, given a pair of k -dimensional submanifolds M and N , there are two ways to naturally combine or add their isotopy classes $[M]$ and $[N]$, in such a manner that minimizing the energy seems to separate the two pieces again.

Most trivial is the disjoint union $[M] \sqcup [N]$, obtained by embedding M and N in disjoint balls in \mathbb{R}^n . It is clear that by placing M and N far apart from each other (or equivalently scaling each one down) we can make their cross-energy $E_f(M, N)$ arbitrarily small. Since the energy of the union is the sum of the self-energies and the cross-energy, it is thus clear that

$$E_f([M] \sqcup [N]) = E_f([M]) + E_f([N]).$$

Even a submanifold of several topological components may not be decomposable in this way as a disjoint union; in this case we say it is essentially linked.

There is also a natural notion of the connected sum $[M] \# [N]$, which is well defined when M and N are both connected. (Of course, if M or N has more than one component, we simply must specify which components are to be connected.) We say a submanifold is prime if the only way it can be decomposed as a connected sum is when one summand is isotopic to a trivial S^k . Using Möbius invariance, we find that

$$E_f([M] \# [N]) \leq E_f([M]) + E_f([N]).$$

To check this, consider $Q \sim M$ and $R \sim N$ each having E_f within any given $\epsilon > 0$ of the respective infimum values. Deform a small neighborhood of

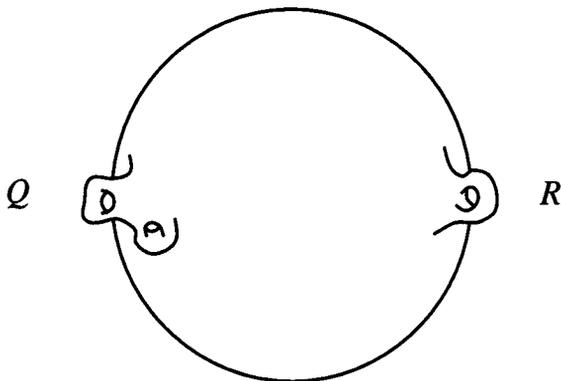


Figure 3: After applying Möbius transformations to Q in $[M]$ and R in $[N]$ to get large round pieces, we can weld them together to make this representative for $[M]\#[N]$, with energy not much more than $E_f(Q) + E_f(R)$.

some point on Q to be flat, changing the energy only by ε , and then apply a Möbius transformation mapping this small piece of a k -plane to almost all of a round S^k . Apply the same procedure to some point on R . Welding the resulting submanifolds together in the obvious way, we get a submanifold P in the class $[M]\#[N]$ (resembling a round k -spherical “head” with copies of Q and R attached as small “ears”—see Figure 3) with

$$E_f(P) \leq E_f(Q) + E_f(R) + 3\varepsilon,$$

where the last ε includes the interaction terms between the two “ears”. Infimizing yields the desired inequality.

In fact, we conjecture equality holds for infima of the energy E , and moreover, that minimization of E leads to a natural “conformal connected sum decomposition” of a submanifold into E -minimizing, essentially linked, prime submanifolds. This phenomenon has been observed in our computational investigations of the energy E for knots, links and surfaces in \mathbb{R}^3 , described in the next section. Figure 4 shows what we expect is a minimizing sequence for the energy of a connected sum of two trefoil knots. (Note that we consider left- and right-handed trefoils to be distinct isotopy classes, and that this matters when taking connected sum. Of course they have the same energy, so are not considered separately in our later knot tables.)

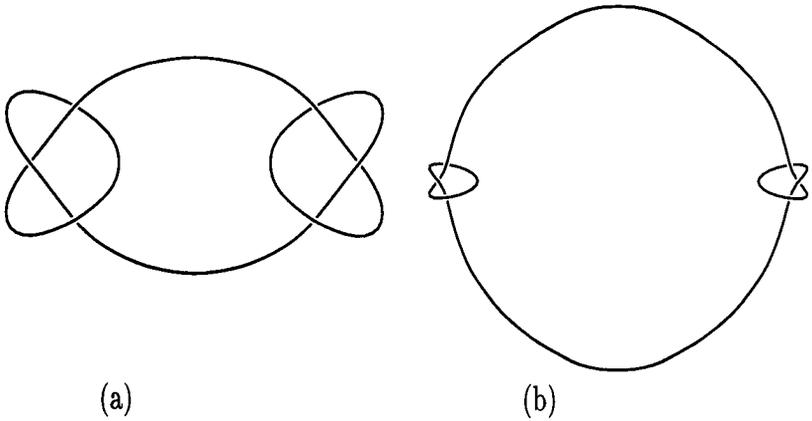


Figure 4: We expect that an energy-minimizing sequence for this “square knot” type, the connected sum of a left-handed and a right-handed trefoil knot, will begin as above. The knot (a) has energy approximately 150, while (b) has energy approximately 142. The limit energy of the sequence is approximately 140.824, or exactly twice the energy of the critical trefoil of Section 9.

6 Discretization and Computer Experiments

In order to gain some intuition into the behavior of the Möbius energy, especially for knots and links, we have implemented various discretizations of E and its gradient flow, and carried out computer experiments using Brakke’s *evolver*. All our discrete models work with polyhedral surfaces or polygonal knots and links. These have infinite energy E , since they have sharp corners, so we must work with some discretization of the energy, which is supposed to model the energy of a nearby smooth curve or surface.

One discretization for curves, the *cosine energy*, places point charges at the midpoints of all edges of the polygon. The charge at x equals the length l_x of the edge. For this energy, we sum $(1 - \cos \alpha)l_x l_y / |x - y|^2$ over all pairs $x \neq y$; here α is the angle between two circles passing through the midpoints x and y . In fact, α can be more easily computed as the angle between the edge at x , and the edge at y reflected in the perpendicular bisector of \overline{xy} (so that it also passes through x).

If we approximate some smooth curve by polygonal segments, this discretization seems to give energy values quite close to the true energy of the smooth curve, even when the polygonal approximation is relatively coarse. However, its gradient flow is problematic, since some edges tend to get very short, and then sometimes fail to line up with their neighbors. We can avoid

| Number of edges | 80 | 160 | 320 | 640 |
|-----------------|-------|-------|-------|-------|
| Cosine energy | 70.85 | 70.52 | 70.44 | 70.42 |
| Edge energy | 72.24 | 71.21 | 70.79 | 70.59 |
| Vertex energy | 69.16 | 69.84 | 70.14 | 70.28 |

Table 1: These discrete energies were computed for minimal polygonal trefoil knots with equal edge lengths in S^3 , approximations to the symmetric critical trefoil whose true energy $E \approx 70.41204$ is computed in Section 9.

this problem by adding a ‘‘Hooke’’ term to the energy, which tends to keep all the edges in the polygon at some fixed length by pretending the edges are stiff springs with that equilibrium length.

The *edge energy* discretization is the same, but without the $(1 - \cos \alpha)$ factor in the summand. This models the unregularized energy E_1 , but of course the discrete sum is finite. The idea is that if we wrote down a discretization for the formula for \bar{E} , the regularization term would depend only on the edge lengths of the polygon, and not on its position in space. This subtracted term in fact is smallest when the edge lengths are equal, so leaving it out of the energy we minimize simply helps keep edge lengths equal (as the Hooke energy did). Note also that the exact integral of $1/|x - y|^2$ over a pair of disjoint line segments in space is not an elementary function of their endpoints; this is why we concentrate the charges at the edge midpoints in the edge and cosine energies.

Finally, the *vertex energy* places a charge at each vertex v of the polygonal knot. The charge l_v equals the average of the adjacent edge lengths. Again we merely sum $l_v l_w / |v - w|^2$ without any regularization. This vertex energy is the one we have used for most of our gradient flows. For informational purposes, we also compute the sum of $l_v l_w / d^2(v, w)$, which is subtracted from the vertex or edge energies as a regularization. We report the values of all three discretized energies for our final knots. All three seem to converge as the number of segments used increases, though the cosine energy is by far the most accurate. Table 1 reports these energies computed for polygonal approximations to the critical trefoil knot whose energy can be computed exactly (see Section 9).

There have been some previous experiments with the energy \bar{E} by other researchers: Kazushi Ahara¹⁹ has used a method like our vertex energy with Hooke terms to compute several simple examples, and a program of Steve Bryson²⁰ was able to approximate the minimal trefoil. Others, including Buck, Orloff and Simon,^{21,22} Fukuhara,²³ and Gunn,²⁴ have done experiments with other repulsive energies for explicitly polyhedral knots; some of these do not

model any energy for smooth knots. Since our experiments were first reported, some other groups^{25,26} have suggested using simulated annealing to find global minimizers for knot energies. We have not found that to be necessary, as the simpler gradient descent methods almost always lead to the global minimum.

Of course, the vertex-based discretization (by which we flow) does not have an infinite barrier to changing knot type. Two segments can cross each other in space if their endpoints stay far away. But in practice, this does not happen as long as we keep edge lengths short enough near tight crossings. We can do this either with the Hooke energy, forcing all edges to be short, or by selectively refining edges whose contribution to the knot energy has become large during the evolution. The latter method seems preferable, as it concentrates the vertices where they are needed, and in the case of links does not constrain the relative lengths of the components. As always, we view our polygons as approximations of smooth curves; this “retriangulation” merely maintains a good approximation.

As always with the `evolver`, formulas for the exact gradients of the discrete energies are programmed into the computer code. Thus, at any given configuration the gradient is known exactly (without testing different perturbations) and the conjugate gradient method is used to flow towards a critical point.

We have also implemented some discretizations of energies for surfaces of dimension two, though not for submanifolds of arbitrary dimension. Here, one discretization is like the vertex energy, ignoring regularization and placing a charge at each vertex equal to one third of the area of the surrounding triangles. This energy seems to work nicely for surfaces in \mathbb{R}^4 , but for surfaces in \mathbb{R}^3 it is too rigid: the high power in the $1/r^4$ repulsive energy (needed for surfaces) means that vertices are influenced mostly just by their nearest neighbors. Thus the discrete surface seems to get locked into a particular, nearly equilateral triangulation, without much freedom to move. It is also not clear if this energy models any E_f .

For surfaces in space, we have had more success with a discretization of the $(1 - \cos \alpha)^2$ energy which places a charge at the center of each face, equal to its area, and computes the angle between the tangent planes at pairs of faces. We have computed, for instance, a tube around a trefoil knot (see Figure 5) with energy about 638, but we have yet to do comprehensive experiments with this energy. As with the similar cosine energy for links, we must pay special attention to keep the triangulation from degenerating during the evolution. Dennis Roseman¹⁵ has made use of both of our discretizations for surface energies to simplify knotted and unknotted surfaces in four space.

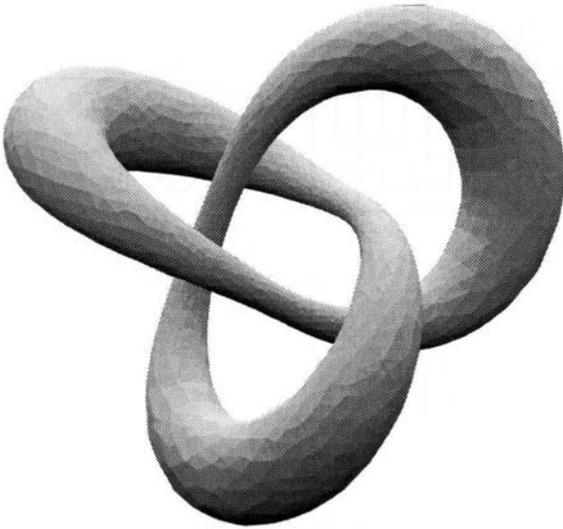


Figure 5: This torus, in the isotopy class of a tubed trefoil knot, seems to minimize energy, with $E \approx 638$. In S^3 , it would evidently be the orbit of a small circle under a rigid rotation.

7 Untangling Unknots

For any prime knot type, the existence of an E -minimizer is guaranteed by the result of Freedman, He and Wang³ already mentioned. But this leaves open the interesting question of whether the same knot type might have other critical points for E , and in particular further local minima. Of course, the E -gradient flow will be most useful for classifying knots if no other minima exist.

Of special interest are tangled unknots. Recently, it has been announced²⁷ that algorithms (based on methods of Haken) can determine if a polygonal curve is unknotted in nondeterministic polynomial time. There is no efficient algorithm known for geometrically untangling an unknotted curve, though according to Hatcher's solution²⁸ of the Smale Conjecture, there is no obstruction to finding a flow which evolves any unknotted curve to a round circle. We would be surprised if the E -flow accomplished this, but we were equally amazed to see that the (discrete) flow did untangle the example³ in Figure 6. Although initially the evolution seems to lead to a large loop caught in a tight slip-knot, perhaps it is the Möbius invariance which lets different parts of the curve grow or shrink as necessary to untangle the unknot to a round circle. The full process is shown in our six-minute video,²⁹ along with other examples

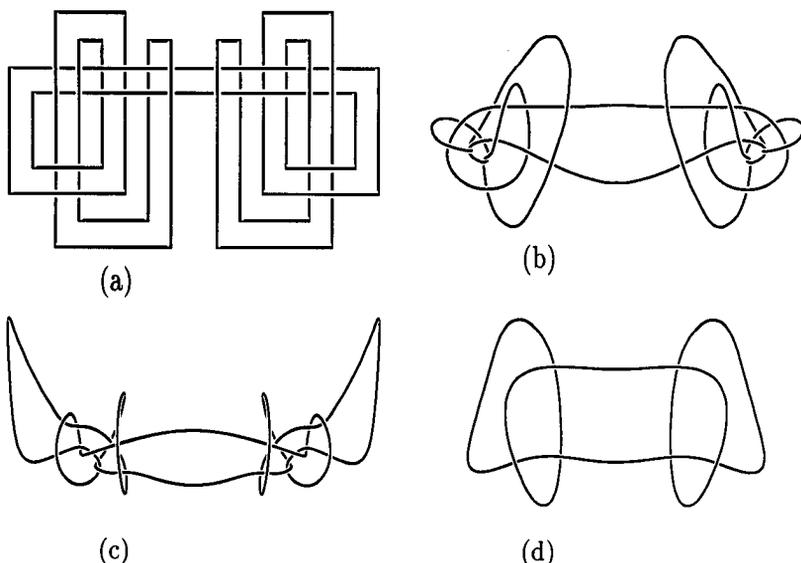


Figure 6: The tangled unknot (a) quickly rounds off to the smooth curve (b). Eventually loops grow as needed to swing out around the knobs, giving the curve (c), which can then shrink these loops to pull them through. The curve (d) clearly has no further obstacles to becoming a round circle, and indeed quickly evolves there.

of evolutions towards E -minimizing knots and links.

Of course, our experiments showing how this one curve untangles leave open the basic question: are there any E -critical unknotted curves besides the round circle? A negative answer would give an elegant analytic proof of Hatcher's theorem.

8 Crossing Numbers and Ropelength

Knots and links are often studied by means of planar projections with marked crossings. Any two projections of equivalent knots can be obtained from each other by a sequence of Reidemeister moves.³⁰ The topological crossing number $c([K])$ of the link type $[K]$ is defined as the minimum number of crossings in any planar projection.

From a three-dimensional geometric perspective, perhaps more interesting is the average crossing number $A(K)$ of the space curve K , which is the average number of crossings in planar projections of the curve, averaged over all possible orthogonal projections. This can be computed by a formula of Gauss:

$$4\pi A(K) = \iint_{K \times K} \frac{|(u \times v) \cdot (x - y)|}{|x - y|^3} ds_K(x) ds_K(y),$$

where u and v are the unit tangent vectors to K at x and y , respectively. In our computations, we discretize this for a polygonal link as a double sum over all pairs of edges, using the edge vectors for u and v , and the edge midpoints for x and y . It is clear that $A(K) \geq c([K])$ for any link.

One result of Freedman, He and Wang³ is a relation between the energy E and the crossing number, namely $E(K) \geq 2\pi c([K])$. This follows from a similar relation between the energy of any curve and its average crossing number, once we apply a Möbius transformation to send one point of the curve to infinity. They also show that $E(K) \geq \frac{6}{11} 2\pi A(K) - \frac{56}{11}$. They prove these results only for knots, although it is easy to check they remain true for our link energy. Note, in this context, their normalization of the energy for links different from ours: they count cross terms between different components only half as much as the self-energy terms. We believe our normalization is more natural, providing a good ordering of all links by energy, independent of how many components they have.

Another scale-invariant (but not Möbius-invariant) geometric measure of the complexity of space curve K is its ropelength $L(K)$, which is the arclength divided by “thickness”—essentially the diameter of the biggest embedded tubular neighborhood.^{31,32} It is known^{33,11,34} that $A(K) \leq CL(K)^{4/3}$ for a universal constant $C \leq 1/4$, and similarly³⁵ that certain energies similar to ours are also at most $L(K)^{4/3}$ times a constant. It would be interesting to prove this for the Möbius energy.

Since the standard projection of any knot type has a minimal number of crossings, we usually like to reduce any projection of that knot to the standard one only using Reidemeister moves which decrease the number of crossings. But sometimes this is impossible. The projection of our initial curve shown in Figure 6(a) has 32 crossings, and it is easy to check that no moves are applicable except the ones which increase crossing number. It seems that at least four extra crossings must be introduced to move this diagram to the zero-crossing picture of the unknot. Therefore, before our experiments, it was reasonable to think that the corresponding three-dimensional (but nearly planar) example might not untangle under the E -flow.

9 Critical Knots and Links from Group Actions

One can also ask about the existence of further critical points for truly knotted curves. Perhaps the simplest of these are the (p, q) torus knots and links. These curves can be realized as (unions of) orbits of points under rigid rotations of S^3 , which rotate one two-plane at speed p and the perpendicular one at speed q . (These rotations can also be viewed as Möbius circle actions on $\mathbb{R}^3 \cup \infty$.)

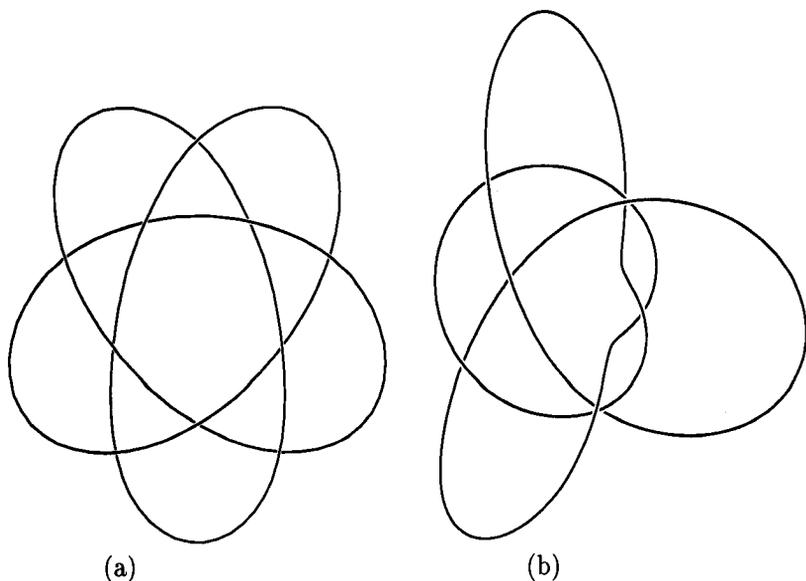


Figure 7: The E -critical orbital $(3, 5)$ torus knot (a), with energy just over 265, is not stable, while the knot (b), with energy numerically computed to be slightly under 260, seems stable and presumably is the minimizer.

In earlier work with Denise Kim,³⁶ explicit E -critical torus knots and links were constructed, by varying the ratio of radii of the torus containing these orbits until we reach a critical ratio, then using the “principle of symmetric criticality”.

Once we have reduced the energy to a function on the space of orbits (which is finite dimensional, and in this case can be parametrized by the ratio of radii), this principle asserts that at a critical point of the reduced energy, the corresponding knot or link is critical for E among *all* variations. To see this, note that if the gradient of E did not vanish, we could simply average it over the orbit to get a variation through orbits which changed (to first order) the reduced energy, contradicting the fact that we are at a critical point on the orbit space.

It can be shown¹⁸ that for large p and q (each at least 3) these orbital (p, q) torus knots and links are unstable for E (compare Figure 7), hence they are not the E -minimizers which are guaranteed to exist by the direct method.³ In particular, there is more than one critical point for E within each of these isotopy types. On the other hand, the orbital $(2, q)$ torus knots do appear to be the E -minimizers, at least experimentally; as $q \rightarrow \infty$, these converge to an E -critical double-helix (see Figure 8) whose pitch, close to 1.454, is a universal constant.

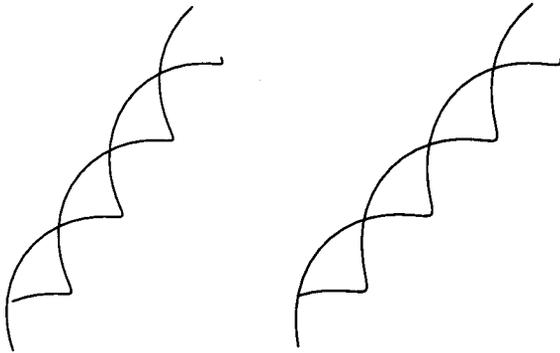


Figure 8: This double helix, critical for E , is a limit of minimizing $(2, q)$ torus knots for large q ; many knot minimizers seem to include segments which look like a piece of this helix. We show a stereo pair of pictures; to see the stereoscopic effect, look at one figure with the left eye and the other with the right eye (either by crossing the eyes or by straightening them, perhaps with the aid of a stereo viewer).

The energies of these orbital critical points can be computed by an analytic formula, derived with Gil Stengle.³⁶ This is obtained by using the circle action to reduce the double integral for E to an explicit single integral of a meromorphic differential around the unit circle $S^1 \subset \mathbb{C}$, and evaluating this as a residue sum. For example, the critical trefoil (or $(2, 3)$ torus knot, see Figure 9) has Möbius energy

$$E = 4\pi^2 \min_r (4r^2 + 9) \sum \operatorname{Res} \frac{-z^2}{r^2 z(z^2 - 1)^2 + (z^3 - 1)^2} \approx 70.41204,$$

where the sum is over poles within the unit circle, and the minimum is achieved at $r \approx 1.857$.

10 Hopf Links and Electrons on S^2

We can find more examples of links with several E -critical points, and presumably several local minima, by examining the special case of Hopf links with the methods of the previous section. Recall that the Hopf fibration of S^3 is given by the orbits of the usual action of $S^1 \subset \mathbb{C}^*$ on $S^3 \subset \mathbb{C}^2$. Each fiber is a great circle, and each pair of fibers has linking number 1. We call the union of any p fibers a p -component geometric Hopf link; for a given p , all such links are isotopic. We can find E -critical points for this link type among such geometric

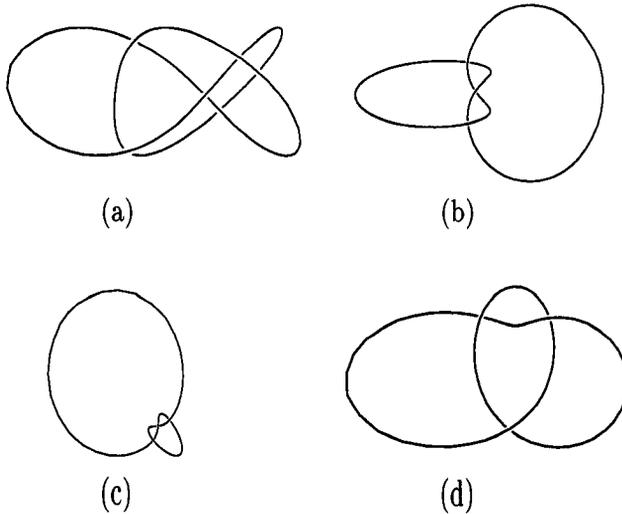


Figure 9: All these are views of the same critical trefoil, the presumed energy minimizer, differing only by Möbius transformations. The view (a), which is a different planar projection of the same space curve shown in the Appendix, is a $(2, 3)$ -torus knot in space, while (b) is dually a $(3, 2)$ -torus knot. The view (c) shows how any curve can be shown as a large circle with a small “ear”, while (d) is a randomly chosen view.

Hopf links by lifting a finite set of points from S^2 , the orbit space for the Hopf action.

Explicitly, suppose that in our p -component Hopf link $(\Gamma_1, \dots, \Gamma_p)$, each component Γ_j corresponds to the point x_j on S^2 . Then we can compute

$$E(\Gamma_1, \dots, \Gamma_p) = 4\pi^2 \sum_{i \neq j} \frac{1}{d_{ij}},$$

where $d_{ij} = |x_i - x_j|$ is the straight-line distance in \mathbb{R}^3 between the points x_i and x_j . In other words, the Möbius energy of a geometric Hopf link is simply (a multiple of) the ordinary Coulomb energy of the corresponding point charges in \mathbb{R}^3 constrained to lie on the round sphere. Thus, in this case, the residue sum for E has a simple geometric interpretation; it would be interesting to know whether this interpretation could be extended beyond Hopf links to torus knots and links.

When the number of components p is less than 4, it is easy to see that there is only one critical configuration of charges on S^2 (or thus of Hopf circles in S^3). This is the global minimum for the Coulomb energy, with the points spaced equally around the equator, corresponding to (p, p) torus links. But when $p = 4$ there are two distinct critical configurations: the equatorial configuration (or geometric $(4, 4)$ torus link) is an unstable equilibrium, while the global minimum has points at the vertices of a regular tetrahedron (the link having

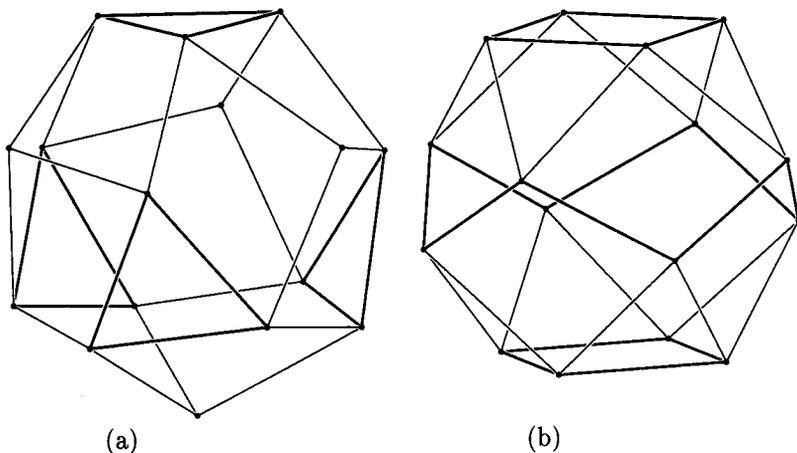


Figure 10: Two stable configurations of sixteen point charges on the sphere, shown with the closest pairs joined with edges. Configuration (a) has tetrahedral symmetry and lower energy; configuration (b) has amphichiral D_8 symmetry and slightly higher energy.

fallen off the torus in S^3).

In the early part of this century, just before the discovery of quantum mechanics, there was interest in this problem, because it was thought that stable configurations of electrons on a sphere might explain the periodic table of elements. With this in mind, configurations of $p \leq 8$ points were analyzed in detail by Föppl.³⁷

To examine the structure of critical geometric Hopf links in general, consider a variation which moves each of the corresponding points x_i by Δx_i , and set $u_{ij} = (x_i - x_j)/d_{ij}$ and $v_{ij} = (\Delta x_i - \Delta x_j)/d_{ij}$. Then to second order, the change in energy is

$$\Delta E = 2\pi^2 \sum_{i \neq j} \frac{1}{d_{ij}} (2v_{ij} \cdot u_{ij} - v_{ij} \cdot v_{ij} + 6(v_{ij} \cdot u_{ij})^2).$$

Of course, if the points x_i are to remain on S^2 , we must also impose on the variation the constraint $\Delta x_i \cdot x_i = 0$ for $i = 1, \dots, p$.

We have done extensive numerical experiments to find such stable configurations of p point charges on S^2 . These indicate that for $p < 16$ there is a single minimum. When $p = 16$ there seem to be two stable configurations, shown in Figure 10. One has tetrahedral A_4 symmetry of order 12 and $E \approx 7336.010$; the other has amphichiral dihedral D_8 symmetry of order 16 and $E \approx 7336.697$; the first local minimum seems to attract about three-quarters of the random

configurations we start with. For configurations of greater numbers of points, not surprisingly there are again usually distinct local minima.

Our experiments also suggest that the corresponding geometric Hopf links are Möbius-energy stable. Thus, the 16-component Hopf link gives the first known example of a link type with distinct local minima for the knot energy E . In general, we conjecture that the Morse index for an equilibrium configuration of point charges equals that of the corresponding geometric Hopf link. The configuration space of distinct points on S^2 has plenty of cohomology, so we expect to find many unstable extrema by viewing the reduced energy as an equivariant Morse function.³⁸ (In this context, Kawazumi³⁹ has investigated critical points of a logarithmic repulsive potential.)

11 Surfaces and Submanifolds

In the case of surfaces and higher dimensional submanifolds, we know much less about the existence of E -stationary examples or E -minimizers. Presumably, for instance, the round S^2 is the unique critical point for spheres in \mathbb{R}^3 (again consistent with the Smale Conjecture), and prime knotted spheres in \mathbb{R}^4 have E -minimizing representatives.

Using a simple scaling argument we are able to prove some partial regularity results in general dimensions. For example, tangent cones (if they exist) to k -dimensional submanifolds with finite E must be flat k -planes.¹⁸ Presumably E -minimizers have tangent cones everywhere, from which it follows that they are at least C^1 submanifolds; in fact, we expect all E -minimizers to be real analytic submanifolds.

To explicitly compute the energy E of a submanifold it is helpful to rewrite the formulas for the angle α . If two points x and y on $M \subset \mathbb{R}^n$ are separated by the vector $r = x - y$, and the tangent spaces to M at these points are spanned by s_1, s_2, \dots, s_k and by t_1, t_2, \dots, t_k , then we do not have to find the spheres $S_x(y)$ and $S_y(x)$ in order to determine how they meet. Instead, note that $S_x(y)$ is tangent to M at x , while the tangent space to $S_y(x)$ there is obtained by reflecting the tangent space $T_y(M)$ in the vector r , and is thus spanned by the vectors $\bar{t}_i := t_i - 2(r \cdot t_i)r/r^2$. Observe that the \bar{t}_i have the same inner products with each other as the t_i do; however, this basis for the tangent space to $S_y(x)$ has the wrong orientation, because of the reflection.

Define a $(k+1) \times (k+1)$ matrix A by setting $A_{00} := r^2/2$, $A_{0j} := r \cdot t_j$,

$A_{i0} := s_i \cdot r$, $A_{ij} := s_i \cdot t_j$. Thus we have

$$A = \begin{pmatrix} r^2/2 & \dots & r \cdot t_j & \dots \\ \vdots & \ddots & \vdots & \\ s_i \cdot r & \dots & s_i \cdot t_j & \dots \\ \vdots & & \vdots & \ddots \end{pmatrix}.$$

By row reduction applied to the leading row and column of A , we find $\det A = \frac{r^2}{2} \det[s_i \cdot \bar{t}_j]$. This latter determinant is the one used to find $\cos \alpha$, so we get

$$\frac{-r^2}{2} \cos \alpha |s_1 \wedge \dots \wedge s_k| |t_1 \wedge \dots \wedge t_k| = \det A,$$

where the minus sign comes from the reversal of orientation by the reflection.

Note that by choosing principal angles we can arrange the s_i and t_j to be orthogonal sets of vectors with the property that $s_i \cdot t_j = 0$ unless $i = j$. Then the formula above simplifies to

$$\frac{-r^2}{2} \cos \alpha \prod_{j=1}^k |s_j| |t_j| = \left(\frac{r^2}{2} - \sum_{j=1}^k \frac{(s_j \cdot r)(r \cdot t_j)}{s_j \cdot t_j} \right) \prod_{j=1}^k s_j \cdot t_j. \quad (1)$$

As an example, let us compute the energy of a k -dimensional Clifford torus of radii r_1, \dots, r_k . We have

$$T^k = S^1(r_1) \times \dots \times S^1(r_k) \subset S^{2k-1} \subset \mathbb{C}^k,$$

parameterized by the map $f(\theta_1, \dots, \theta_k) = (r_1 e^{i\theta_1}, \dots, r_k e^{i\theta_k})$. This embedding is homogeneous, so the energy density is constant; thus we can compute $E(T)$ with a single integral over T :

$$E(T) = \iint_{T \times T} \left(\frac{1 - \cos \alpha}{r^2} \right)^k dx dy = \text{vol}(T) \int_T \left(\frac{1 - \cos \alpha}{r^2} \right)^k dx \quad (2)$$

for any fixed y , say $y = f(0, \dots, 0)$. We find it easier, though, to compute this integral not by fixing this y and letting $x = f(\phi_1, \dots, \phi_k)$ vary, but by rotating so that $x = f(\theta_1, \dots, \theta_k)$ and $y = f(-\theta_1, \dots, -\theta_k)$, where $2\theta_j = \phi_j$. The tangent vectors at these points x and y are $s_j = (0, \dots, 0, ir_j e^{i\theta_j}, 0, \dots, 0)$ and $t_j = (0, \dots, 0, ir_j e^{-i\theta_j}, 0, \dots, 0)$. Their difference vector is $r = x - y = 2i(r_1 \sin \theta_1, \dots, r_k \sin \theta_k)$.

It is a simple matter to compute the inner products:

$$s_i \cdot s_j = t_i \cdot t_j = \delta_{ij} r_j^2, \quad s_i \cdot t_j = \delta_{ij} r_j^2 \cos \phi_j,$$

$$s_j \cdot r = t_j \cdot r = r_j^2 \sin^2 \theta_j,$$

$$r \cdot r = 4 \sum r_j^2 \sin^2 \theta_j = 2 \sum r_j^2 (1 - \cos \phi_j).$$

Define an angle ϕ by $\cos \phi = \prod \cos \phi_j$. Then using the formula (1) for $\cos \alpha$, we find that

$$\cos \alpha = \cos \phi \frac{\sum r_j^2 (1/\cos \phi_j - 1)}{\sum r_j^2 (1 - \cos \phi_j)},$$

which gives

$$\frac{1 - \cos \alpha}{r \cdot r} = \frac{1}{2} \frac{\sum r_j^2 (1 - \cos \phi_j)(1 - \cos \phi / \cos \phi_j)}{(\sum r_j^2 (1 - \cos \phi_j))^2}.$$

For $k = 2$ we can now explicitly evaluate the integral (2). We find

$$\begin{aligned} E(T) &= \text{area}(T) \int_0^{2\pi} \int_0^{2\pi} \left(\frac{1 - \cos \alpha}{r \cdot r} \right)^2 r_1 d\phi_1 r_2 d\phi_2 \\ &= 4\pi^2 r_1^2 r_2^2 \int_0^\pi \int_0^\pi \left(\frac{(r_1^2 + r_2^2)(1 - \cos \phi_1)(1 - \cos \phi_2)}{(r_1^2(1 - \cos \phi_1) + r_2^2(1 - \cos \phi_2))^2} \right)^2 d\phi_1 d\phi_2 \\ &= \frac{2\pi^2 r_1^2 (r_1^2 + r_2^2)^2}{r_2} \int_0^\pi \frac{\sqrt{1 - \cos \phi_2} (r_1^2 + 3(1 - \cos \phi_2)r_2^2)}{(2r_1^2 + (1 - \cos \phi_2)r_2^2)^{\frac{7}{2}}} d\phi_2 \\ &= \frac{\pi^3}{6} \frac{3r_1^4 + 14r_1^2 r_2^2 + 3r_2^4}{r_1 r_2 (r_1^2 + r_2^2)}. \end{aligned}$$

This is a rational function of the r_j , homogeneous of degree 0. It has a global minimum at $r_1 = r_2$, corresponding to the minimal Clifford torus in S^3 , the lift of the equator under the Hopf map. We conjecture that this surface (see Figure 11) has the minimum Möbius energy for any unknotted torus in S^3 ; its energy is $5\pi^3/3$. In fact, this should be the absolute minimum for E among all nonspherical embedded surfaces in S^3 (or \mathbb{R}^3).

By the technique of symmetric criticality mentioned earlier, we do know that this surface is a critical point for E ; and for any k , in fact, the k -torus of equal radii in S^{2k-1} is a critical point for the k -dimensional energy E .

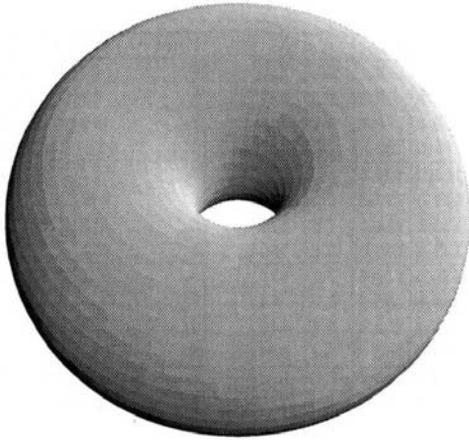


Figure 11: The minimal Clifford torus in S^3 , shown stereographically projected to \mathbb{R}^3 , has energy $E = 5\pi^3/3$, presumably the lowest of any nonspherical surface.

12 A Table of Knots and Links Minimizing Möbius Energy

We have computed experimentally (with help from a group of undergraduate students at the Five Colleges Geometry Institute⁴⁰) what seem to be E -minimizers for all the essential prime knots and links with less than nine crossings; these are pictured in the Appendix. Most of these knots have a two-bridge or rational tangle decomposition into segments where two particular strands twist around each other a certain number of half-turns.⁴¹ This decomposition seems to be reflected in the shapes of the energy minimizers: each twisted piece resembles several half-turns of the E -critical double-helix (see Figure 8), up to Möbius transformations (which can send a half-turn of the helix into a pair of large arcs near infinity). It would be nice to prove that this is the shape of a minimizer, but at present it is not even known whether, for instance, a minimizer is a real analytic curve.

We computed an approximation to each knot or link by evolving at least 9000 steps with the conjugate gradient method, refining as necessary when edges had high energy. We included all the knots and links through eight crossings as well as a few nine-crossing knots and the $(2, 15)$ torus knot.

In Table 2, the first column lists the name of the knot or link from the standard tables⁴² and its Conway notation.⁴³ The next three columns list the energy of our approximate minimizer, computed with each of the three discretizations. Finally, we list the average crossing number, and the number

| Knot or Link | E_{cos} | E_{edge} | E_{vert} | A | N_e | |
|------------------------------|-------------|------------|------------|-------|-------|------|
| 0 ₁ | 0.0 | 0.1 | -0.1 | 0.0 | 256 | |
| 2 ₁ ² | 2 | 39.5 | 39.9 | 39.2 | 2.3 | 192 |
| 3 ₁ | 3 | 70.4 | 70.8 | 70.1 | 4.1 | 264 |
| 4 ₁ ² | 4 | 99.1 | 99.5 | 98.7 | 5.5 | 320 |
| 4 ₁ | 22 | 104.9 | 105.2 | 104.6 | 5.5 | 602 |
| 5 ₁ | 5 | 126.8 | 127.3 | 126.3 | 6.9 | 360 |
| 5 ₂ | 32 | 134.6 | 135.0 | 134.1 | 7.0 | 574 |
| 6 ₃ ³ | 2,2,2- | 136.8 | 137.1 | 136.5 | 6.9 | 534 |
| 5 ₁ ² | 212 | 138.9 | 139.3 | 138.4 | 7.1 | 674 |
| 6 ₁ ² | 6 | 154.1 | 154.6 | 153.6 | 8.2 | 584 |
| 6 ₁ | 42 | 162.8 | 163.4 | 162.1 | 8.4 | 640 |
| 6 ₂ ² | 33 | 164.5 | 164.9 | 163.8 | 8.4 | 684 |
| 7 ₇ ² | 3,2,2- | 166.0 | 166.4 | 165.7 | 8.5 | 560 |
| 6 ₂ | 312 | 168.5 | 169.1 | 167.9 | 8.6 | 746 |
| 6 ₃ ² | 222 | 170.0 | 169.8 | 169.4 | 8.5 | 722 |
| 6 ₃ | 2112 | 172.9 | 173.2 | 172.4 | 8.7 | 914 |
| 7 ₈ ² | 21,2,2- | 173.9 | 174.3 | 173.5 | 8.9 | 592 |
| 6 ₁ ³ | 2,2,2 | 174.6 | 175.2 | 174.2 | 8.6 | 756 |
| 7 ₁ | 7 | 181.0 | 181.7 | 180.4 | 9.7 | 504 |
| 7 ₂ | 52 | 190.3 | 191.0 | 189.6 | 9.8 | 728 |
| 7 ₃ | 43 | 192.7 | 193.2 | 191.9 | 9.7 | 756 |
| 8 ₃ ⁷ | 4,2,2- | 194.2 | 194.7 | 193.8 | 9.8 | 612 |
| 7 ₁ ² | 412 | 196.6 | 197.2 | 195.9 | 9.9 | 820 |
| 8 ₁₉ | 3,3,2- | 197.0 | 197.3 | 196.7 | 10.2 | 680 |
| 7 ₄ | 313 | 197.7 | 198.1 | 197.1 | 9.9 | 940 |
| 6 ₂ ³ | 6* | 197.9 | 198.3 | 197.4 | 9.2 | 492 |
| 7 ₃ ² | 232 | 198.7 | 199.1 | 198.0 | 9.8 | 796 |
| 7 ₅ | 322 | 199.7 | 200.0 | 199.0 | 9.9 | 902 |
| 8 ₁₅ ² | 22,2,2- | 203.3 | 203.8 | 202.9 | 10.2 | 826 |
| 7 ₂ ³ | 3112 | 203.6 | 203.8 | 203.2 | 10.1 | 1224 |
| 7 ₆ | 2212 | 203.7 | 204.0 | 203.1 | 10.2 | 970 |
| 8 ₈ ³ | 31,2,2- | 203.7 | 204.3 | 203.2 | 10.3 | 700 |
| 8 ₂₀ | 3,21,2- | 203.9 | 204.4 | 203.4 | 10.4 | 720 |
| 7 ₄ ² | 3,2,2 | 204.4 | 205.1 | 203.7 | 10.0 | 808 |
| 7 ₃ ⁴ | 2,2,2+ | 205.4 | 206.0 | 204.7 | 10.3 | 890 |
| 8 ₄ ³ | 2,2,2,2-- | 206.0 | 206.5 | 205.5 | 9.8 | 672 |
| 7 ₇ | 21112 | 207.1 | 207.0 | 206.5 | 10.2 | 1080 |
| 8 ₁ ² | 8 | 207.9 | 208.7 | 207.1 | 11.0 | 512 |
| 8 ₁₆ ² | 211,2,2- | 208.2 | 208.5 | 207.6 | 10.3 | 912 |
| 7 ₅ ² | 21,2,2 | 208.9 | 209.4 | 208.2 | 10.1 | 916 |
| 8 ₂₁ | 21,21,2- | 209.5 | 210.0 | 209.0 | 10.5 | 852 |
| 8 ₅ ³ | (2,2)(2,2-) | 216.3 | 216.5 | 215.8 | 10.5 | 758 |
| 8 ₄ ⁴ | 2,2,2,2- | 217.2 | 217.7 | 216.6 | 10.9 | 752 |
| 8 ₁₀ ³ | (2,2)-(2,2) | 217.4 | 217.7 | 216.9 | 10.6 | 720 |

| Knot or Link | E_{cos} | E_{edge} | E_{vert} | A | N_e | |
|------------------------------|------------|------------|------------|-------|-------|------|
| 8 ₁ | 62 | 217.4 | 218.3 | 216.7 | 11.1 | 798 |
| 8 ₂ ² | 53 | 220.3 | 220.9 | 219.5 | 11.1 | 776 |
| 8 ₃ | 44 | 220.9 | 221.1 | 220.1 | 11.1 | 836 |
| 9 ₄₃ ² | 5,2,2- | 221.9 | 224.1 | 223.7 | 11.3 | 1568 |
| 8 ₂ | 512 | 224.1 | 224.8 | 223.3 | 11.3 | 898 |
| 8 ₆ ² | 242 | 226.2 | 226.8 | 225.4 | 11.4 | 934 |
| 8 ₄ | 413 | 226.2 | 226.7 | 225.4 | 11.3 | 878 |
| 8 ₅ ² | 422 | 227.9 | 227.4 | 227.2 | 11.2 | 998 |
| 8 ₆ | 332 | 228.6 | 229.2 | 227.8 | 11.3 | 914 |
| 8 ₄ ² | 323 | 229.6 | 229.7 | 228.9 | 11.2 | 974 |
| 7 ₆ ² | 6*2 | 231.1 | 231.6 | 230.6 | 10.7 | 760 |
| 8 ₇ | 4112 | 231.3 | 231.2 | 230.5 | 11.4 | 880 |
| 8 ₈ | 2312 | 232.4 | 232.9 | 231.6 | 11.5 | 1002 |
| 8 ₁ ³ | 4,2,2 | 232.7 | 233.5 | 231.8 | 11.3 | 862 |
| 8 ₉ | 3113 | 233.2 | 233.3 | 232.4 | 11.4 | 906 |
| 8 ₁₁ | 3212 | 233.7 | 233.8 | 232.9 | 11.4 | 1034 |
| 8 ₅ | 3,3,2 | 234.0 | 234.6 | 233.2 | 11.3 | 954 |
| 8 ₃ ³ | 2,2,2++ | 234.1 | 234.0 | 233.4 | 11.5 | 1220 |
| 9 ₁ | 9 | 234.6 | 235.6 | 233.7 | 12.3 | 576 |
| 8 ₁₂ | 2222 | 235.1 | 234.6 | 234.4 | 11.4 | 1036 |
| 8 ₅ ² | 3122 | 235.2 | 235.3 | 234.8 | 11.5 | 1406 |
| 8 ₇ ² | 21212 | 237.8 | 237.5 | 237.1 | 11.6 | 1176 |
| 8 ₁₃ | 31112 | 237.9 | 238.1 | 237.1 | 11.5 | 982 |
| 8 ₁₀ | 3,21,2 | 238.7 | 239.3 | 237.9 | 11.5 | 1006 |
| 8 ₁₄ | 22112 | 238.7 | 238.7 | 237.9 | 11.6 | 956 |
| 8 ₂ ³ | 31,2,2 | 238.9 | 239.4 | 238.1 | 11.5 | 900 |
| 8 ₂ ² | 21,2,2+ | 239.3 | 239.1 | 238.6 | 11.7 | 1180 |
| 8 ₆ ² | 22,2,2 | 239.6 | 240.0 | 238.9 | 11.6 | 1044 |
| 8 ₁₁ ² | 3,2,2+ | 241.6 | 241.3 | 241.2 | 11.6 | 1700 |
| 8 ₁₀ ² | 211,2,2 | 243.0 | 243.2 | 242.3 | 11.6 | 1114 |
| 8 ₁ ⁴ | 2,2,2,2 | 243.1 | 243.8 | 242.3 | 11.5 | 844 |
| 8 ₁₅ | 21,21,2 | 243.1 | 243.0 | 242.5 | 11.7 | 1154 |
| 8 ₄ ³ | (2,2)(2,2) | 245.8 | 246.5 | 245.1 | 11.5 | 906 |
| 8 ₃ ² | 211112 | 246.0 | 248.1 | 247.7 | 11.7 | 1868 |
| 8 ₃ ³ | 6*3 | 260.5 | 261.2 | 259.8 | 12.0 | 828 |
| 8 ₆ ³ | 6*2..20 | 264.4 | 264.8 | 263.7 | 12.1 | 916 |
| 8 ₁₆ | 6*2.20 | 264.5 | 264.8 | 263.9 | 12.4 | 862 |
| 8 ₁₄ ² | 6*2..2 | 264.8 | 265.1 | 264.1 | 12.1 | 894 |
| 8 ₁₇ | 6*2.2 | 265.1 | 265.7 | 264.4 | 12.3 | 832 |
| 8 ₁₃ ² | 6*21 | 266.0 | 266.2 | 265.3 | 12.2 | 998 |
| 8 ₁₈ | 8* | 283.9 | 284.3 | 283.4 | 12.6 | 844 |
| 9 ₃₁ | 2111112 | 284.1 | 284.2 | 283.9 | 13.3 | 2616 |
| 9 ₄₀ | 9* | 329.6 | 329.9 | 329.4 | 14.3 | 1640 |
| 15 ₁ | 15 | 394.2 | 395.6 | 392.8 | 20.2 | 1022 |

Table 2: Approximate Möbius energies of links through eight crossings (see text).

of edges for this polygonal link. Note that the average crossing number would change somewhat if we applied a Möbius transformation to our link, though the other values should stay constant. If we started with a different initial configuration, we might get to the same minimizer in a different picture, and the value shown for A would be different.

We believe that the cosine energy is significantly more accurate than the other discretizations, and have ordered the table by its values. Note that the edge energy tends to be a bit higher, and the vertex energy a bit lower.

Recall³ that the minimum energy E for a knot type is at least $2\pi c$, where c is the topological crossing number. Our experiments suggest this inequality is far from sharp: the minimum of the ratio E/c seems closer to $2\pi^2$, achieved by the Hopf link 2_1^2 . However, we still expect our ordering of knots by Möbius energy to list knots of small crossing number first. In fact, this is reflected in our table, though it seems that nonalternating links have significantly less energy than their crossing number would suggest—evidently two over-crossings in a row require less twisting from a three-dimensional perspective. The ordering of rational (or two-bridge) knots by our energy seems quite predictable from their Conway names (their structure as rational tangles). Indeed, for each k , the lowest energy alternating k -crossing link in the table is the $(2, k)$ torus link with notation k . The highest energy two-bridge link is $21 \cdots 12$.

Many of the non-alternating links in the table have the notation $p, q, 2-$ and in each case, this link is very close in energy to $p1q$, a link with one less crossing. Note that the link $p, 2, 2-$ consists of a $(2, p)$ torus link together with the core circle of the torus which links it twice.

Suppose we look at the highest energy knots and links for a fixed number of crossings. These in general seem to be knots based (in Conway's nomenclature) on the planar diagrams 6^* , 8^* , etc., in which all regions have at least three sides. In fact, 6^* (the Borromean rings, 6_2^3) and 8^* (8_{18}), which each have significantly more energy than all others with the same number of crossings, each fit into the class of so-called "Turk's-head" knots, with very symmetric planar diagrams. The energy minimizers stay close to this plane, with the strands weaving up and down only slightly, and presumably this accounts for the high energy. The highest-energy link of seven crossings, 6^*2 (7_6^2) is in fact the Borromean rings with one of the six crossings replaced by a double half-twist. It thus also has a symmetric planar picture, although we would have to apply Möbius transformations to the picture in the Appendix to see this.

We have included in the table four knots of nine crossings, which we believe to be extreme for energy. We expect that $5,2,2-$ (9_{43}^2) has the lowest energy of any nine crossing link. The $(2, 9)$ torus knot 9 (9_1) presumably has the lowest energy among alternating knots and links. The knot 2111112 (9_{31})

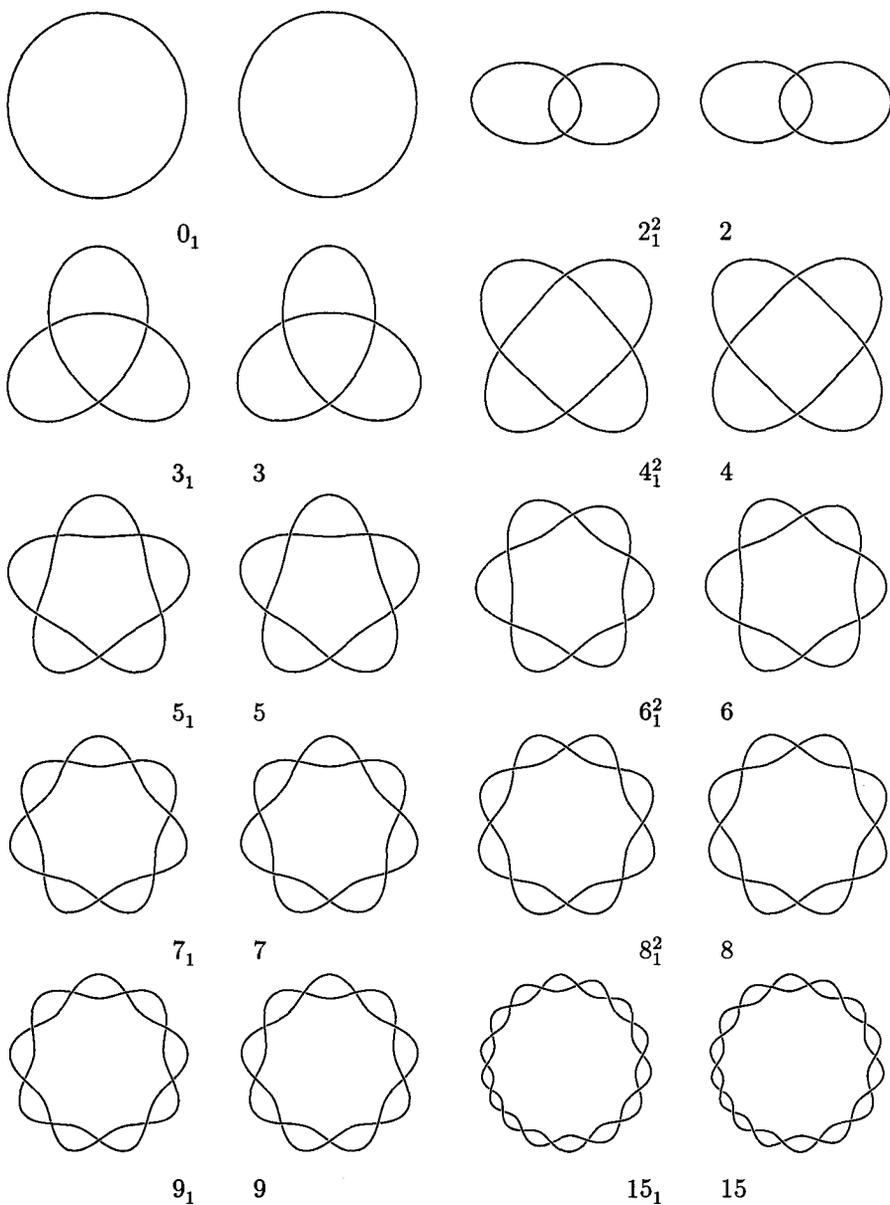
is presumably highest-energy among arithmetic nine-crossing knots. And 9^* (9_{40}) should have the highest energy overall. We have also included the $(2, 15)$ torus knot to indicate the limiting helical behavior.

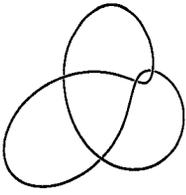
The Appendix shows stereo pictures of the approximate minimizer for each link. We made no particular effort to choose an optimal projection or Möbius representative for the minimizers; often, as for 7^2_3 , there is some conformal symmetry that fails to be Euclidean for our representative. To see the stereoscopic effect, look at the left figure with the right eye and the right one with the left eye (by crossing the eyes). Please see our report¹ in *Geometric Topology* for a version of these pictures printed instead for straight-eyed viewing.

Acknowledgments

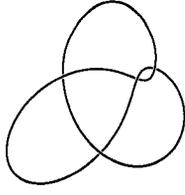
This research was supported in part by the NSF (Kusner) and by the Geometry Center (Sullivan), and at MSRI by NSF grant DMS-9022140. We would like to thank Ken Brakke for his help in adding features to the `evolver`, and John Conway, Peter Doyle, and Bill Thurston for useful discussions. This paper is an updated version of our report¹ in *Geometric Topology*.

Appendix

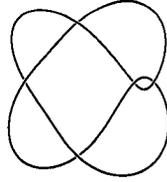
(2, q) torus knots and links



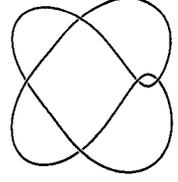
4_1



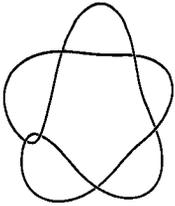
22



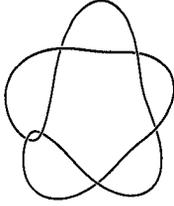
5_2



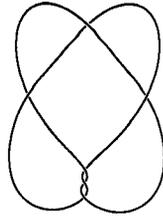
32



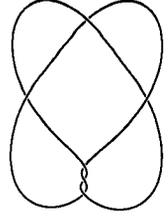
6_1



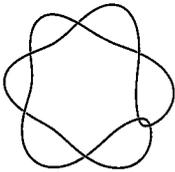
42



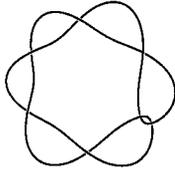
6_2^2



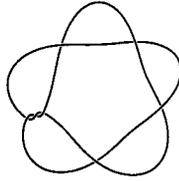
33



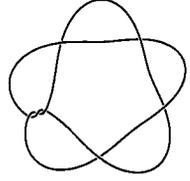
7_2



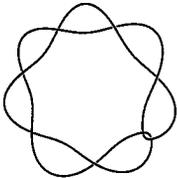
52



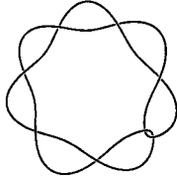
7_3



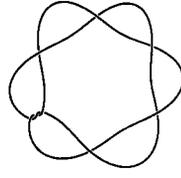
43



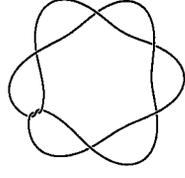
8_1



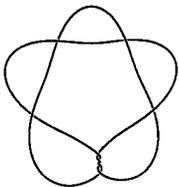
62



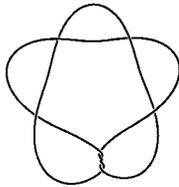
8_2^2



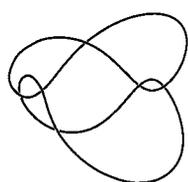
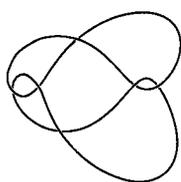
53



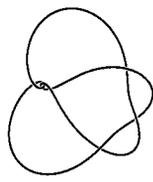
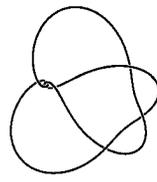
8_3



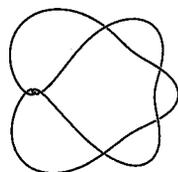
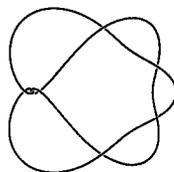
44

 5_1^2 

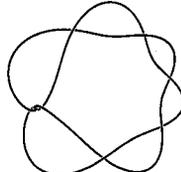
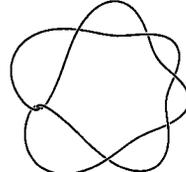
212

 6_2 

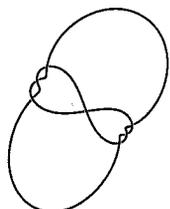
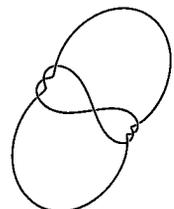
312

 7_1^2 

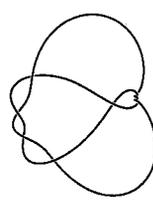
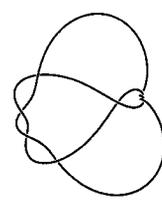
412

 8_2 

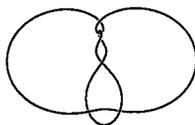
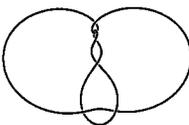
512

 7_4 

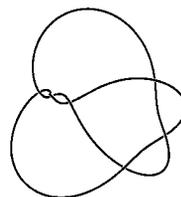
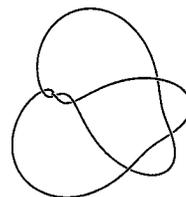
313

 8_4 

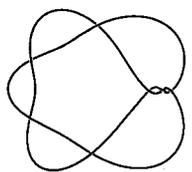
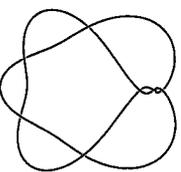
413

 6_3^2 

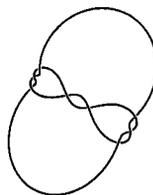
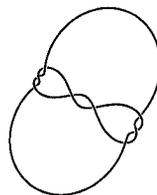
222

 7_5 

322

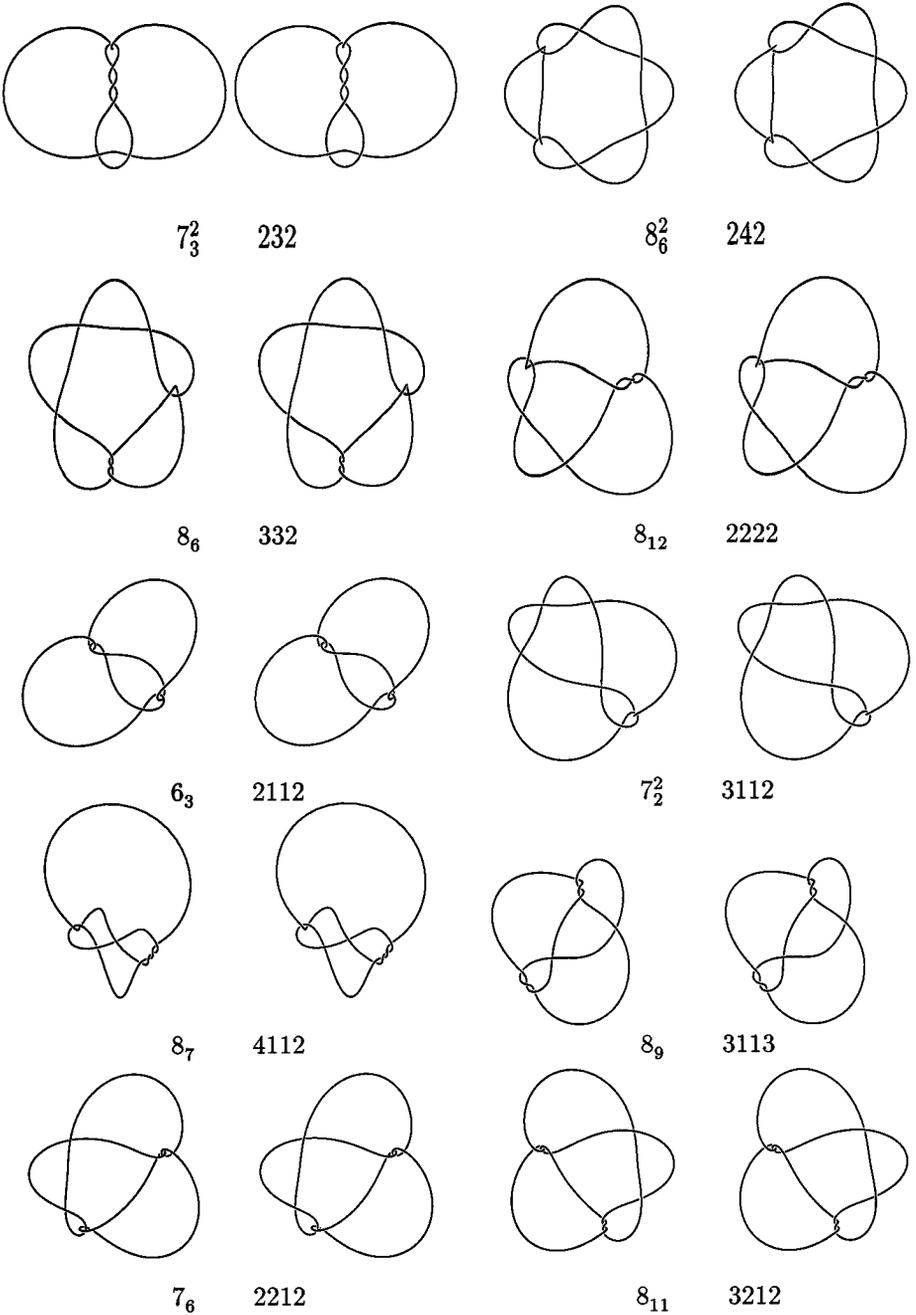
 8_3^2 

422

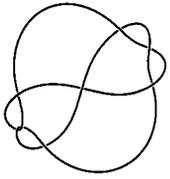
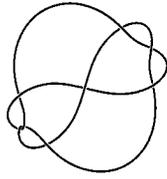
 8_4^2 

323

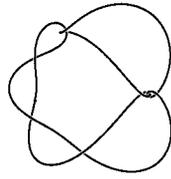
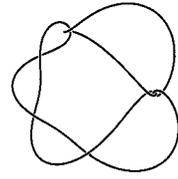
Most rational links with three-term continued fractions



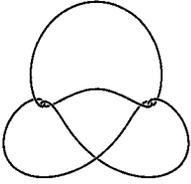
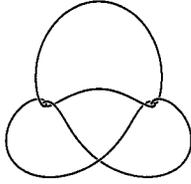
More rational links with three- or four-term continued fractions

 8_5^2 

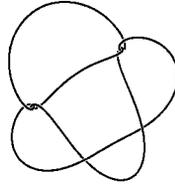
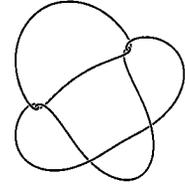
3122

 8_8 

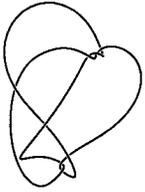
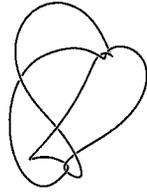
2312

 7_7 

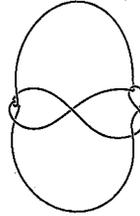
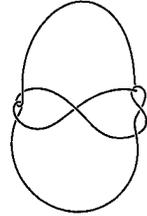
21112

 8_7^2 

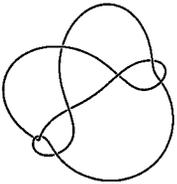
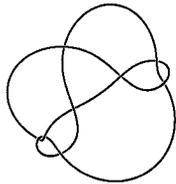
21212

 8_{13} 

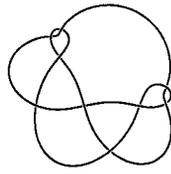
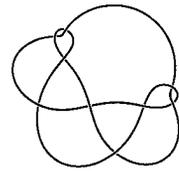
31112

 8_{14} 

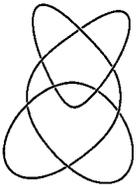
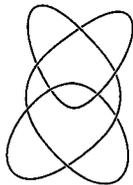
22112

 8_8^2 

211112

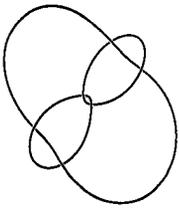
 9_{31} 

2111112

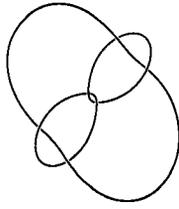
 8_{19} 

3,3,2-

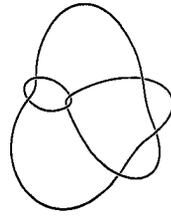
Rational links with long continued fractions, and the (3, 4) torus knot



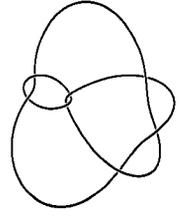
6_1^3



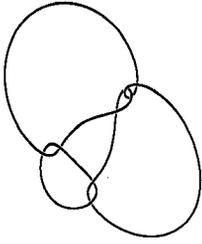
2,2,2



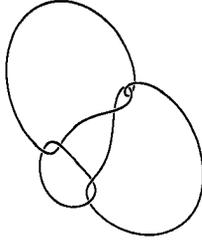
7_4^2



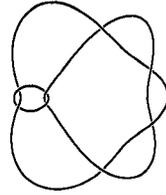
3,2,2



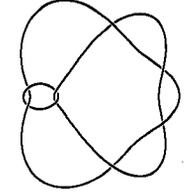
7_5^2



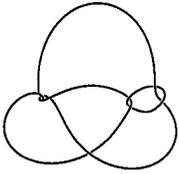
21,2,2



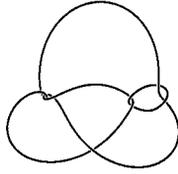
8_1^3



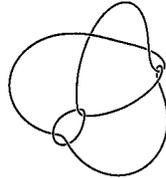
4,2,2



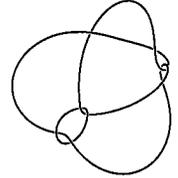
8_{10}^2



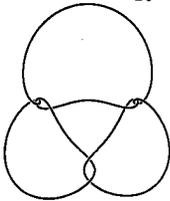
211,2,2



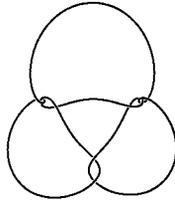
8_9^2



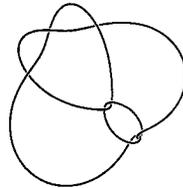
22,2,2



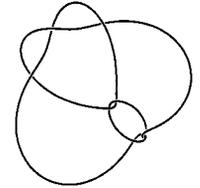
8_{15}



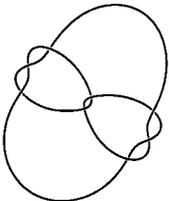
21,21,2



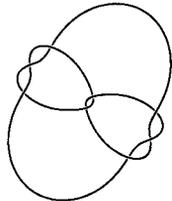
8_{10}



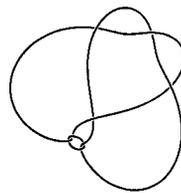
3,21,2



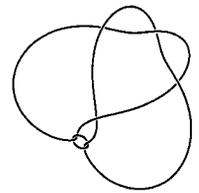
8_5



3,3,2

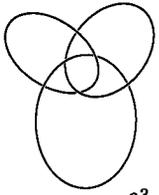
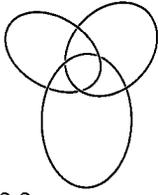
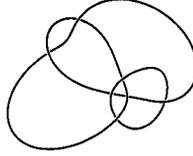
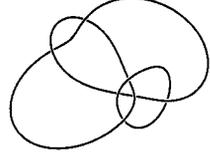
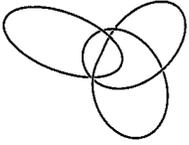
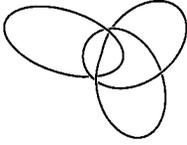
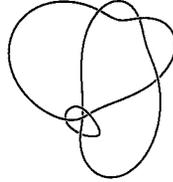
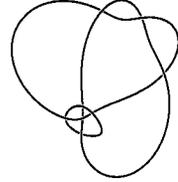
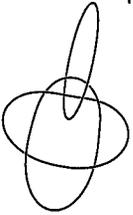
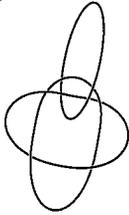
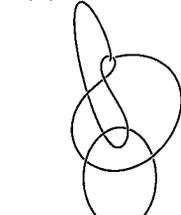
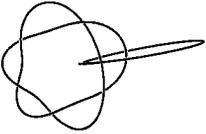
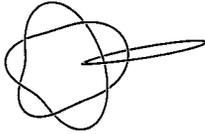
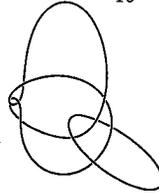
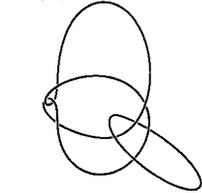
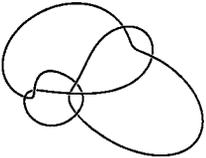
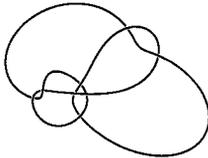


8_3^3

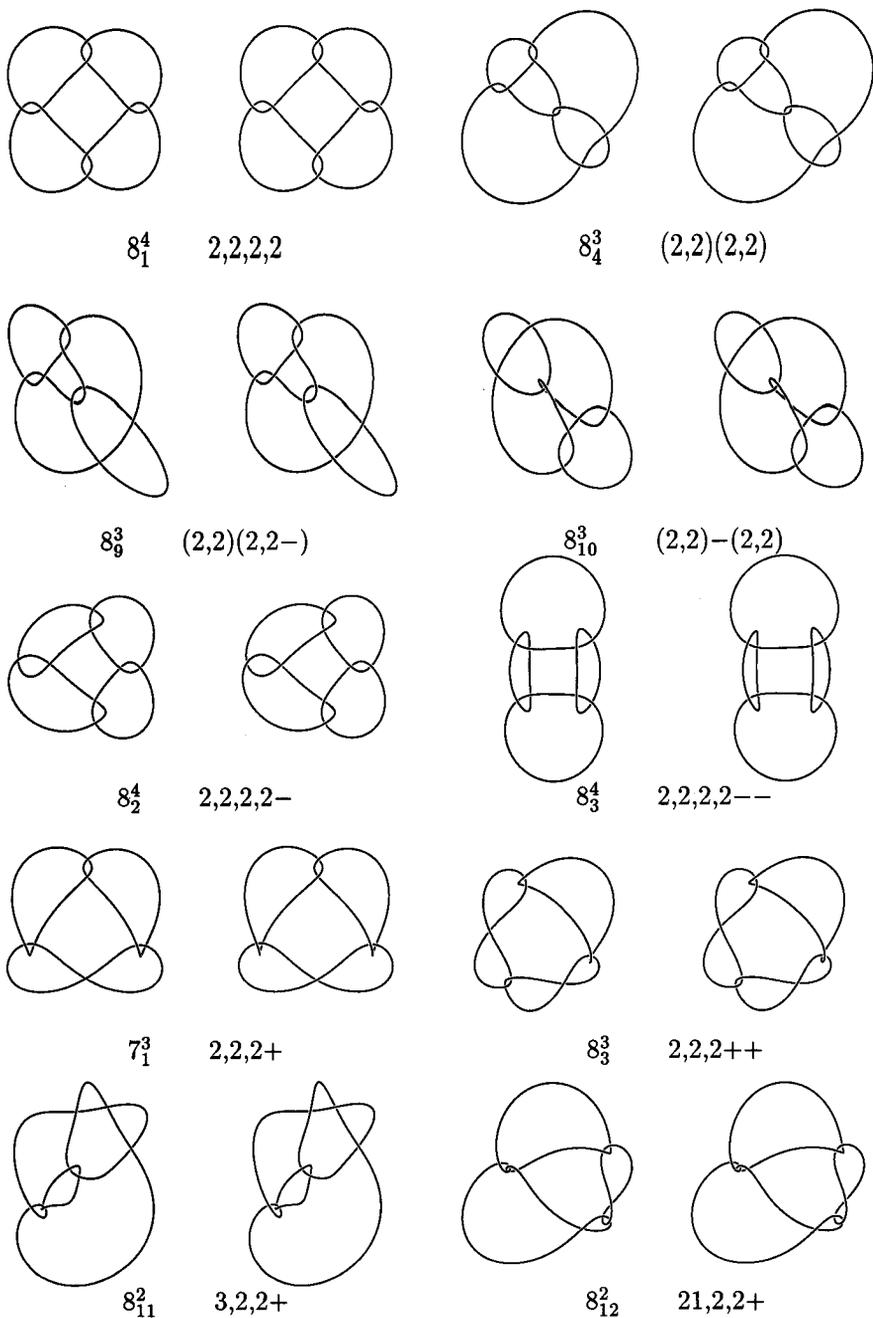


31,2,2

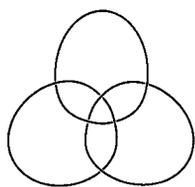
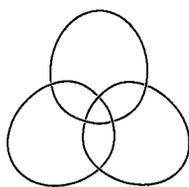
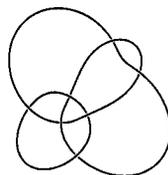
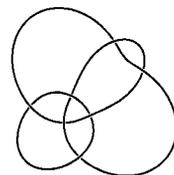
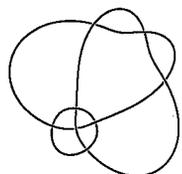
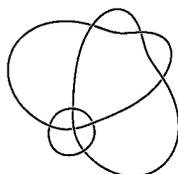
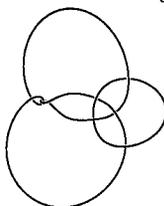
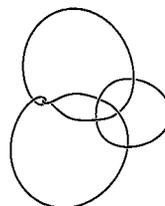
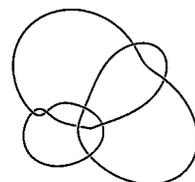
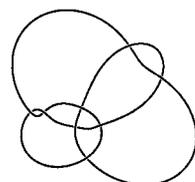
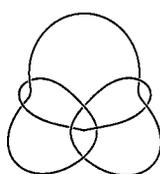
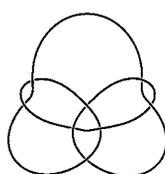
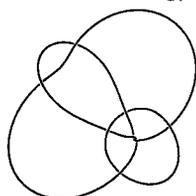
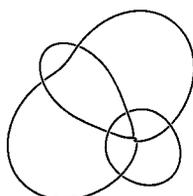
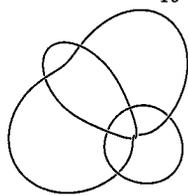
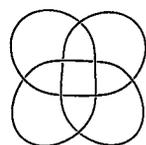
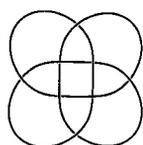
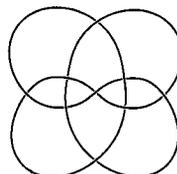
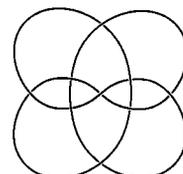
Links with Conway notation a, b, c

 6_3^3  $2,2,2-$  7_8^2  $21,2,2-$  7_7^2  $3,2,2-$  8_8^3  $31,2,2-$  8_7^3  $4,2,2-$  8_{16}^2  $211,2,2-$  9_{43}^2  $5,2,2-$  8_{15}^2  $22,2,2-$  8_{21}  $21,21,2-$  8_{20}  $3,21,2-$

Most nonalternating links



The remaining links based on 1*

 6_2^3  6^*  7_6^2  6^*2  8_5^3  6^*3  8_{13}^2  6^*21  8_{17}  $6^*2.2$  8_{16}  $6^*2.20$  8_{14}^2  $6^*2:.2$  8_6^3 $6^*2:.20$  8_{18}  8^*  9_{40}  9^*

The links based on 6^* , 8^* or 9^*

References

1. Robert B. Kusner and John M. Sullivan. Möbius energies for knots and links, surfaces and submanifolds. In Willam H. Kazez, editor, *Geometric Topology*, pages 570–604, Providence, 1997. Amer. Math. Soc./Int'l Press.
2. Jun O'Hara. Energy of a knot. *Topology*, 30:241–247, 1991.
3. Michael H. Freedman, Zheng-Xu He, and Zhenghan Wang. On the Möbius energy of knots and unknots. *Annals of Math.*, 139(1):1–50, 1994.
4. Jean-Luc Brylinski. The beta function of a knot. MSRI dg-ga preprint 9710011, 1997.
5. Kenneth A. Brakke. The Surface Evolver. *Experimental Math.*, 1:141–165, 1992.
6. Michael H. Freedman and Zheng-Xu He. Divergence free fields: energy and asymptotic crossing number. *Annals of Math.*, 134(1):189–229, 1991.
7. A. Y. K. Chui and H. K. Moffatt. The energy and helicity of knotted magnetic flux tubes. *Proc. R. Soc. Lond. A*, 451:609–629, 1995.
8. Andrzej Stasiak, Vsevolod Katritch, Jan Bednar, Didier Michoud, and Jacques Dubochet. Electrophoretic mobility of DNA knots. *Nature*, 384(14):122, November 1996.
9. Vsevolod Katritch, Jan Bednar, Didier Michoud, Robert G. Scharein, Jacques Dubochet, and Andrzej Stasiak. Geometry and physics of knots. *Nature*, 384(14):142–145, November 1996.
10. H. K. Moffatt. Pulling the knot tight. *Nature*, 384(14):114, November 1996.
11. Jason Cantarella, Robert B. Kusner, and John M. Sullivan. Crossing numbers of tight knots. *Nature*, 1997. To appear.
12. Lucas Hsu, Rob Kusner, and John M. Sullivan. Minimizing the squared mean curvature integral for surfaces in space forms. *Experimental Mathematics*, 1(3):191–207, 1992.
13. Thomas J. Willmore. A survey on Willmore immersions. In *Geometry and Topology of Submanifolds, IV (Leuven, 1991)*, pages 11–16. World Sci. Pub., 1992.
14. Xavier Michalet and David Bensimon. Observations of stable shapes and conformal diffusion in genus 2 vesicles. *Science*, 269:666–668, August 1995.
15. Dennis Roseman. What should a surface in 4-space look like? In Hans-Christian Hege and Konrad Polthier, editors, *Visualization and Mathematics*, pages 67–82. Springer Verlag, Heidelberg, 1997.

16. Peter Doyle and Oded Schramm. Personal communications.
17. David Auckly and Lorenzo Sadun. A family of Möbius invariant 2-knot energies. In Willam H. Kazez, editor, *Geometric Topology*, pages 235–258, Providence, 1997. Amer. Math. Soc./Int'l Press.
18. Robert B. Kusner and John M. Sullivan. A nonlocal, Möbius-invariant energy for embedded submanifolds. In preparation.
19. Kazushi Ahara. The numerical analysis of an energy of a knot. Preprint with video, Univ. of Tokyo.
20. Steve Bryson, Michael H. Freedman, Zheng-Xu He, and Zhenghan Wang. Möbius invariance of knot energy. *Bulletin Amer. Math. Soc.*, 28:99–103, 1993.
21. Greg Buck and J. Orloff. Computing canonical conformations for knots. *Topology Appl.*, 51:247–253, 1993.
22. Greg Buck and Jonathan Simon. Knots as dynamical systems. *Topology Appl.*, 51:229–246, 1993.
23. Shinji Fukuhara. Energy of a knot. In Y. Matsumoto, T. Mizutani, and S. Morita, editors, *A Fête of Topology*, pages 443–451. Academic Press, 1988.
24. Charlie Gunn. Personal communication.
25. Terry J. Ligocki and James A. Sethian. Recognizing knots using simulated annealing. *J. Knot Theory Ramifications*, 3(4):477–495, 1994.
26. Robert P. Grzeszczuk, Milana Huang, and Louis H. Kauffman. Physically-based stochastic simplification of mathematical knots. *IEEE Trans. on Visualization and Comp. Graphics*, 3(3):262–272, 1997.
27. Joel Hass, Jeffrey Lagarias, and Nicholas Pippenger. The computational complexity of knot and link problems. Preprint, 1997.
28. Allen E. Hatcher. A proof of the Smale conjecture $\text{Diff}(S^3) \simeq O(4)$. *Annals of Math.*, 117:553–607, 1983.
29. John M. Sullivan. Knots minimizing a Möbius-invariant energy. VHS video tape (Nov. 1993) available from the author.
30. Colin C. Adams. *The Knot Book*. W. H. Freeman, 1994.
31. Robert B. Kusner and John M. Sullivan. On distortion and thickness of knots. In S. Whittington, D. Sumners, and T. Lodge, editors, *Topology and Geometry in Polymer Science*, IMA Volumes in Mathematics and its Applications, 103, pages 67–78, New York, 1997. Springer.
32. Richard A. Litherland, Jon Simon, Oguz Durumeric, and Eric Rawdon. Thickness of knots. Preprint, 1996.
33. Jason Cantarella, Dennis DeTurck, and Herman Gluck. Upper bounds for writhe and helicity. Preprint, 1997.
34. Greg Buck and Jon Simon. Thickness and crossing number of knots.

- Topol. Appl.* to appear.
35. Greg Buck. Applications of the projection energy for knots. Preprint, 1994.
 36. Denise Kim and Rob Kusner. Torus knots extremizing the Möbius energy. *Experimental Math.*, 2:1–9, 1993.
 37. L. Föppl. Stabile Anordnungen von Elektronen im Atom. *J. reine angew. Math.*, 141:251–302, 1912.
 38. Robert B. Kusner and John M. Sullivan. Möbius energy of Hopf links, and electrons on the sphere. In preparation.
 39. Nariya Kawazumi. On the homotopy type of the moduli space of n -point sets of P^1 . *J. Fac. Sci., Univ. Tokyo, IA*, 37:263–287, 1990.
 40. Rob Kusner, Peter Norman, and John Sullivan. Möbius energy minimizing knot catalog. Preprint III.18, GANG/UMass Amherst, 1993.
 41. Gerhard Burde and Heiner Zieschang. *Knots*. Number 5 in Studies in Math. de Gruyter, Berlin, 1985.
 42. Dale Rolfsen. *Knots and Links*. Publish or Perish, Wilmington, 1976.
 43. John H. Conway. An enumeration of knots and links and some of their related properties. In J. Leech, editor, *Computational Problems in Abstract Algebra*, pages 329–358. Pergamon, 1970. Proc. Conf. Oxford, 1967.

CHAPTER 18

AN INTRODUCTION TO HARMONIC KNOTS

AARON K. TRAUTWEIN

Department of Mathematics

Carthage College

2001 Alford Park Drive, Kenosha, WI 53140, USA

E-mail: akt@carthage.edu

<http://www.carthage.edu/~trautwn>

An harmonic knot is a knot in three-dimensional space that can be expressed parametrically in the form $(x(t), y(t), z(t))$ where each coordinate function is a trigonometric polynomial. Since every knot is ambient isotopic to an harmonic knot, we are able to define the harmonic index of a knot as the smallest degree of any harmonic knot in the knot class. The harmonic index of a knot is related to its superbridge index and crossing number. In particular, there is a bound on harmonic index computed from the crossing number and vice versa. This bound allows us to compute the superbridge index of the figure-eight knot and the granny knot and to show that only finitely many knot types occur with any given harmonic index. We provide explicit harmonic parametrizations of the trefoil knot, the figure-eight knot, the $(2, 5)$ and the $(3, 4)$ torus knots, and the granny knot.

1 Introduction

In this paper, we define and give examples of harmonic knots. An harmonic knot is a knot in three-dimensional space that can be expressed parametrically in the form $(x(t), y(t), z(t))$ where each coordinate function is a finite trigonometric polynomial. Harmonic knots provide nice representations of smooth knots because of the simplicity of the coordinate functions involved. An example of an harmonic knot is the trivial knot since it can be parametrized in \mathcal{R}^3 as $k(t) = (\sin(t), \cos(t), 1)$.

Even though harmonic knots provide particularly nice representations of smooth knots, up to now little research has been done to find harmonic knots and to determine their properties. The only harmonic knots that have studied substantially are Lissajous knots. A Lissajous knot is an harmonic knot in which each of the coordinate functions is a trigonometric monomial. Bogle, Hearst, Jones, and Stoilov showed in ¹ that a Lissajous knot necessarily has Kervaire invariant zero so that the trefoil knot, the figure-eight knot, and the $(2, 5)$ torus knot are not Lissajous. However, we will show that the trefoil knot, the figure-eight knot, and the $(2, 5)$ torus knot can be represented as harmonic knots.

We give the formal definition of an harmonic knot in section 2 and show that every knot is equivalent to an harmonic knot. In sections 3, 4, and 5 we establish relationships between the harmonic index of a knot and knot theoretic indices such as the crossing number and superbridge index. The technique used to find explicit parametrizations of harmonic knots and examples of such parametrizations are given in section 6. Harmonic parametrizations for all knots through eight crossings are available in ² and via the internet at ³.

2 Harmonic knots – definition, existence, and index

Let $f_N(t)$, $g_N(t)$, and $h_N(t)$ be trigonometric polynomials of degree less than or equal to N . We say $T_N: [0, 2\pi) \rightarrow \mathfrak{R}^3$ defined by

$$T_N(t) = (f_N(t), g_N(t), h_N(t))$$

is an N^{th} degree trigonometric polynomial parametrization if at least one of $f_N(t)$, $g_N(t)$, or $h_N(t)$ has a nonzero N^{th} degree coefficient. We call the image of T_N in \mathfrak{R}^3 an N^{th} degree trigonometric polynomial curve. If $T_N: [0, 2\pi) \rightarrow \mathfrak{R}^3$ is an N^{th} degree trigonometric polynomial parametrization, we say T_N is regular if T_N is 1:1, its image in \mathfrak{R}^3 is a simple, closed curve, and $T_N'(t) \neq (0, 0, 0)$ for all t in $[0, 2\pi)$. We call the image of T_N in this case a regular N^{th} degree trigonometric polynomial curve.

Definition 2.1 An N^{th} degree harmonic knot is a knot with a regular N^{th} degree trigonometric polynomial parametrization.

The trivial knot (unknot) is a 1st degree harmonic knot since it can be represented parametrically by $k(t) = (\sin(t), \cos(t), 1)$.

Definition 2.2 We say two knots K_1 and K_2 are equivalent, denoted $K_1 \approx K_2$, if they are ambient isotopic. If two knots are equivalent, we say the knots are in the same knot class or that the knots are of the same type. The collection of all knots equivalent to a knot K is denoted by $[K]$.

Definition 2.3 A knot K is tame if it is ambient isotopic to a simple, closed polygon in \mathfrak{R}^3 . A knot K is wild if it is not tame.

Assume that a closed curve C is represented by a continuous, vector function $r(t) = (r_1(t), r_2(t), r_3(t))$ of period L which is not constant in any interval. For each unit vector \mathbf{u} , define $m(C, \mathbf{u})$ to be the number of maxima of the function $\mathbf{u} \cdot r(t)$ (i.e.

the number of parameter values t_0 for which $\mathbf{u} \cdot \mathbf{r}(t_0) \geq \mathbf{u} \cdot \mathbf{r}(t)$ for t within some neighborhood of t_0 in a fundamental period. The crookedness of C , denoted $m(C)$, is the $\min \{m(C, \mathbf{u}) : \mathbf{u} \text{ is a unit vector}\}$. In ⁴ Milnor showed a knot K is tame if its crookedness is finite. By finding the number of extrema (the crookedness) of an harmonic parametrization, we obtain the following theorem.

Theorem 2.4 *All harmonic knots are tame.*

Below in Examples 2.5 and 2.6, we give the parametric functions for two harmonic knots. Each knot is identified by its common name according to its knot type. After each example is a figure representing the image of the function in \mathbb{R}^3 as a “knotted tube,” that is, this simple, closed curve has been drawn as the center of a tube having a nonzero radius.

Example 2.5 Harmonic parametrization of the trefoil (3_1 knot).

$$x(t) = 41\cos(t) - 18\sin(t) - 83\cos(2t) - 83\sin(2t) - 11\cos(3t) + 27\sin(3t)$$

$$y(t) = 36\cos(t) + 27\sin(t) - 113\cos(2t) + 30\sin(2t) + 11\cos(3t) - 27\sin(3t)$$

$$z(t) = 45\sin(t) - 30\cos(2t) + 113\sin(2t) - 11\cos(3t) + 27\sin(3t)$$

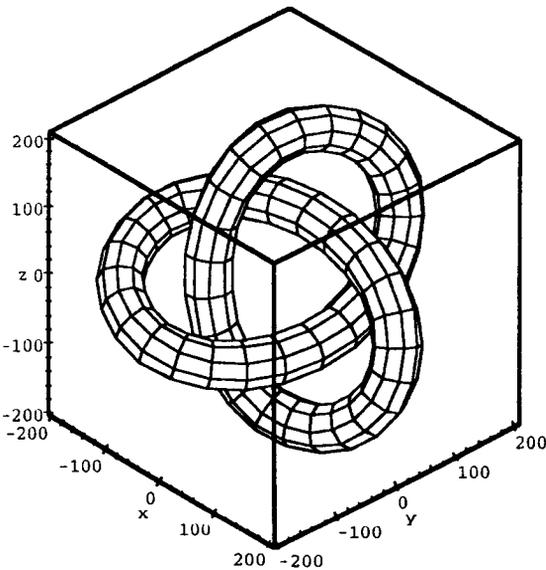


Figure 1: Harmonic trefoil knot.

Example 2.6 Harmonic parametrization of the figure-eight knot (4_1 knot).

$$x(t) = 32\cos(t) - 51\sin(t) - 104\cos(2t) - 34\sin(2t) + 104\cos(3t) - 91\sin(3t)$$

$$y(t) = 94\cos(t) + 41\sin(t) + 113\cos(2t) - 68\cos(3t) - 124\sin(3t)$$

$$z(t) = 16\cos(t) + 73\sin(t) - 211\cos(2t) - 39\sin(2t) - 99\cos(3t) - 21\sin(3t)$$

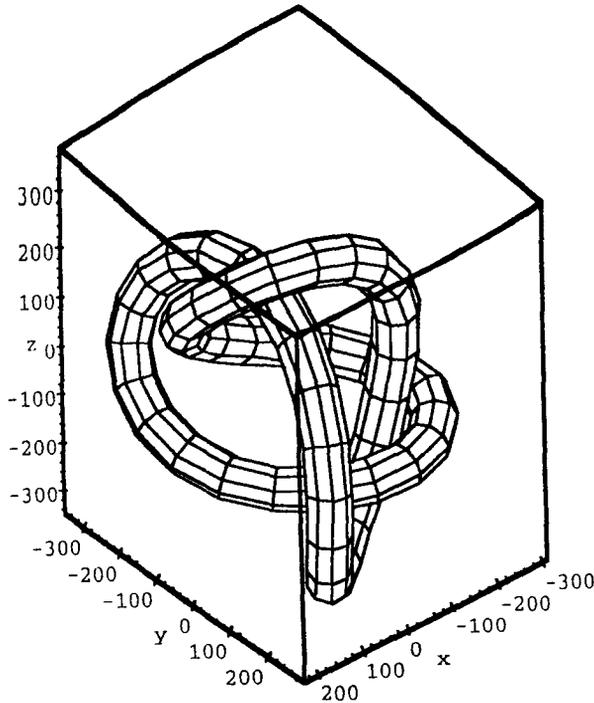


Figure 2: Harmonic figure-eight knot.

In ² we show every tame knot is equivalent to an harmonic knot. Using basic theorems about first order Fourier approximations of smooth (C^1) functions (see ⁵), we show that given a simple, closed, smooth curve K in \mathfrak{R}^3 , it is possible to find a regular trigonometric polynomial curve that lies “close enough” to K to be of the same knot type. In particular, we show there is a regular trigonometric polynomial curve lying in a normal disk tubular neighborhood of K that has the property that it never reverses direction with respect to the orientation induced on this tube by the curve K . The following theorem results.

Theorem 2.7 *If K is a tame knot then for some N there is an N^{th} degree harmonic knot K^* of the same knot type as K .*

Since each knot class contains harmonic knots, we are able to define a new knot invariant based on the degrees of the harmonic knots in a knot class called the *harmonic index*.

Definition 2.8 Let K be a tame knot. The *harmonic index* of K , denoted $d[K]$, is the minimum integer N such that there is an N^{th} degree harmonic knot of the same knot type as K .

For all knots, the harmonic index is greater than or equal to 1 by definition. The harmonic index of the trivial knot is 1 since the trivial knot has an harmonic parametrization given by $k(t) = (\sin(t), \cos(t), 1)$.

3 Superbridge index and harmonic index

If K is a tame knot, K can be represented by a continuous real-valued vector function $k(t) = (x(t), y(t), z(t))$ of period 2π . We can count the number of local maximum points relative to a given direction, and this is a measure of the complexity of the knot. To allow for whole intervals of such points, we must be careful in the method used in our counting. For a continuous map $f: \mathfrak{R} \rightarrow \mathfrak{R}$, a critical component is a nondegenerate interval J (necessarily closed) such that f is constant on J and there is no interval J^* such that f is constant on J^* , $J \neq J^*$, and $J \subset J^*$. A critical point is a point A at which f has a local maximum or minimum. Note that this does make sense whether or not K is smooth, though for smooth knots, “critical” has a different meaning.

Let \mathbf{u} be a fixed unit vector in \mathfrak{R}^3 . The dot product $\mathbf{u} \cdot k(t)$ is a smooth real-valued function of t of period 2π . If $\mathbf{u} \cdot k(t)$ is constant on \mathfrak{R} , i.e. K lies in a plane, define $b_{\mathbf{u}}(K) = 1$. If $\mathbf{u} \cdot k(t)$ has infinitely many critical components or infinitely many critical points, define $b_{\mathbf{u}}(K) = \infty$. Otherwise, we can choose α in \mathfrak{R} so that α is not contained in a critical component of $\mathbf{u} \cdot k(t)$ and α is not a critical point of $\mathbf{u} \cdot k(t)$. In this case define $\mathbf{u} \cdot k(t)$ to be the number of connected critical components of $\mathbf{u} \cdot k(t)$ added to the number of critical points where $\mathbf{u} \cdot k(t)$ has local maxima in the interval $[\alpha, \alpha + 2\pi)$. The minimum of $b_{\mathbf{u}}(K)$ over all directions \mathbf{u} was defined by Schubert in ⁶ to be the *bridge number* of K denoted $b(K)$. Kuiper examined Schubert's bridge number in ⁷ and defined a new knot invariant, the *superbridge number*. Kuiper's definition of this new invariant is given below.

Definition 3.1 Let K be a tame knot. Then the *superbridge number* of K , denoted $sb(K)$ is

$$\max \{ b_u(K) : \mathbf{u} \text{ is a direction vector in } \mathfrak{R}^3 \}.$$

The trefoil knot K in Figure 1 has only two maxima in the direction \mathbf{u} pointing toward the positive z -axis so. However, it is impossible to draw a trefoil so that in all directions we see only two maxima. In particular, in Figure 1 there are directions where we see three maxima (but no number higher). Hence the superbridge number of this trefoil is 3. Using the definitions of the bridge number of a knot K and the superbridge number of a knot K , we are now able to define bridge index and the superbridge index of K .

Definition 3.2 Let K be a tame knot. Then the *bridge index* and the *superbridge index* of K are

$$b[K] = \min \{ b(K) : K^* \text{ is of the same knot type as } K \}$$

and

$$sb[K] = \min \{ sb(K) : K^* \text{ is of the same knot type as } K \},$$

respectively.

For a tame knot K , Kuiper showed in ⁷ that $b[K] < sb[K]$ (strict-inequality). This inequality also follows from Milnor's analysis of total curvature in ⁴. By finding the maximum possible number of connected critical components where a trigonometric polynomial can have local maxima, we obtain the following theorem.

Theorem 3.3 Let K be a tame knot. Then $sb[K] \leq d[K]$.

Since the $b[K] < sb[K]$ as a simple corollary we see that the bridge index of a knot type is strictly less than its harmonic index. In particular, we can place a lower bound on the harmonic index of a non-trivial knot.

Corollary 3.4 All non-trivial knots K have harmonic index ≥ 3 .

In general, for non-torus knots the superbridge index is not known. In some cases, however, since the superbridge index of a knot is less than or equal to its harmonic index, we can use the harmonic index of a knot as an upper bound for the knot's superbridge index and harmonic index. Below in Corollary 3.5, we find the superbridge index of the figure-eight knot and the granny knot using the low degree parametrizations for these knots given in section 6.

Corollary 3.5 *The superbridge index and the harmonic index of the figure-eight knot are 3. The superbridge index and the harmonic index of the granny knot are 4.*

4 Torus knots and harmonic index

Definition 4.1 A torus knot is a knot whose image lies on the surface of a torus in \mathfrak{R}^3 . Torus knots are identified by giving two numbers p and q ($2 \leq p < q$, p and q coprime) which represent the number of times the knot crosses a meridian and a longitude on a torus, respectively. A (p, q) torus knot can be represented parametrically on a torus of radius $1/4$ by $K_{p,q}(t) = (x(t), y(t), z(t))$ where

$$\begin{aligned}x(t) &= \cos(pt) + (1/4) \cos(pt)\cos(qt) \\y(t) &= \sin(pt) + (1/4)\sin(pt)\cos(qt) \\z(t) &= (1/4)\sin(qt)\end{aligned}$$

Using standard trigonometric identities, we see that this is a $p + q$ degree trigonometric polynomial curve in \mathfrak{R}^3 . The parametrization of the (p, q) torus knot above does not necessarily give the minimum harmonic degree. Kuiper showed in ⁷ that if $K_{p,q}$ is a (p, q) torus knot then $sb[K_{p,q}]$ is the minimum of $\{2p, q\}$. For example, a $(2, 5)$ torus knot can be drawn so that in every direction there are less than or equal to four local maxima. From Kuiper's result and Theorem 3.3 we obtain the following result.

Corollary 4.2 *If $K_{p,q}$ is a (p, q) torus knot ($2 \leq p < q$, p and q coprime) then $\min\{2p, q\} \leq d[K_{p,q}] \leq p + q$.*

In some cases, the superbridge index of a (p, q) torus knot can actually be used to determine the knot's harmonic index. In particular, in Corollary 4.3 below we are able to find the harmonic index of for the 3_1 knot (the $(2, 3)$ torus knot), the 5_1 knot (the $(2, 5)$ torus knot), and the 8_{19} knot (the $(3, 4)$ torus knot) since appropriate low degree harmonic parametrizations exist for these knots (see section 6).

Corollary 4.3 *The harmonic index of the 3_1 knot, the 5_1 knot, and the 8_{19} knot are 3, 4, and 4, respectively.*

5 Crossing number and harmonic index

Another measure of the complexity of a knot is the knot's crossing number. In this section we will relate the crossing number of a knot to its harmonic index. Before we discuss this relationship, we recall the definition of the crossing number of a knot class. The crossing number of a knot class is the smallest crossing number of all diagrams representing knots from the given class, denoted $c[K]$.

Since the $sb[K] \leq d[K]$, and they coincide for the trivial knot, the trefoil knot, the (2, 5) and (3, 4) torus knots, the figure-eight knot, and the granny knot, one might expect that $sb[K] = d[K]$ for all knots K . However, by relating the crossing number of a knot to its superbridge index we will show this is not true. The theorem below proven in ² uses Bocher's work on Finite Fourier series ⁵ and Bezout's Theorem ⁸ to relate the harmonic index and crossing number of a knot. Using this theorem we will show that for (2, q) torus knots with q large that $sb[K_{2,q}]$ is strictly less than $d[K_{2,q}]$.

Theorem 5.1 *If K is a tame knot then*

(a) $c[K] \leq 2(d[K])^2$ and

(b) $d[K] \leq 2(c[K] + 1)^2$

In particular, if we bound either $d[K]$ or $c[K]$, we bound the other.

The superbridge index of the (2, 49) knot is 4 by ⁷. The crossing number of the (2, 49) knot is 49 (see ⁹ or ¹⁰), so by Theorem 5.1(a) above the harmonic index of the (2, 49) knot must be 5 or greater. Hence, in general for a tame knot K , $sb[K]$ is not equal to $d[K]$.

Corollary 5.2 *The superbridge index of the (2, 49) torus knot does not equal its harmonic index.*

Since the number of knot types having a given crossing number is finite (see ¹⁰), Theorem 5.1(b) also implies the following corollary.

Corollary 5.3 *The number of knot types having a given harmonic index is finite.*

6 Examples of harmonic knots

In this section we give examples of harmonic knots having minimal harmonic degree. Each parametrization was constructed by approximating the coordinate function of a polygonal knot using finite Fourier series. To get the lowest possible harmonic degree when Fourier approximating a polygonal knot, the total number of maxima and minima in most directions must be “relatively” the same. To obtain a polygonal knot with this property, an energy minimizing algorithm was applied to the polygonal knot to find a low energy conformation. This low energy conformation was another polygonal knot equivalent to the original knot which had the desired “spreading” of maxima and minima. The algorithm used was developed by Simon in ¹¹ and ¹².

After a minimum energy polygonal knot was found, the coordinate functions of this polygonal knot (expressed as 2π periodic, continuous, piecewise linear functions) were approximated using finite Fourier series. The resulting harmonic curve was then checked to be of the same knot type as the original knot by graphing using Maple ¹³. If the harmonic curve was of a different type, a higher degree harmonic curve was tried. Sometimes when using this technique to find harmonic knots, we iterated the process passing from minimum energy polygons to Fourier approximations of polygons to inscribed polygons (perhaps with more edges). Harmonic parametrizations for all knots through eight crossings are available in ² and ³.

Example 6.1 5_1 Knot – (2, 5) torus knot

$$\begin{aligned}x(t) &= 88\cos(t) + 115\sin(t) - 475\cos(2t) - 127\sin(2t) - 87\cos(3t) + 36\sin(3t) + \\ &\quad 11\cos(4t) - 19\sin(4t) \\ y(t) &= 89\cos(t) - 32\sin(t) - 172\cos(2t) + 294\sin(2t) + 76\cos(3t) + 102\sin(3t) - \\ &\quad 61\cos(4t) + 113\sin(4t) \\ z(t) &= 44\cos(t) - 69\sin(t) + 34\cos(2t) + 223\sin(2t) + 16\cos(3t) + 120\sin(3t) + \\ &\quad 42\cos(4t) - 125\sin(4t)\end{aligned}$$

Example 6.2 8_{19} Knot – (3, 4) torus knot

$$\begin{aligned}x(t) &= -26\cos(t) + 6\sin(t) + 8\cos(2t) + 12\sin(2t) + 49\cos(3t) - 68\sin(3t) - \\ &\quad 36\cos(4t) - 13\sin(4t) \\ y(t) &= 23\cos(t) + 66\sin(t) + 2\sin(2t) + 77\cos(3t) + 67\sin(3t) - 9\cos(4t) - 3\sin(4t) \\ z(t) &= -15\cos(t) + 16\sin(t) - 11\cos(2t) - 15\sin(2t) + 47\cos(3t) - 37\sin(3t) + \\ &\quad 51\cos(4t) + 19\sin(4t)\end{aligned}$$

Example 6.3 Granny Knot

$$x(t) = -22\cos(t) - 128\sin(t) - 44\cos(3t) - 78\sin(3t)$$

$$y(t) = -10\cos(2t) - 27\sin(2t) + 38\cos(4t) + 46\sin(4t)$$

$$z(t) = 70\cos(3t) - 40\sin(3t)$$

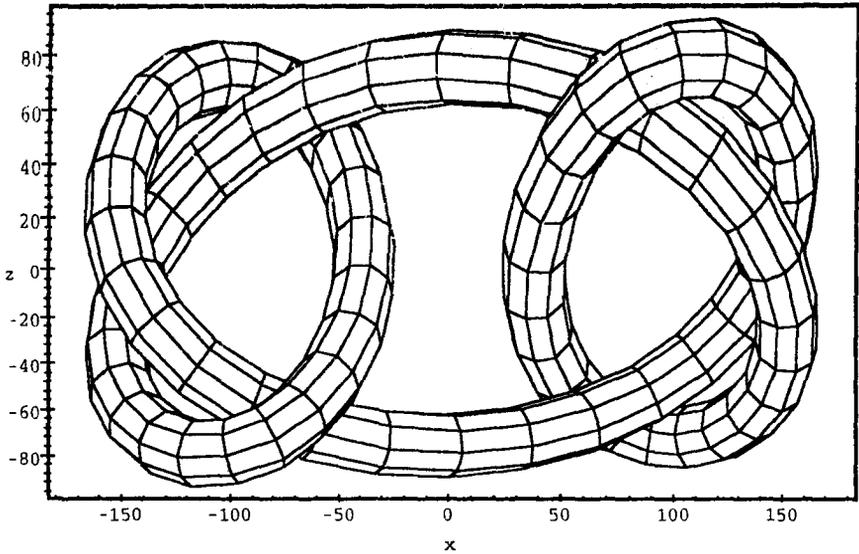


Figure 3: Harmonic granny knot.

7 Acknowledgements

The research for this article was completed while the author was a graduate student at the University of Iowa. The author is grateful to his thesis advisor Jonathan Simon for his help with this research and for the use of his polygonal knot energy program which was instrumental in obtaining conformations of knots with minimal harmonic degree.

8 References

1. M. Bogle, J. Hearst, V. Jones, and L. Stoilov, *Lissajous knots*, preprint, 1994.
2. A. Trautwein, *Harmonic Knots*, preprint, 1997 (also available as *Harmonic Knots*, Ph.D. thesis, University of Iowa, 1995).
3. A. Trautwein, *Harmonic Knot Research*, <http://www.carthage.edu/~trautwn>.

4. J. Milnor, *On the total curvature of knots*, *Annals of Mathematics*, Volume 52, 1950, 248–257.
5. M. Boucher, *Introduction to the Theory of Fourier Series*, *Annals of Mathematics*, Volume 7, 1906, 81–162.
6. H. Schubert, *Über eine numerische Knoteninvariante*, *Math. Z.*, Volume 61, 1954, 245–288.
7. N. Kuiper, *A New Knot Invariant*, *Mathematische Annalen*, Volume 278, 1976, 193–209.
8. J. Harris, *Algebraic Geometry, A First Course*, Springer–Verlag, New York, 1992.
9. G. Burde and H. Zieschang, *Knots*, Walter de Gruyter Publisher, Berlin and New York, 1985.
10. D. Rolfsen, *Knot and Links*, Publish or Perish Press, Berkeley, 1976.
11. J. Simon, *Energy functions for polygonal knots*, *Journal of Knot Theory and its Ramifications*, Volume 3, 1985, 299–320.
12. J. Simon, *Energy functions for knots: beginning to predict physical behavior*, preprint, 1995.
13. Symbolic Computation Group, Department of Computer Science, University of Waterloo, *Maple V – Release 3*, Waterloo Maple Software, 1994.

CHAPTER 19

Fourier Knots

Louis H. Kauffman

*Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
851 South Morgan Street
Chicago, IL 60607-7045*

This paper introduces the concept of Fourier knot. A Fourier knot is a knot that is represented by a parametrized curve in three dimensional space such that the three coordinate functions of the curve are each finite Fourier series in the parameter. That is, the knot can be regarded as the result of independent vibrations in each of the coordinate directions with each of these vibrations being a linear combination of a finite number of pure frequencies.

1 Introduction

The previously studied Lissajous knots [1] constitute the case of a single frequency in each coordinate direction. Not all knots are Lissajous knots, and in fact the trefoil knot and the figure eight knot are the first examples of non-Lissajous knots. The first section of this paper sketches the proof that every tame knot is a Fourier knot. Subsequent sections give robust examples of Fourier representations for the trefoil, the figure eight and a class of knots that we call Fibonacci knots. In the case of the trefoil we have given a minimal Fourier representation in the sense that it has single frequencies in two of the coordinate directions and a combination of frequencies in the third direction. The paper ends by pointing out the usual compact non-linear trigonometric formula for torus knots, and raises the question of the finite Fourier series representations for these knots.

On completing an early draft of this paper, we learned that an extensive study of Fourier knots (there called Harmonic Knots) has been carried out by Aaron Trautwein in his 1995 PhD Thesis [7] at the University of Iowa under the direction of Jon Simon. While the independently obtained results of the present paper are primarily illustrative of the idea of Fourier knots, Trautwein's pioneering work establishes relationships between the complexity of the harmonic representation and knot theoretic indices such as crossing number and superbridge index. The interested reader should consult this work.

Acknowledgment. The author thanks the National Science Foundation for support of this research under grant number DMS-9205277.

2 Every Knot is a Fourier Knot

In considering problems about knots it is interesting to have an equation for the knot or link under consideration. By an equation for a knot I mean the specification of a parametrized curve in three dimensional space of the form

$$X(t) = A(t)$$

$$Y(t) = B(t)$$

$$Z(t) = C(t)$$

where A, B, C are smooth functions of the variable t with (for a knot) a total period of $P > 0$, specifying an embedding of the circle into three dimensional space. Here the circle is the quotient space of the interval from 0 to P , $[0, P]$, obtained by identifying the ends of the interval to each other with the quotient topology.

A function $F(t)$ is said to be a finite Fourier series if it has the form

$$F(t) = A_1 \text{Cos}(K_1 T + L_1) + \dots + A_N \text{Cos}(K_N T + L_N)$$

where A_1, A_2, \dots, A_N and K_1, K_2, \dots, K_N and $L_1, L_2, L_3, \dots, L_N$ are given constants and K_1, K_2, \dots, K_N are each rational numbers. Of course, $F(t)$ can also be expressed in terms of the *Sin* function or as a combination of *Sin* and *Cos* functions, since $\text{Sin}(X + \pi/2) = \text{Cos}(X)$.

Definition. A knot K embedded in three dimensional space will be called a *Fourier knot* if it has an equation (as described above) with each of the functions A, B, C a finite Fourier series.

Definition. A knot K is said to be *tame* if every point p of K in R^3 (Euclidean three-space) has a neighborhood such that the intersection of K with that neighborhood is equivalent to a standard pair of three-dimensional ball and diameter-arc of that ball.

It is well known that every tame knot can be represented topologically by equations where A, B and C are smooth (i.e. infinitely differentiable) functions. It then follows by standard approximation theorems for Fourier series that A, B and C can be taken to be finite Fourier series. Thus we have proved the

Theorem. Every tame knot is (topologically equivalent to) a Fourier knot.

3 Lissajous Knots and the Arf Invariant

One class of Fourier knots that have been studied are the Lissajous knots. In a Lissajous knot there is one term in each of the Fourier series. Thus a Lissajous knot has the form

$$X(t) = A_1 \text{Cos}(K_1 t + L_1)$$

$$Y(t) = A_2 \text{Cos}(K_2 t + L_2)$$

$$Z(t) = A_3 \text{Cos}(K_3 t + L_3)$$

In it is proved that the Arf invariant of a Lissajous knot is necessarily equal to zero. This means that many knots are not Lissajous [1]. In particular the trefoil knot and the figure eight knot are not Lissajous knots. This leads to the question: if a tame knot K is not Lissajous, what is the "simplest" representation of K in terms of finite Fourier series?

In [1] the next section we shall give a definite answer to this question in the case of the trefoil knot, and a conjecture in the case of the figure eight knot. In all cases, when one answers this question there is also the parallel question of obtaining Fourier equations for the knot K that are robust in the sense that plots of these equations yield pleasing images that can be explored geometrically.

Since knots of non-zero Arf invariant are necessarily not Lissajous, it will be useful for us to recall one definition of the Arf invariant. An interested reader can apply this definition to find examples of Fourier knots that are not Lissajous knots. There are algebraic definitions of the Arf invariant and a fundamental geometric definition as well. See [2] or [3] for more details.

We recall an algebraic definition of the Arf invariant by first defining an integer valued invariant, $a(K)$, associated to any oriented knot K . The invariant $a(K)$ is defined by the (recursive) equation (*)

$$a(K_+) = a(K_-) = Lk(K_0)$$

where K_+ , K_- and K_0 are three diagrams that differ at a single crossing as shown in Figure 1, and $Lk(K_0)$ denotes the linking number of the link of two components K_0 . K_+ and K_- are each knot diagrams, differing from each other by a single switched crossing. K_0 is obtained from either K_+ or K_- by smoothing that crossing. A *smoothing* is accomplished by reconnecting the strands at the crossing so that the arcs no longer cross over one another (as

shown in Figure 1). Both the switching and the smoothing operations can change the topological type of the diagram. Smoothing always replaces a knot by a link of two components. Thus K_0 is a such a link. By definition, $a(K)$ is equal to zero if K is topologically equivalent to an unknotted circle.

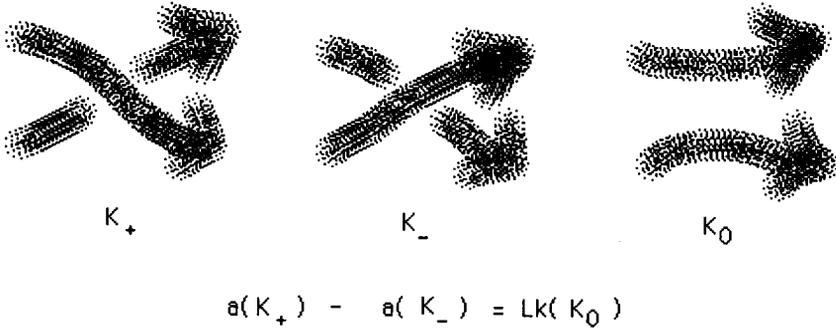


Figure 1: Switching Relation

Note that in Figure 1 we have implicitly assigned signs of +1 or -1 to the two types of oriented crossings. This number, +1 or -1, is called the sign of the crossing. The linking number of a link L is defined by the equation

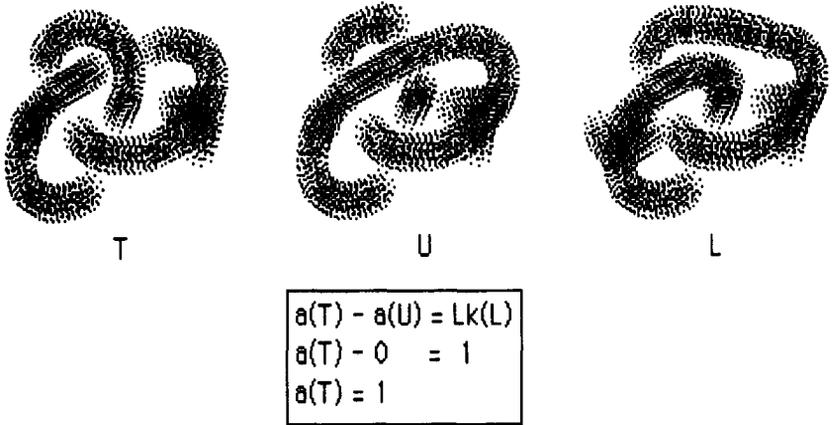
$$Lk(L) = \sum_p e(p)/2$$

where the summation runs over all the crossings in K that are between two components of K . Crossings of any given component with itself are not counted in this summation.

It is a (non-obvious) fact that the recursive equation (*) defines a topological invariant of knots and links. It is, in fact part of a much larger scheme of things. For example it is the second coefficient of the Conway polynomial. One way to define the Arf invariant, $Arf(K)$, is by the equation

$$Arf(K) = a(K)(mod 2).$$

Thus the Arf invariant of K is either 0 or 1 depending upon the parity of $a(K)$. See Figure 2 for a sample calculation of the Arf invariant of the trefoil knot.

Figure 2: $a(\text{Trefoil})$

It is a remarkable fact that Lissajous knots have Arf invariant zero. I do not know if every knot of vanishing Arf invariant is a Lissajous knot.

4 A Fourier Trefoil Knot

Consider the following equations

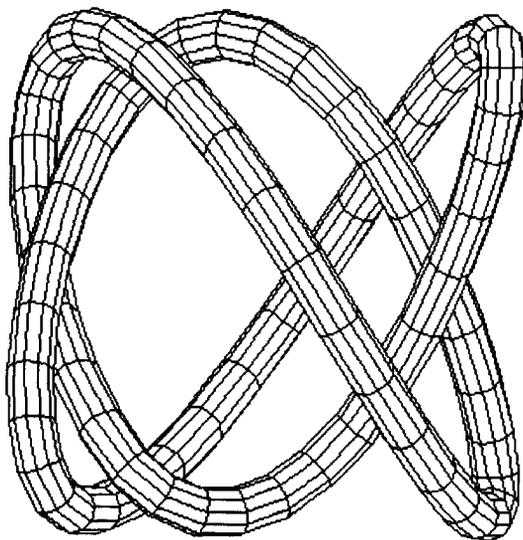
$$x = \text{Cos}(2T),$$

$$y = \text{Cos}(3T + (1/2)),$$

$$z = (1/2)\text{Cos}(5T + (1/2)) + (1/2)\text{Sin}(3T + (1/2)).$$

These equations define a trefoil knot, showing that the trefoil knot is a Fourier knot where only one coordinate needs to be a combination of frequencies. The proof that these equations give a trefoil knot is left to the reader. One way to verify this is to use a computer to draw the pictures in three dimensions and then examine the results. Figure 3 illustrates a computer drawing of this Fourier trefoil. The drawing illustrates what I mean by a robust representation of the knot. The knot does not come ambiguously close to itself, and the form of the drawing is aesthetically pleasing. The author wishes to acknowledge Lynnclaire Dennis [4] for inspiring him to search for the Fourier trefoil. In her book "The Pattern" Ms. Dennis draws a picture of a knot (the Pattern knot)

that closely resembles our Fourier trefoil. In projection the Pattern knot looks like a Lissajous figure with frequencies 2 and 3 and the Pattern knot is a trefoil knot. This led to trying combinations of frequencies for the third coordinate, and eventually to the equations above with pure frequencies 2 and 3 in two directions and the combination of frequencies 5 and 3 in the third direction. The Pattern knot is more spherically symmetrical than the Fourier trefoil, and does not have an obvious equation.

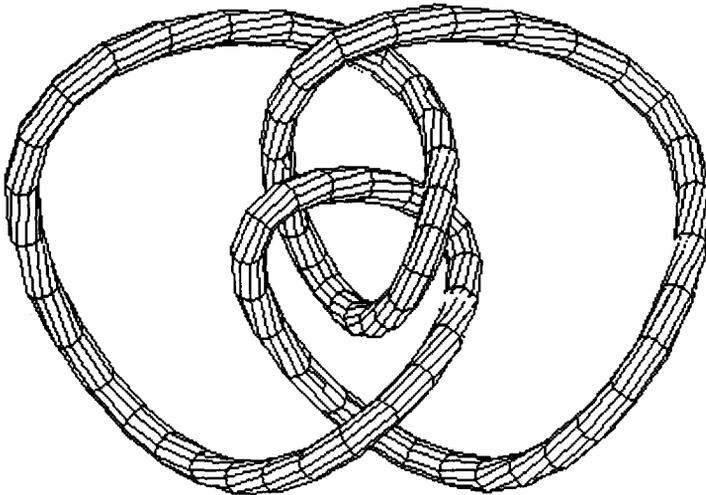


$$\begin{aligned}
 x &= \cos(2*t) \\
 y &= \cos(3*t + 1/2) \\
 z &= (1/2)(\cos(5*t + 1/2) + \sin(3*t + 1/2))
 \end{aligned}$$

Figure 3: The Fourier Trefoil

I would also like to mention an experiment that I performed with the Fourier trefoil in the form of a (hand-drawn) diagram corresponding to the knot in Figure 3. I gave this diagram as input to Ming, a knot energy program written by Ying-Quing Wu at the University of Iowa. (A diagrammatic interface for Ming was written by Milana Huang at the Electronic Visualisation Lab at the University of Illinois). Ming sets the knot on a descending energy trajec-

tory, following Jon Simon's energy for the knot. The result of this experiment is that the flat knot diagram quickly unfurls into a three dimensional geometry very similar to the Fourier trefoil and nearly stabilizes in this form. Then slowly the knot moves off this slightly higher energy level and settles into the familiar symmetry of the (empirically) known energy minimum for the trefoil knot. Thus there appears to be a "point of inflection" in this particular way of descending to minimum energy for the trefoil knot. This experiment points to a wide range of possible explorations, investigating the gradient descent for knot energy from particular starting configurations for a knot.



$$X = 10 * \cos(T) + 10 * \cos(3 * T)$$

$$Y = 6 * \sin(T) + 10 * \sin(3 * T)$$

$$Z = 4 * \sin(3 * T) - 10 * \sin(6 * T)$$

Figure 4: The Fourier Figure Eight Knot

5 A Fourier Figure Eight Knot

The following equations describe a figure eight knot.

$$\begin{aligned}x &= \text{Cos}(t) + \text{Cos}(3t), \\y &= .6\text{Sin}(t) + \text{Sin}(3t), \\z &= .4\text{Sin}(3t) - \text{Sin}(6t).\end{aligned}$$

See Figure 4.

I do not know if there is a simpler Fourier representation for this knot.

6 A Series of Fibonacci Fourier Knots

Recall the Fibonacci series

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \dots$$

The n -th term, f_n , of the series is equal to the sum of the previous two terms, and $f_1 = f_2 = 1$. Consider the equations

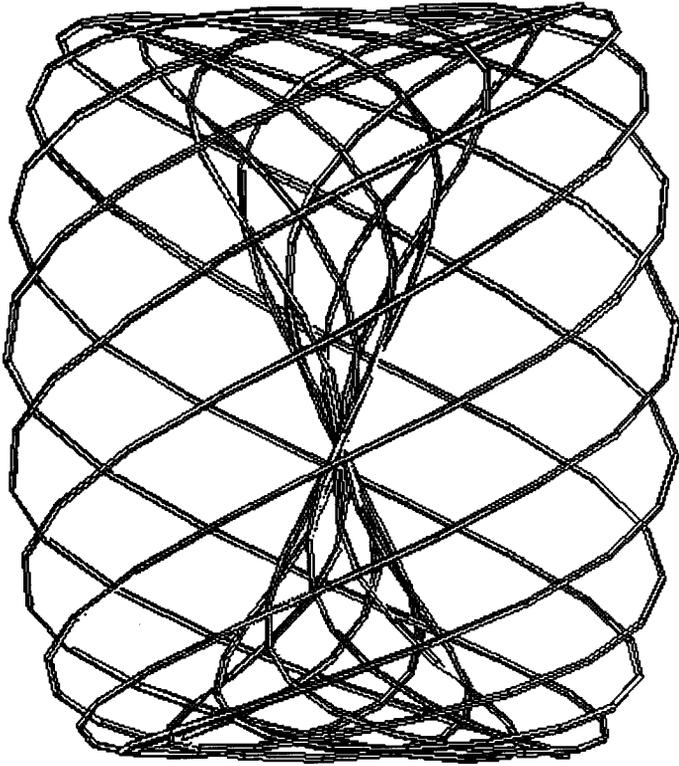
$$\begin{aligned}x &= \text{Cos}(8T), \\y &= \text{Cos}(13T + .5), \\z &= .5\text{Cos}(21T + .5) + .5\text{Sin}(13T + .5),\end{aligned}$$

and more generally

$$\begin{aligned}x &= \text{Cos}(f_n T), \\y &= \text{Cos}(f_{n+1} T + .5), \\z &= .5\text{Cos}(f_{n+2} T + .5) + .5\text{Sin}(f_{n+1} T + .5).\end{aligned}$$

The last set of equations defines a knot that we shall dub $F(n)$, the n -th Fibonacci knot. Thus $F(3)$ is the Fourier trefoil of Section 3, and the first equations we have written in this section denote $F(6)$. In Figure 5 we illustrate a computer drawing of $F(6)$.

The Fibonacci Fourier knots provide a strong class of knots for investigation using computer graphics.



$$\begin{aligned} X &= \cos(8 \cdot T) \\ Y &= \cos(13 \cdot T + .5) \\ Z &= .5 \cdot \cos(21 \cdot T + .5) + .5 \cdot \sin(13 \cdot T + .5) \end{aligned}$$

Figure 5: A Fibonacci Fourier Knot

7 Torus Knots

Recall that a knot that winds P times around a torus in one direction and Q times in the other direction - a torus knot of type (P, Q) - has the equation

$$\begin{aligned}x &= \text{Cos}(T)(1 + .5\text{Cos}((Q/P)T)), \\y &= \text{Sin}(T)(1 + .5\text{Cos}((Q/P)T)), \\z &= .5\text{Sin}((Q/P)T).\end{aligned}$$

Now use the trigonometric identities

$$\begin{aligned}\cos(a)\cos(b) &= .5(\cos(a + b) + \cos(a - b)), \\ \sin(a)\cos(b) &= .5(\sin(a + b) + \sin(a - b)).\end{aligned}$$

The equations above then become

$$\begin{aligned}x &= \cos(T) + .25\cos((1 + Q/P)T) + .25\cos((1 - Q/P)T) \\y &= \sin(T) + .25 * \sin((1 + Q/P)T) + .25 * \sin((1 - Q/P)T) \\z &= .5\sin((Q/P)T)\end{aligned}$$

Thus torus knots are Fourier knots, and we can ask if these are simplest Fourier representations for torus knots. The parametrization shown above appears in [6]. I am indebted to Peter Roegen for pointing this out.

References

1. Bogle, Hearst, Jones and Stoilov, Lissajous Knots, *Journal of Knot Theory and Its Ramifications*, Vol.3, No. 2 (June 1994), pp. 121- 140.
2. L. Kauffman, *Formal Knot Theory*, Princeton Lecture Notes 30, Princeton University Press (1983).
3. L. Kauffman, *On Knots*, Princeton Annals Vol. 115, Princeton University Press (1987).
4. L. Dennis., *The Pattern*, Integral Press, Lower Lake, CA (1997).
5. G. Buck and J. Simon, Knots as dynamical systems, *Topology Appl.*, 51, 1993.
6. Sueli I. Rodriques Costa, On closed twisted curves, *Proc. Amer. Math. Soc.*, 109, No. 1 (1990).
7. Aaron Trautwein, *Harmonic Knots*, PhD Thesis, July 1995, Department of Mathematics, University of Iowa, Iowa City, Iowa.

CHAPTER 20

Symmetry is a vast subject, significant in art and nature. Mathematics lies at its roots, and it would be hard to find a better one on which to demonstrate the working of the mathematical intellect. [Hermann Weyl, Symmetry [We]]

Symmetric knots and billiard knots.

by Józef H. Przytycki

Symmetry of geometrical figures is reflected in regularities of their algebraic invariants. Algebraic regularities are often preserved when the geometrical figure is topologically deformed. The most natural, intuitively simple but mathematically complicated, topological objects are *knots*.

We present in this papers several examples, both old and new, of regularity of algebraic invariants of knots. Our main invariants are the Jones polynomial (1984) and its generalizations.

In the first section, we discuss the concept of a symmetric knot, and give one important example – a torus knot. In the second section, we give review of the Jones type invariants. In the third section, we gently and precisely develop the periodicity criteria from the Kauffman bracket (ingenious version of the Jones polynomial). In the fourth section, we extend the criteria to skein (Homflypt) and Kauffman polynomials. In the fifth section we describe r^q periodicity criteria using Vassiliev-Gusarov invariants. We also show how the skein method may be used for r^q periodicity criteria for the classical (1928) Alexander polynomial. In the sixth section, we introduce the notion of Lissajous and billiard knots and show how symmetry principles can be applied to these geometric knots. Finally, in the seventh section, we show how symmetry can be used to gain nontrivial information about knots in other 3-manifolds, and how symmetry of 3-manifolds is reflected in manifold invariants.

1 Symmetric knots.

We analyze, in this paper, symmetric knots and links, that is, links which are invariant under a finite group action on S^3 or, more generally, a 3-manifold.

For example, a torus link of type (p, q) (we call it $T_{p,q}$) is preserved by an action of a group $Z_p \oplus Z_q$ on S^3 (c.f. Fig. 1.1 and Example 1.1).

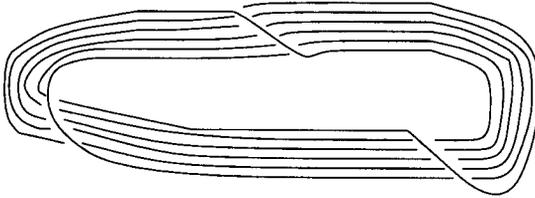


Fig. 1.1. The torus link of type $(3, 6)$, $T_{3,6}$

Example 1.1 Let $S^3 = \{z_1, z_2 \in C \times C : |z_1|^2 + |z_2|^2 = 1\}$. Let us consider an action of $Z_p \oplus Z_q$ on S^3 which is generated by T_p and T_q , where $T_p(z_1, z_2) = (e^{2\pi i/p} z_1, z_2)$ and $T_q(z_1, z_2) = (z_1, e^{2\pi i/q} z_2)$. Show that this action preserves torus link of type (p, q) . This link can be described as the following set $\{(z_1, z_2) \in S^3 : z_1 = e^{2\pi i(\frac{t}{p} + \frac{k}{p})}, z_2 = e^{2\pi i t/q}\}$, where t is an arbitrary real number and k is an arbitrary integer.

If p is co-prime to q then $T_{p,q}$ is a knot and can be parameterized by:

$$R \ni t \mapsto (e^{2\pi i t/p}, e^{2\pi i t/q}) \in S^3 \subset C^2.$$

In this case $Z_p \oplus Z_q = Z_{pq}$ with a generator $T = T_p T_q$.

Subsequently, we will focus on the action of a cyclic group Z_n . We will mainly consider the case of an action on S^3 with a circle of fixed points. The new link invariants, (see Section 2), provide efficient criteria for examining such actions.

Definition 1.2 A link is called n -periodic if there exists an action of Z_n on S^3 which preserves the link¹ and the set of fixed points of the action

¹More precisely we require an existence of an embedding of circles, realizing the link, which is equivariant under the group action.

is a circle disjoint from the link. If, moreover, the link is oriented then we assume that a generator of Z_n preserves the orientation of the link or changes it globally (that is on every component).

2 Polynomial invariants of links

We will describe in this section polynomial invariants of links, crucial in periodicity criteria. We start from the skein (Jones-Conway or Homflypt) polynomial, $P_L(v, z) \in Z[v^{\pm 1}, z^{\pm 1}]$ (i.e. $P_L(v, z)$ is a Laurent polynomial in variables v and z).

Definition 2.1 *The skein polynomial invariant of oriented links in S^3 can be characterized by the recursive relation (skein relation):*

- (i) $v^{-1}P_{L_+}(v, z) - vP_{L_-}(v, z) = zP_{L_0}(v, z)$, where L_+, L_- and L_0 are three oriented link diagrams, which are the same outside a small disk in which they look as in Fig. 2.1,

and the initial condition

- (ii) $P_{T_1} = 1$, where T_1 denotes the trivial knot.

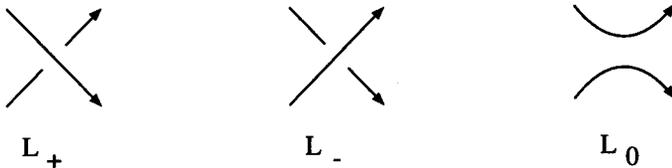


Fig. 2.1

We need some elementary properties of the skein polynomial:

Proposition 2.2 (a) $([L-M]) z^{com(L)-1} P_L(v, z) \in Z[v^{\pm 1}, z]$ where $com(L)$ is the number of link components of L . That is we do not use negative powers of z . Furthermore the constant term, with respect to variable z , is non-zero.

- (b) $P_L(v, z) - P_{T_{com(L)}}$ is divisible by $(v^{-1} - v)^2 - z^2$. Here T_i denotes the trivial link of i components.

- (c) Let $\hat{P}_L(v, z) = z^{\text{com}(L)-1} P_L(v, z)$, then $\hat{P}_L(v, z) - (v^{-1} - v)^{\text{com}(L)-1}$ is divisible by $(v^{-1} - v)^2 - z^2$
- (d) $P_L(v, z) \equiv (v^3 z)^j P_{T_k}(v, z) \pmod{(\frac{v^6-1}{v^2-1}, z^2 + 1)}$ for some j and k where $j \equiv (k - \text{com}(L)) \pmod 2$. In particular $P_L \equiv \delta P_{\bar{L}} \pmod{(\frac{v^6-1}{v^2-1}, z^2 + 1)}$ where δ equals to 1 or -1 and \bar{L} denotes the mirror image of L .

Proof:

- (a)-(c) Initial conditions and recurrence relation give immediately that $\hat{P}_L(v, z) \in Z[v^{\pm 1}, z]$. (c) holds for trivial links from the definition (the difference is zero). To make inductive step, first notice that $\hat{P}_L(v, z)$ satisfies the skein relation $v^{-1} \hat{P}_{L_+}(v, z) - v \hat{P}_{L_-}(v, z) = z^{2\epsilon} \hat{P}_{L_0}(v, z)$ where ϵ is equal to 0 in the case of the selfcrossing of L_+ and $\epsilon = 1$ in the mixed crossing case. We can rewrite the skein relation so that the inductive step is almost obvious:

$$v^{-1}(\hat{P}_{L_+} - (v^{-1} - v)^{\text{com}L_+ - 1}) - v(\hat{P}_{L_-} - (v^{-1} - v)^{\text{com}L_- - 1}) = z^{2\epsilon}(\hat{P}_{L_0} - (v^{-1} - v)^{\text{com}L_0 - 1}) + (v^{-1} - v)^{\text{com}L_+ - 2\epsilon}(v^{-1} - v)^{2\epsilon} - z^{2\epsilon}$$

Namely the last term in the equality is always divisible by $(v^{-1} - v)^2 - z^2$ therefore if two other are divisible then the last one is. We completed the proof of (c). From (c) follows that $\hat{P}_L(1, 0) = 1$, thus the second part of (a) follows. (b) is just a weaker version of (c).

- (d) It can be derived from the relation of $P_L(e^{2s\pi i/6}, \pm 1)$ and the first homology group (modulo 3) of the 2-fold branched cyclic cover of (S^3, L) [L-M, Yo-1, Ve].

□

Definition 2.3 Let L be an unoriented diagram of a link. Then the Kauffman bracket polynomial $\langle L \rangle \in Z[A^{\mp 1}]$ is defined by the following properties:

1. $\langle \bigcirc \rangle = 1$

$$2. \langle \bigcirc \sqcup L \rangle = -(A^2 + A^{-2})\langle L \rangle$$

$$3. \langle \times \rangle = A\langle \smile \rangle + A^{-1}\langle \frown \rangle$$

The Kauffman bracket polynomial is a variant of the Jones polynomial for oriented links. Namely, for $A = t^{-\frac{1}{4}}$ and D being an oriented diagram of L we have

$$V_L(t) = (-A^3)^{-w(D)} \langle D \rangle \tag{1}$$

where $w(D)$ is the *planar writhe* (twist or Tait number) of D equal to the algebraic sum of signs of crossings.

Kauffman noted that $V_L(t) = P_L(t, \sqrt{t} - \frac{1}{\sqrt{t}})$. In particular $t^{-1}V_{L_+} - tV_{L_-} = (\sqrt{t} - \frac{1}{\sqrt{t}})V_{L_0}$. Proposition 2.2 gives us:

Corollary 2.4 (a) [Jo]. For a knot K , $V_K(t) \in Z[t^{\pm 1}]$ and $V_K(t) - 1$ is divisible by $(t - 1)(t^3 - 1)$.

(b) [Yo-1]. For a knot K , $V_K(t) - \delta V_K(t^{-1})$ is divisible by $\frac{t^3+1}{t+1}$, where δ equals to 1 or -1 .

In the summer of 1985 (two weeks before discovering the “bracket”), L. Kauffman invented another invariant of links [Ka], $F_L(a, z) \in Z[a^{\pm 1}, z^{\pm 1}]$, generalizing the polynomial discovered at the beginning of 1985 by Brandt, Lickorish, Millett and Ho [B-L-M, Ho]. To define the Kauffman polynomial we first introduce the polynomial invariant of link diagrams $\Lambda_D(a, z)$. It is defined recursively by:

(i) $\Lambda_o(a, z) = 1,$

(ii) $\Lambda_{\smile}(a, z) = a\Lambda_+(a, z); \Lambda_{\frown}(a, z) = a^{-1}\Lambda_-(a, z),$

(iii) $\Lambda_{D_+}(a, z) + \Lambda_{D_-}(a, z) = z(\Lambda_0(a, z) + \Lambda_{D_\infty}(a, z)).$

The Kauffman polynomial of oriented links is defined by

$$F_L(a, z) = a^{-w(D)} \Lambda_D(a, z)$$

where D is any diagram of an oriented link L .

Remark 2.5 Let $F^*(a, z)$ be the Dubrovnik variant of the Kauffman polynomial ². The polynomial F^* satisfies initial conditions $F_{T_n}^* = (\frac{a-a^{-1}}{z} + 1)^{n-1}$ and the recursive relation $a^{w(D_+)} F_{D_+}^* - a^{w(D_-)} F_{D_-}^* = z(a^{w(D_0)} F_{D_0}^* - a^{w(D_\infty)} F_{D_\infty}^*)$. Lickorish noted that the Dubrovnik polynomial is just a variant of the Kauffman polynomial:

$$F_L^*(a, z) = (-1)^{com(L)-1} F_L(ia, -iz), ([Li]).$$

Proof.

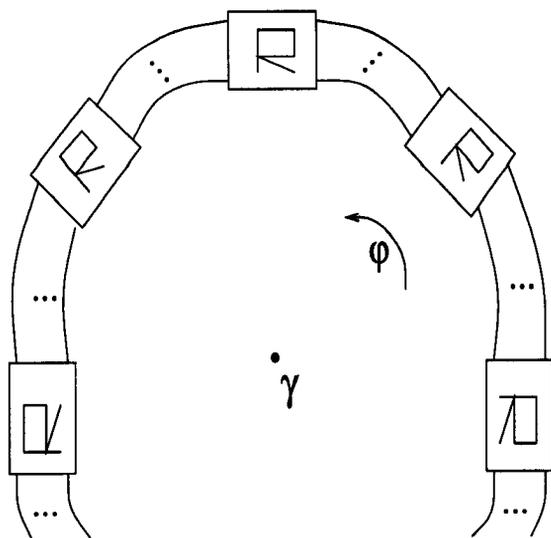
We check the formula for trivial links and then show that if it holds for three terms of the skein relation it also holds for the fourth one.

- (i) $F_{T_n}(ia, -iz) = (-1)^{(n-1)} (\frac{a-a^{-1}}{-z} - 1)^{n-1} = F_{T_n}^*(a, z)$.
- (ii) $(ia)^{w(D_+)} F_{D_+}(ia, -iz) + (ia)^{w(D_-)} F_{D_-}(ia, -iz) = (-iz)((ia)^{w(D_0)} F_{D_0} + (ia)^{w(D_\infty)} F_{D_\infty})$. This reduces to $a^{w(D_+)} F_{D_+}^* - a^{w(D_-)} F_{D_-}^* = z(a^{w(D_0)} F_{D_0}^* + (-1)^{com(D_0)-com(D_\infty)} a^{w(D_\infty)} F_{D_\infty}^*)$, which gives the skein relation for $F_D^*(a, z)$.

3 Periodic links and the Jones polynomial.

Periodicity of links is reflected in the structure of new polynomials of links. We will describe this with details in the case of the Jones polynomial (and the Kauffman bracket), mostly following [Mu-2, T-1, P-3, Yo-1] but also proving new results. Let D be an r -periodic diagram of an unoriented r -periodic link, that is $\varphi(D) = D$ where φ denote the rotation of R^3 along the vertical axis by the angle $2\pi/r$. In the coordinates of R^3 given by a complex number (for the first two real co-ordinates) and a real number (for the third) one gets: $\phi(z, t) = (e^{2\pi i/r} z, t)$. φ is a generator of the group Z_r acting on R^3 (and $S^3 = R^3 \cup \infty$). In particular $\varphi^r = Id$. See Fig. 3.1.

²Kauffman described the polynomial F^* on a postcard to Lickorish sent from Dubrovnik in September '85.



...

Fig. 3.1

By the positive solution to the Smith Conjecture [Sm, Thur], every r -periodic link has an r -periodic diagram, thus we can restrict ourselves to considerations of these, easy to grasp, diagrams. With the help of elementary group theory we have the following fundamental lemma.

Lemma 3.1 *Let D be an unoriented r -periodic link diagram, r prime. Then the Kauffman bracket polynomial satisfies the following “periodic” formula:*

$$D_{sym}(\times) \equiv A^r D_{sym}(\smile) + A^{-r} D_{sym}(\frown) \pmod{r}$$

where $D_{sym}(\times)$, $D_{sym}(\smile)$ and $D_{sym}(\frown)$ denote three φ -invariant diagrams of links which are the same outside of the Z_r -orbit of a neighborhood of a fixed single crossing, c (i.e. $c, \varphi(c), \dots, \varphi^{r-1}(c)$) at which they differ by replacing \times by \smile or \frown , respectively.

Proof: Let us build the binary computational resolving tree of D using the Kauffman bracket skein relation for every crossing of the orbit

(under Z_r action), $c, \varphi(c), \dots, \varphi^{r-1}(c)$. The tree has therefore 2^r leaves and a diagram at each leaf contributes some value (polynomial) to the Kauffman bracket polynomial of $D = D_{sym}(\text{X})$, see Fig. 3.2. The idea of the proof is that only extreme leaves, $D_{sym}(\text{~})$ and $D_{sym}(\text{~})$, contribute to $\langle D \rangle \pmod r$.

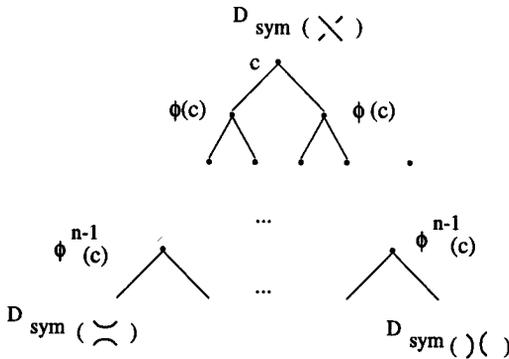


Fig. 3.2

We need now some, elementary, group theory:

Let a finite group G acts on a set X , that is every element $g \in G$ “moves” X ($g : X \rightarrow X$) and $hg(x) = h(g(x))$ for any $x \in X$ and $h, g \in G$. Furthermore we require that the identity element of G is not moving X ($e(x) = x$ where e is the identity element of G). The orbit of an element x_0 in X is the set of all elements of X which can be obtained from x_0 by acting on it by G ($\mathcal{O}_{x_0} = \{x \in X \mid \text{there is } g \text{ such that } g(x_0) = x\}$). The standard, but important, fact of elementary group theory is that the number of elements in an orbit divides the order (number of elements) of the group. In particular if the group is equal to Z_r , r prime, then orbits of Z_r action can have one element (such orbits are called fixed points) or r elements.

After this long group theory digression, we go back to our leaf diagrams of the binary computational resolving tree of D . We claim that Z_r acts on the leaf diagrams and the only fixed points of the action are extreme leaves $D_{sym}(\text{~})$ and $D_{sym}(\text{~})$. All other orbits have r elements and

they would cancel their contribution to $\langle D \rangle$ modulo r . To see these we introduce an adequate notation: Let $c_i = \varphi(c)$ and $D_{s_0, \dots, s_{r-1}}^{c_0, \dots, c_{r-1}}$, where $s_i = \smile$ or \frown , denote the diagram of a link obtained from D by smoothing the orbit of crossings, c_0, \dots, c_{r-1} according to indices s_i . $D_{s_0, \dots, s_{r-1}}^{c_0, \dots, c_{r-1}}$ are leaves of our binary computational resolving tree of D , and the Z_r action can be fully described by the action on the indices s_0, \dots, s_{r-1} . Namely $\varphi(s_0, \dots, s_{r-2}, s_{r-1}) = (s_{r-1}, s_0, \dots, s_{r-2})$. From this description it is clear that the only fixed point sequences are $(\smile, \dots, \smile, \smile)$ and $(\frown, \dots, \frown, \frown)$ and that diagrams from the given orbit represent equivalent (ambient isotopic) links, they just differ by the rotation of R^3 . From this follows that the contribution to $\langle D \rangle$ of an r element orbit is equal to 0 modulo r . Thus all leaves, but the fixed points, do not contribute to $\langle D \rangle$ modulo r . The contribution of the fixed point leaves is expressed in the formula of Lemma 3.1. \square

Corollary 3.2 *Let D° be an oriented r -periodic link diagram, r prime, and D the same diagram, orientation forgotten.*

(i) *Then the Kauffman bracket polynomial satisfies the following formula:*

$$A^r \langle D_{sym(\smile)} \rangle - A^{-r} \langle D_{sym(\frown)} \rangle \equiv (A^{2r} - A^{-2r}) \langle D_{sym(\smile)} \rangle \pmod{r}$$

(ii) *Let $f_{D^\circ}(A) = (-A^3)^{-w(D^\circ)} \langle D \rangle$ then*

$$-A^{4r} f_{D^\circ_{sym(+)}} + A^{-4r} f_{D^\circ_{sym(-)}} \equiv (A^{2r} - A^{-2r}) f_{D^\circ_{sym(0)}} \pmod{r}$$

Here $D^\circ_{sym(+)} = D^\circ_{sym(\nearrow)}$, $D^\circ_{sym(-)} = D^\circ_{sym(\nwarrow)}$ and $D^\circ_{sym(0)} = D_{sym(\bowtie)}$ denote three φ -invariant oriented diagrams of links which are the same outside of the Z_r -orbit of a fixed single crossing and which at a neighborhood of the crossing differ by replacing \nearrow by \nwarrow or \bowtie ;

(iii)

$$t^{-r}V_{D_{sym(+)}^{\circ}} - t^rV_{D_{sym(-)}^{\circ}} \equiv (t^{r/2} - t^{-r/2})V_{D_{sym(0)}^{\circ}} \pmod{r}$$

Proof:

(i) Use Lemma 3.1 for $D_{sym(\times)}$ and $D_{sym(\times)}$ and reduce the term $D_{sym(\circ)}$.

(ii) (i) can be written as:

$$(-A^3)^{w(D_{sym(+)}^{\circ})} A^r f_{D_{sym(+)}^{\circ}} - (-A^3)^{w(D_{sym(-)}^{\circ})} A^{-r} f_{D_{sym(-)}^{\circ}} \equiv (A^{2r} - A^{-2r})(-A^3)^{w(D_{sym(0)}^{\circ})} f_{D_{sym(0)}^{\circ}} \pmod{r}$$

and using the equality $w(D_{sym(+)}^{\circ}) = w(D_{sym(-)}^{\circ}) + 2r = w(D_{sym(0)}^{\circ}) + r$, one gets the congruence (ii).

(iii) This follows from (ii) by putting $V_L(t) = f_L(A)$, for $t = A^{-4}$.

□

Lemma 3.1 and Corollary 3.2 have several nice applications: to symmetric knots, periodic 3-manifolds and to analysis of connections between skein modules of the base and covering space in a covering (see Section 7). Below are two elementary but illustrative applications to periodic links.

Theorem 3.3 ([T-1, P-3]) (i) *If L is an r -periodic oriented link (r is a prime), then its Jones polynomial satisfies the relation*

$$V_L(t) \equiv V_L(t^{-1}) \pmod{r, t^r - 1}.$$

(ii) *Let us consider a polynomial $\hat{V}_L(t) = (t^{\frac{1}{2}})^{-3lk(L)}V_L(t)$ which is an invariant of an ambient isotopy of unoriented links. $lk(L)$ denotes here the global linking number of L , any orientation of L gives the same $\hat{V}_L(t)$.*

If L is an r -periodic unoriented link (r is an odd prime), then $\hat{V}_L(t) \equiv \hat{V}_L(t^{-1}) \pmod{r, t^r - 1}$.

Theorem 3.4 *Let K be an r -periodic knot with linking number k with the fixed point set axis. Then*

$$t^{r(k-1)/2}V_K(t) \equiv \frac{(t^{(k+1)/2} - t^{(-k-1)/2}) - (t^{(k-1)/2} - t^{(1-k)/2})}{t - t^{-1}} \pmod{(r, t^r - 1)}.$$

In particular:

(a) *If k is odd then* $V_K(t) \equiv \frac{t^{k/2} + t^{-k/2}}{t^{1/2} + t^{-1/2}} \equiv t^{(k-1)/2} - t^{(k-3)/2} + \dots - t^{(3-k)/2} + t^{(1-k)/2} \pmod{(r, t^r - 1)}$,

(b) *If k is even then* $t^{r/2}V_K(t) \equiv \frac{t^{k/2} + t^{-k/2} - 2(-1)^{k/2}}{t^{1/2} + t^{-1/2}} + (-1)^{k/2} \frac{t^r + 1}{t^{1/2} + t^{-1/2}} \pmod{(r, t^r - 1)}$.

Proof:

3.3(i) From Corollary 3.2(iii) follows that $V_{D_{sym(+)}^\circ} \equiv V_{D_{sym(-)}^\circ} \pmod{(r, t^r - 1)}$. On the other hand we can change D° to its mirror image \bar{D}° by a sequence of changes of type $D_{sym(+)}^\circ \leftrightarrow D_{sym(-)}^\circ$. Therefore $V_{D^\circ}(t) \equiv V_{\bar{D}^\circ}(t) \pmod{(r, t^r - 1)}$. Theorem 3.3(i) follows because $V_{\bar{D}^\circ}(t) = V_{D^\circ}(t^{-1})$.

3.3(ii) The link L may be oriented so that φ (a generator of the Z_r action) preserves the orientation of L (first we orient $L_* = L/Z_r$ the quotient of L under the group action and then we lift the orientation up to L). For such an oriented L we get from (i):

$$V_L(t) \equiv V_L(t^{-1}) \pmod{(r, t^r - 1)}$$

and thus

$$(t^{\frac{1}{2}})^{3lkL} \hat{V}_L(t) \equiv (t^{\frac{1}{2}})^{-3lkL} \hat{V}_L(t^{-1}) \pmod{(r, t^r - 1)}$$

and consequently

$$\hat{V}_L(t) \equiv t^{-3lk(L)} \hat{V}_L(t^{-1}) \pmod{(r, t^r - 1)}.$$

For $r > 2$, $lk(L) \equiv 0 \pmod{r}$ so Theorem 3.3(ii) follows.

3.4 We proceed as in the proof of 3.3(i) except that instead of aiming at mirror image we aim the appropriate torus knot. To see this it is convenient to think of our link L being in a solid torus and simplify its quotient $L_* = L/Z_n$ in the solid torus (see Section 7). The formula for the Jones polynomial of the torus knot, $T_{(r,k)}$, was found by Jones [Jo]:

$$V_{T_{(r,k)}}(t) = t^{r(k-1)/2} \frac{t^{(k+1)/2} - t^{(-k-1)/2} - t^r(t^{(k-1)/2} - t^{(1-k)/2})}{t - t^{-1}}.$$

In Section 7, we discuss elementary proof of the formula and give a short proof of the generalization of the Jones formula to the solid torus (modulo r) using Lemma 3.1.

□

Theorem 3.3 is strong enough to allow Traczyk [T-1] to complete periodicity tables for knots up to 10 crossings (for $r > 3$), that is, to decide whether the knot 10_{101} is 7-periodic.

Example 3.5 *The knot 10_{101} (Fig.3.3) is not r -periodic for $r \geq 5$.*

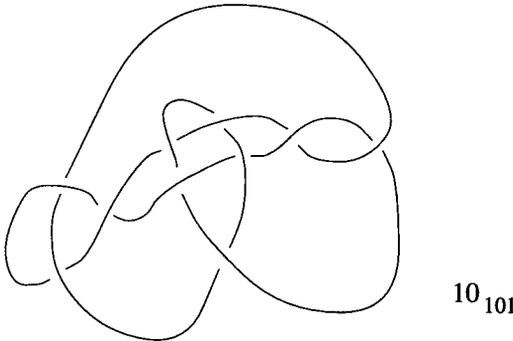


Fig. 3.3

The Jones polynomial for 10_{101} is equal to

$$V_{10_{101}}(t) = t^2 - 3t^3 + 7t^4 - 10t^5 + 14t^6 - 14t^7 + 13t^8 - 11t^9 + 7t^{10} - 4t^{11} + t^{12}.$$

Thus, for $r \geq 5$ it follows that $V_{10_{101}}(t) \not\equiv V_{10_{101}}(t^{-1} \bmod (r, t^r - 1))$. In particular, it follows that $V_{10_{101}}(t) \equiv t^{-1} + 2 + 3t^3 \bmod (5, t^5 - 1)$ and also $V_{10_{101}}(t) \equiv 3t^{-3} + 5t^{-2} + 6t + 4t^2 + 4t^3 \bmod (7, t^7 - 1)$.

Remark 3.6 *Theorem 3.3 and 3.4 do not work for knots and $r = 3$. It is the case because by Corollary 2.4, for any knot, $V_K(t) \equiv 1 \pmod{t^3 - 1}$. One should add that the classical Murasugi criterion using the Alexander polynomial, Theorem 5.3 ([Mu-1]) is working for $r = 3$, and Traczyk developed the method employing the skein polynomial, Theorem 4.10(a) ([T-2, T-3]).*

Yokota proved in [Yo-1] the following criterion for periodic knots which generalize Theorem 3.3 and is independent of Theorem 3.4.

Theorem 3.7 *Let K be an r -periodic knot (r is an odd prime) with linking number k with the fixed point set axis. Then*

- (a) *If k is odd then $V_K(t) \equiv V_K(t^{-1}) \pmod{r, t^{2r} - 1}$,*
- (b) *If k is even then $V_K(t) \equiv t^r V_K(t^{-1}) \pmod{r, \frac{t^{2r}-1}{t+1}}$.*

We can extend Theorems 3.3, 3.4 and 3.7 (or rather show its limits) by considering the following operations on link diagrams:

Definition 3.8 (i) *A t_k move is an elementary operation on an oriented link diagram L resulting in the diagram $t_k(L)$ as shown on Fig. 3.4.*



Fig. 3.4

(ii) *A \bar{t}_k move, k even, is an elementary operation on an oriented link diagram L resulting in the diagram $\bar{t}_k(L)$ as shown on Fig. 3.5.*

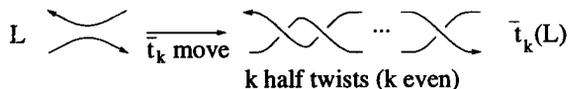


Fig. 3.5

(iii) *The local change in a link diagram which replaces parallel lines by k positive half-twists is called a k -move; see Fig.3.6.*

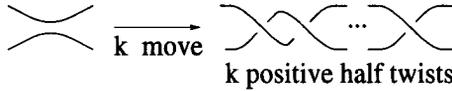


Fig. 3.6

Lemma 3.9 *Let L_k be an unoriented link obtained from L_0 by a k move. Then $\langle L_k \rangle = A^k \langle L_0 \rangle + A^{-3k+2} \frac{A^{4k} - (-1)^k}{A^4 + 1}$.*

Proof: It follows by an induction on k . \square

Corollary 3.10 *If L and L' are oriented t_{2r}, \bar{t}_{2r} equivalent links (that is L and L' differ by a sequence of t_{2r}, \bar{t}_{2r} -moves) then*

$$V_L(t) \equiv t^{rj} V_{L'}(t) \pmod{\left(r, \frac{t^{2r}-1}{r+1}\right)}, \text{ for some integer } j.$$

Proof: For $k = 2r$ we have from Lemma 3.9 $\langle L_{2r} \rangle = A^{2r} \langle L_0 \rangle + A^{-6r+2} (A^{4r} - 1) \frac{A^{4r} + 1}{A^4 + 1} \langle L_\infty \rangle$. Thus $\langle L_{2r} \rangle \equiv A^{2r} \langle L_0 \rangle \pmod{(A^{4r} - 1) \frac{A^{4r} + 1}{A^4 + 1}}$ and Corollary 3.10 easily follows. \square

4 Periodic links and the generalized Jones polynomials.

We show here how periodicity of links is reflected in regularities of skein and Kauffman polynomials. We explore the same ideas which were fundamental in Section 3, especially we use variations of Lemma 3.1.

Let \mathcal{R} be a subring of the ring $Z[v^{\pm 1}, z^{\pm 1}]$ generated by $v^{\pm 1}, z$ and $\frac{v^{-1}-v}{z}$. Let us note that z is not invertible in \mathcal{R} .

Lemma 4.1 *For any link L its skein polynomial $P_L(v, z)$ is in the ring \mathcal{R} .*

Proof: For a trivial link T_n with n components we have $P_{T_n}(v, z) = \left(\frac{v^{-1}-v}{z}\right)^{n-1} \in \mathcal{R}$. Further, if $P_{L_+}(v, z)$ (respectively $P_{L_-}(v, z)$) and $P_{L_0}(v, z)$ are in \mathcal{R} then $P_{L_-}(v, z)$ (respectively $P_{L_+}(v, z)$) is in \mathcal{R} as well. This observation enables a standard induction to conclude 4.1. Now we can formulate our criterion for r -periodic links. It has an especially simple form for a prime period (see Section 5 for a more general statement). \square

Theorem 4.2 *Let L be an r -periodic oriented link and assume that r is a prime number. Then the skein polynomial $P_L(v, z)$ satisfies the relation*

$$P_L(v, z) \equiv P_L(v^{-1}, -z) \pmod{(r, z^r)}$$

where (r, z^r) is an ideal in \mathcal{R} generated by r and z^r .

In order to apply Theorem 4.2 effectively, we need the following fact.

Lemma 4.3 *Suppose that $w(v, z) \in \mathcal{R}$ is written in the form $w(v, z) = \sum_i u_i(v)z^i$, where $u_i(v) \in Z[v^{\mp 1}]$. Then $w(v, z) \in (r, z^r)$ if and only if for any $i \leq r$ the coefficient $u_i(v)$ is in the ideal $(r, (v^{-1} - v)^{r-i})$.*

Proof: \Leftarrow Suppose $u_i(v) \in (r, (v^{-1} - v)^{r-i})$ for $i \leq r$. Now $u_i(v)z^i \equiv (v^{-1} - v)^{r-i}z^i p(v) \equiv (\frac{v^{-1}-v}{z})^{r-i}z^r p(v) \pmod r$, where $p(v) \in Z[v^{\pm 1}]$. Thus $u_i(v)z^i \in (r, z^r)$ and finally $w(v, z) \in (r, z^r)$.

\Rightarrow Suppose that $w(v, z) \in (r, z^r)$, that is, $w(v, z) \equiv z^r \bar{w}(v, z) \pmod r$ for some $\bar{w}(v, z) \in \mathcal{R}$. The element $\bar{w}(v, z)$ can be uniquely written as a sum $\bar{w}(v, z) = z\bar{u}(v, z) + \sum_{j \geq 0} (\frac{v^{-1}-v}{z})^j \bar{u}_j(v)$, where $\bar{u}(v, z) \in Z[v^{\mp 1}, z]$ and $\bar{u}_j(v) \in Z[v^{\mp 1}]$. Thus, for $i \leq r$ ($j = r - i$) we have $u_i(v) \equiv (v^{-1} - v)^{r-i} \bar{u}_{r-i}(v) \pmod r$ and finally $u_i(v) \in (r, (v^{-1} - v)^{r-i})$ for $i \leq r$.

□

Example 4.4 *Let us consider the knot 11_{388} , in Perko's notation [Per], see Fig.4.1. The skein polynomial $P_{11_{388}}(v, z)$ is equal to*

$$(3 - 5v^{-2} + 4v^{-4} - v^{-6}) + (4 - 10v^{-2} + 5v^{-4})z^2 + (1 - 6v^{-2} + v^{-4})z^4 - v^{-2}z^6.$$

Let us consider the polynomial $P_{11_{388}}(v, z) - P_{11_{388}}(v^{-1}, -z)$. The coefficient $u_0(v)$ for this polynomial is equal to $5(-v^{-2} + v^2) + 4(v^{-4} - v^4) - v^{-6} + v^6$ and thus for $r \geq 7$ we have $u_0(v) \notin (r, (v^{-1} - v)^r) = (r, v^{-r} - v^r)$. Now, from Lemma 4.3 we have $P_{11_{388}}(v, z) - P_{11_{388}}(v^{-1}, -z) \notin (r, z^r)$. Therefore from Theorem 4.2 it follows that the knot 11_{388} is not r -periodic for $r \geq 7$.

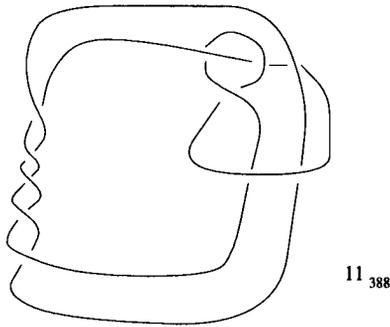


Fig. 4.1

Theorem 3.3 is a corollary of Theorem 4.2 as $V_L(t) = P_L(t, t^{1/2} - t^{-1/2})$.

The periodicity criterion from Theorem 3.3 is weaker than the one from Theorem 4.2: the knot 11_{388} from Example 4.4 has a symmetric Jones polynomial

$$V_{11_{388}}(t) = V_{11_{388}}(t^{-1}) = t^{-2} - t^{-1} + 1 - t + t^2,$$

and therefore Theorem 3.3 can not be applied in this case. Theorem 3.4 is also not sufficient in this case (linking number 5 cannot be excluded as $V_{11_{388}}(t) = \frac{t^{5/2} + t^{-5/2}}{t^{1/2} + t^{-1/2}} \equiv V_{T(r,5)} \pmod{(t^r - 1)}$).

Now let us consider the Kauffman polynomial $F_L(a, z)$. Let \mathcal{R}' be a subring of $Z[a^{\mp 1}, z^{\mp 1}]$ generated by $a^{\mp 1}$, z and $\frac{a+a^{-1}}{z}$. It is easy to check that Kauffman polynomials of links are in \mathcal{R}' (compare Lemma 4.3).

Theorem 4.5 *Let L be an r -periodic oriented link and let r be a prime number. Then the Kauffman polynomial F_L satisfies the following relation*

$$F_L(a, z) \equiv F_L(a^{-1}, z) \pmod{(r, z^r)},$$

where (r, z^r) is the ideal in \mathcal{R}' generated by r and z^r .

In order to apply Theorem 4.5 we will use the appropriate version of Lemma 4.3.

Lemma 4.6 Suppose that $w(a, z) \in \mathcal{R}'$ is written in the form $w(a, z) = \sum_i v_i(a)z^i$, where $v_i(a) \in Z[a^{\mp 1}]$. Then $w(a, z) \in (r, z^r)$ if and only if for any $i \leq r$ the coefficient $v_i(a)$ is in the ideal $(r, (a + a^{-1})^{r-i})$.

Example 4.7 Let us consider the knot 10_{48} from Rolfsen's book [Ro], see Fig.4.2. This knot has a symmetric skein polynomial, that is $P_{10_{48}}(v, z) = P_{10_{48}}(v^{-1}, -z)$. Consequently, Theorem 4.2 can not be applied to examine the periodicity of this knot. So let us apply the Kauffman polynomial to show that the knot 10_{48} is not r -periodic for $r \geq 7$. It can be calculated ([D-T, P-1]) that $F_{10_{48}}(a, z) - F_{10_{48}}(a^{-1}, z) = z(a^5 + 3a^3 + 2a - 2a^{-1} - 3a^{-3} - a^{-5}) + z^2(\dots)$.

Now let us apply Lemma 4.6 for $i = 1$ and let us note that for $r \geq 7$ we have $a^5 + 3a^3 + 2a - 2a^{-1} - 3a^{-3} - a^{-5} \notin (r, (a + a^{-1})^{r-1})$ (Note that for $r = 5$ we have $a^5 + 3a^3 + 2a - 2a^{-1} - 3a^{-3} - a^{-5} = a(a + a^{-1})^4 - a^{-1}(a + a^{-1})^4 = (a - a^{-1})(a + a^{-1})^4 \in (5, (a + a^{-1})^4)$).

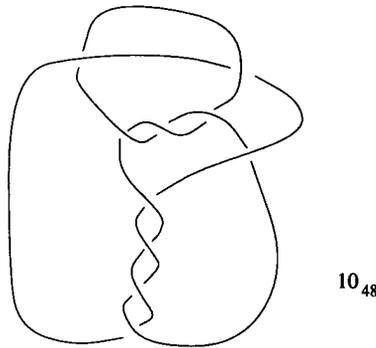


Fig. 4.2

We can reformulate Theorem 4.5 in terms of the Dubrovnik version of the Kauffman polynomial:

Theorem 4.8 Let L be an r -periodic oriented link and let r be a prime number. Then the Dubrovnik polynomial $F_L^* \in \mathcal{R}'' = Z[a^{\pm 1}, \frac{a-a^{-1}}{z}, z] \subset Z[a^{\pm 1}, z^{\pm 1}]$ satisfies the following relation

$$F_L^*(a, z) \equiv F_L^*(a^{-1}, -z) \text{ mod } (r, z^r),$$

where (r, z^r) is the ideal in \mathcal{R}'' generated by r and z^r .

In the last part of this section we will show how to strengthen Theorems 4.2 and 4.5.

One can also modify our method so it applies to symmetric links with one fixed point component. We consider more general setting in Section 7 (symmetric links in the solid torus).

Proof of Theorems 4.3 and 4.5 is very similar to that of Theorem 3.3. Instead of Lemma 3.1 we use the following main lemma (which is of independent interest), proof of which is again following the same principle (Z_r acting on the leaves of a computational tree) as the proof of Lemma 3.1.

Lemma 4.9 (i) Let $L_{sym(\nearrow)}$, $L_{sym(\searrow)}$ and $L_{sym(\sphericalangle)}$ denote three φ -invariant diagrams of links which are the same outside of the Z_r -orbit of a fixed single crossing and which at the crossing differ by replacing \nearrow by \searrow or \sphericalangle , respectively. φ , as before denotes the generation of the Z_r action. Then

$$a^r P_L_{sym(\nearrow)}(a, z) + a^{-r} P_L_{sym(\searrow)}(a, z) = z^r P_L_{sym(\sphericalangle)}(a, z) \pmod r.$$

(ii) Consider four r -periodic unoriented diagrams $L_{sym(\times)}$, $L_{sym(\otimes)}$, $L_{sym(\smile)}$ and $L_{sym(\frown)}$. Then the Kauffman polynomial of unoriented diagrams $\Lambda_L(a, z)$, satisfies:

$$\Lambda_{sym(\times)} + \Lambda_{sym(\otimes)} \equiv z^r (\Lambda_{sym(\smile)} + \Lambda_{sym(\frown)}) \pmod r.$$

Traczyk [T-3] and Yokota [Yo-2] substantially generalized Theorems 4.2 and 4.5. Lemma 4.9 is still crucial in their proofs, but detailed study of the skein polynomial of the torus knots is also needed. I simplified their proof by using Jaeger composition product [P-5].

Theorem 4.10 Let K be an r -periodic knot (r an odd prime number) with linking number with the rotation axis equal to k and $P_K(v, z) = \sum_{i=0} P_{2i} z^{2i}$ then:

- (a) (Traczyk) If $P_0(K) = \sum a_{2i}v^{2i}$ then $a_{2i} \equiv a_{2i+2} \pmod r$ except possibly when $2i + 1 \equiv \pm k \pmod r$.
- (b) (Yokota) $P_{2i}(K) \equiv b_{2i}P_0(K) \pmod r$ for $2i < r - 1$, where numbers b_{2i} depends only on r and $k \pmod r$.

One can also use the Jaeger's skein state model for Kauffman polynomial to give periodicity criteria yielded by the Kauffman polynomial. I have never written details of the above idea (from June 1992), and Yokota proved independently criteria yielded by the Kauffman polynomial [Yo-3, Yo-4].

Theorem 4.11 ((Yokota)) *Let K be an r -periodic knot (r an odd prime number) with the linking number with the rotation axis equal to k , and let $(\frac{a-a^{-1}}{z} + 1)F^*(a, z) = \sum_{i=0} F_i z^i$ ($F_0(a, z) = P_0(v, z)$ for $v = a^{-1}$), then for $2i \leq r - 3$:*

(i) $F_{2i}(K) \equiv b_{2i}F_0(K) \pmod r$,
 $F_{2i+1} \equiv 0 \pmod r$ except $i = 0$ where $F_1 \in Z[a^{\pm r}] \pmod r$

(ii) Let $P^*(a, z) = P(v, z)$ for $a = v^{-1}$ and $J_K(a, z) = (\frac{a-a^{-1}}{z} + 1)F_K^*(a, z) - (\frac{a-a^{-1}}{z})P_K^*(a, z)$ and let $J_K(a, z) = \sum_{i=0} J_i z^i$. Then for $0 \leq i \leq r - 1$ $J_i(a, k) \equiv 0 \pmod r$ except for $i = 1$ when $J_1(a; K) \in Z[a^{\pm r}] \pmod r$

Define $J_{i,l}(a; K)$ as a polynomial obtained by gathering all terms in $J_i(a; K)$ which have degree $\pm l \pmod r$. Then:

For each l and for $0 \leq 2i \leq r - 3$

$$J_{r+2i,l}(a; K) \equiv b_{i,k}J_{r,k}(a, K) \pmod r$$

$$J_{r+2i+1}(a; K) \equiv 0 \pmod r \text{ except } J_{r+1}(a; K) \text{ which modulo } r \text{ is in } Z_r[a^{\pm r}].$$

5 r^q -periodic links and Vassiliev invariants.

The criteria of r -periodicity, which we have discussed before, can be partially extended to the case of r^q -periodic links. We assume that r is a prime number and the fixed point set of the action of Z_{r^q} is a circle disjoint from the link in question (trivial knot by the Smith Conjecture)

We will not repeat here all criteria where r is generalized to r^q [P-3], but instead we will list one pretty general criterion using Vassiliev-Gusarov invariants (compare [P-4]).

Definition 5.1 Let \mathcal{K}^{sg} denote the set of singular oriented knots in S^3 where we allow only immersion of S^1 with, possibly, double points, up to ambient isotopy. Let $Z\mathcal{K}^{sg}$ denote the free abelian group generated by elements of \mathcal{K}^{sg} (i.e. formal linear combinations of singular links). In the group $Z\mathcal{K}^{sg}$ we consider resolving singularity relations $\sim: K_{cr} = K_+ - K_-$; see Fig. 5.1.

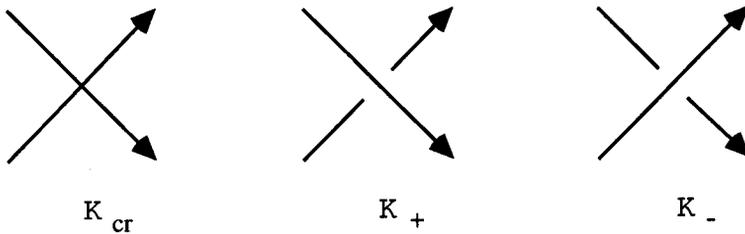


Fig. 5.1

$Z\mathcal{K}^{sg} / \sim$ is clearly \mathbb{Z} -isomorphic to $Z\mathcal{K}$. Let C_m be a subgroup of $Z\mathcal{K}^{sg} / \sim = Z\mathcal{K}$ generated by immersed knots with m double points. Let A be any abelian group. The m 'th Vassiliev-Gusarov invariant is a homomorphism $f: Z\mathcal{K} \rightarrow A$ such that $f(C_{m+1}) = 0$.

Theorem 5.2 (i) Let K be an oriented r^q -periodic knot and f a Vassiliev-Gusarov invariant of degree $m < r^q$. Then $f(K) \equiv f(\bar{K}) \pmod r$, where \bar{K} is the mirror image of K .

(ii) Let K be an oriented r^q -periodic knot with the linking number k with the fixed point set axis, and let f be a Vassiliev-Gusarov invariant of degree $m < r^q$. Then $f(K) \equiv f(T_{(r^q, k)})$ where $T_{(r^q, k)}$ is the torus knot of type (r^q, k) .

Proof of Theorem 5.2 is similar to the previous one and again bases on the fundamental observation that $f(K_{sym(+)} - f(K_{sym(-)}) \equiv 0 \pmod r$.

Our method allows also to prove quickly the classical Murasugi congruence for r^q periodic knot, using the Alexander polynomial.

Theorem 5.3 *Show that by applying our method to Alexander polynomial we obtain the following version of a theorem of Murasugi:*

$$\Delta_L(t) \equiv \Delta_{L_*}^q(t)(1+t+t^2+\dots+t^{\lambda-1})^{r^q-1} \pmod r$$

where L_* is the quotient of an r^q -periodic link L and λ is the linking number of L and z axis.

Proof: Sketch.

Construct the binary computational tree of the Alexander polynomial of L_* and the associated binary tree for Alexander polynomial L modulo r . For the Alexander polynomial of L_* we use the skein relation $\Delta_{L_{*+}}(t) - \Delta_{L_{*-}}(t) = (t^{1/2} - t^{-1/2})\Delta_{L_{*0}}(t)$ and for the r^q periodic link L the congruence (related to Lemma 4.8(i)):

$$\Delta_{L_{sym(+)}}(t) - \Delta_{L_{sym(-)}}(t) \equiv (t^{r^q/2} - t^{-r^q/2})\Delta_{L_{sym(0)}}(t) \pmod r.$$

Finally we use the fact that Alexander polynomial of a split link is zero, and that $\Delta_{T(r^q, \lambda)} \equiv (1+t+t^2+\dots+t^{\lambda-1})^{r^q-1} \pmod r$. \square

Remark 5.4 *If $r^q = 2$ then the formula from the previous exercise reduces to:*

$$\Delta_L(t) \equiv \Delta_{L_*}^2(t)(1+t+t^2+\dots+t^{\lambda-1}) \pmod 2.$$

Similar formula can be proven, for other Z_2 -symmetry of links. Namely: a knot (or an oriented link) in R^3 is called strongly plus amphicheiral if it has a realization in R^3 which is preserved by a (changing orientation) central symmetry $((x, y, z) \rightarrow (-x, -y, -z))$; "plus" means that the involution is preserving orientation of the link. One can show, using "skein" considerations, as in the case of Theorem 5.3, that if L is a strongly + amphicheiral link, then modulo 2 the polynomial $\Delta_L(t)$ is a square of another polynomial.

Hartley and Kawachi [H-K] proved that $\Delta_L(t)$ is a square in general. For example if $L(2m+1)$ is a Turks head link - the closure of the 3-string braid $(\sigma_1\sigma_2^{-1})^{2m+1}$, then its Alexander polynomial satisfies:

$$\Delta_{L(2m+1)} = \left(\frac{a^{m+1} - a^{-m-1}}{a - a^{-1}} \right)^2$$

where $a + a^{-1} = 1 - t - t^{-1}$ and the Alexander polynomial is described up to an invertible element in $Z[t^{\pm 1}]$. This formula follows immediately by considering the Burau representation of the 3-string braid group [Bi, Bu].

6 Lissajous knots and billiard knots.

A Lissajous knot K is a knot in R^3 given by the parametric equations

$$x = \cos(\eta_x t + \phi_x)$$

$$y = \cos(\eta_y t + \phi_y)$$

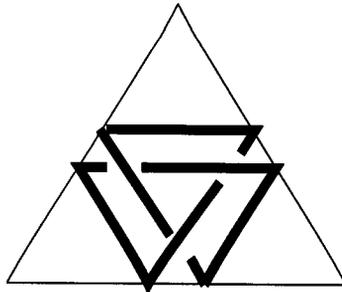
$$z = \cos(\eta_z t + \phi_z)$$

for integers η_x, η_y, η_z . A *Lissajous link* is a collection of disjoint Lissajous knots.

The fundamental question was asked in [BHJS]: which knots are Lissajous?

It was shown in [BHJS] and [J-P] that a Lissajous knot is a Z_2 -symmetric knot (2-periodic with a linking number with the axis equal to ± 1 or strongly plus amphicheiral) so a "random" knot is not Lissajous (for example a nontrivial torus knot is not Lissajous). Lamm constructed infinite family of different Lissajous knots [La].

One defines a *billiard knot* (or racquetball knot) as the trajectory inside a cube of a ball which leaves a wall at rational angles with respect to the natural frame, and travels in a straight line except for reflecting perfectly off the walls; generically it will miss the corners and edges, and will form a knot. We show in [J-P] that these knots are precisely the same as the Lissajous knots. We define general billiard knots, e.g. taking another polyhedron instead of the ball, considering a non-Euclidean metric, or considering the trajectory of a ball in the configuration space of a flat billiard. We will illustrate these by various examples. For instance, the trefoil knot is not a Lissajous knot but we can easily realize it as a billiard knot in a room with a regular triangular floor.



The left handed trefoil knot in a room
with a regular triangular floor ("Odin's triangle")

Fig. 6.1

Theorem 6.1 *Lissajous knots and billiard knots in a cube are the same up to ambient isotopy.*

A billiard knot (or link), is a simple closed trajectory (trajectories) of a ball in a 3-dimensional billiard table. The simplest billiards to consider would be polytope (finite convex polyhedra in R^3). But even for Platonian bodies we know nothing of the knots they support except in the case of the cube. It seems that polytopes which are the products of polygons and the interval $([-1, 1])$ (i.e. polygonal prisms) are more accessible. This is the case because diagrams of knots are billiard trajectories in 2-dimensional tables. We will list some examples below (compare [Ta]).

Example 6.2 (i) *The trivial knot and the trefoil knot are the trajectories of a ball in a room (prism) with an acute triangular floor. In Fig.6.2(a), the diagram of the trivial knot is an inscribed triangle Δ_I whose vertices are the feet of the triangle's altitudes. If we move the first vertex of Δ_I slightly, each of its edges splits into two and we get the diagram of the trefoil. We should be careful with the altitude of the trajectory: We start from level 1 at the vertex close to the vertex of Δ_I and opposite to the shortest edge of Δ_I . Then we choose the vertical parameter so that the trajectory has 3 maxima and three minima (Fig.6.2(b)).*

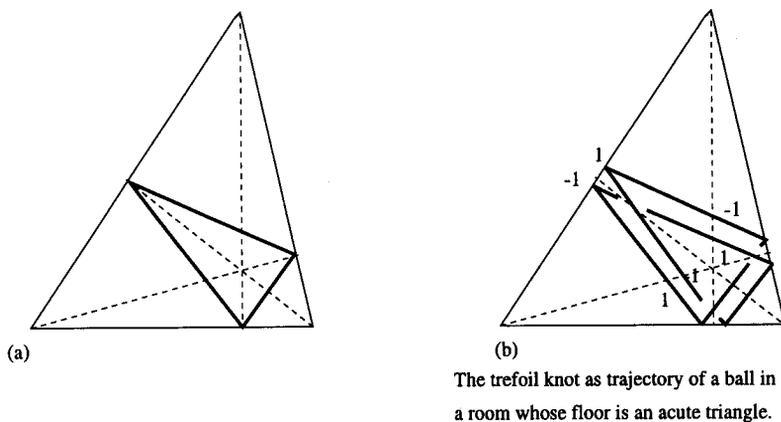
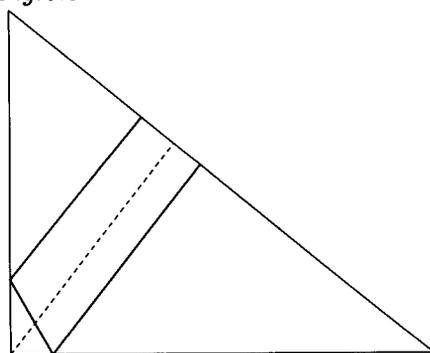


Fig. 6.2

- (ii) The trivial knot is a trajectory of a ball in a room with an right triangular floor, Fig.6.3.



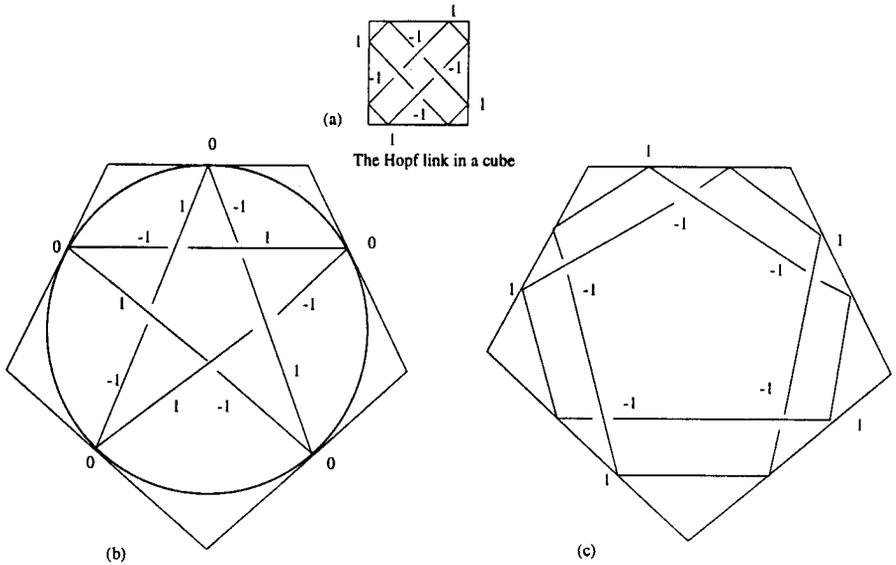
The trivial knot can be realized as a trajectory of a ball in any room with a right triangular floor.

Fig. 6.3

- (iii) If the floor of a room is a general obtuse triangle, it is an open problem whether any knot can be realized as the trajectory of a ball in it. However we have general theorem that periodic points are dense (in the phase space of the billiard flow) in a rational polygon (that is, all polygonal angles are rational with respect to π) [BGKT].

Example 6.2(i) is of interest because it was shown in [BHJS] that the trefoil knot is not a Lissajous knot and thus it is not a trajectory of a ball in a room with a rectangular floor. More generally we show in Section 3 that no nontrivial torus knot is a Lissajous knot. However, we can construct infinitely many torus knots in prisms and in the cylinder.

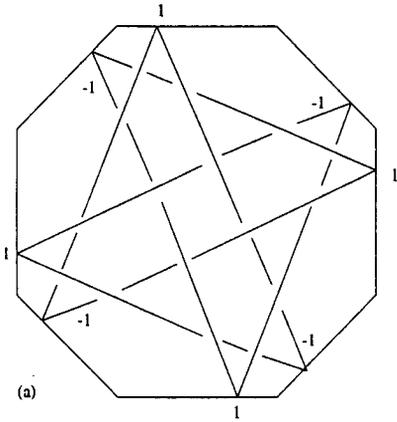
Example 6.3 (i) Any torus knot (or link) of type $(n, 2)$ can be realized as a trajectory of a ball in a room whose floor is a regular n -gon ($n \geq 3$). Fig.6.1 shows the $(3, 2)$ torus knot (trefoil) in the regular triangular prism; Fig.6.4(a) depicts the $(4, 2)$ torus link in the cube; and Fig.6.4(b)(c) illustrates the $(5, 2)$ torus knot in a room with a regular pentagonal floor.



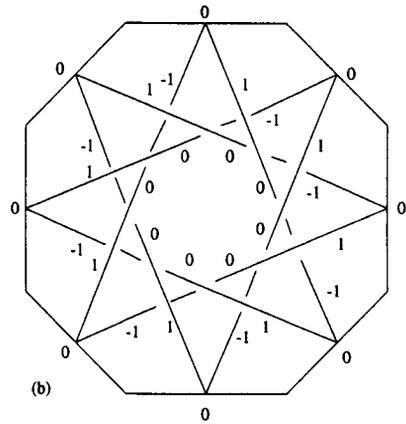
The torus knot of type $(5,2)$ as a trajectory of a ball in a room with a regular pentagonal floor.

Fig. 6.4

(ii) The $(4,3)$ torus knot is a trajectory of a ball in a room with the regular octagonal floor; Fig.6.5(a).



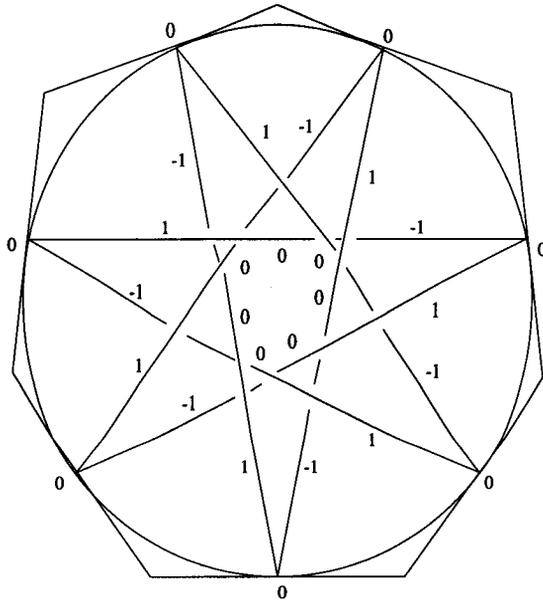
(a) The torus knot of type $(4,3)$ as a trajectory of a ball in a room with a regular octagonal floor.



(b) The torus knot of type $(8,3)$ as a trajectory of a ball in a room with a regular octagonal floor.

Fig. 6.5

- (iii) Figures 6.5(b) and 6.6 illustrate how to construct a torus knot (or link) of type $(n, 3)$ in a room with a regular n -gonal floor for $n \geq 7$.
- (iv) Any torus knot (or link) of type (n, k) , where $n \geq 2k + 1$, can be realized as a trajectory of a ball in a room with a regular n -gonal floor. The pattern generalizes that of Figures 6.4(b), 6.5(b) and 6.6. Edges of the diagram go from the center of the i^{th} edge to the center of the $(i + k)^{\text{th}}$ edge of the n -gon. The ball bounces from walls at altitude 0 and its trajectory has n maxima and n minima. The whole knot (or link) is n -periodic.



The torus knot of the type $(7,3)$ realized as a trajectory of a ball in a room with a regular heptagonal floor.

Fig. 6.6

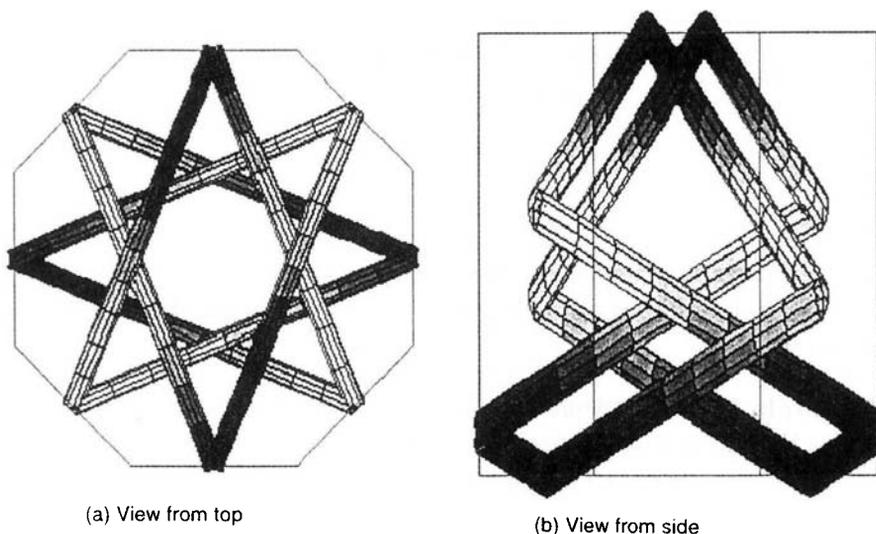
Example 6.4 Let D be a closed billiard trajectory on a 2-dimensional polygonal table. If D is composed of an odd number of segments, then we can always find the “double cover” closed trajectory $D^{(2)}$ in the neighborhood of D (each segment will be replaced by two parallel segments on the opposite sides of the initial segment). This idea can be used to construct, for a given billiard knot K in a polygonal prism (the projection D of K having an odd number of segments), a 2-cable $K^{(2)}$ of K as a billiard trajectory (with projection $D^{(2)}$). This idea is illustrated in Fig. 6.1 and 6.4(c) (the $(5,2)$ torus knot as a 2-cable of a trivial one). Starting from Example 2.3(iv) we can construct a 2-cable of a torus knot of the type (n, k) in a regular n -gonal prism, for n odd and $n \geq 2k + 1$.

It follows from [BHJS] that 3-braid alternating knots of the form $(\sigma_1 \sigma_2^{-1})^{2k}$ are not Lissajous knots as they have a non-zero Arf invariant

(Corollary 6.12). For $k = 1$ we have the figure eight knot and for $k = 2$ the 8_{18} knot [Ro].

Example 6.5 (i) *The Listing knot (figure eight knot) can be realized as a trajectory of a ball in a room with a regular octagonal floor, Fig. 6.7.*

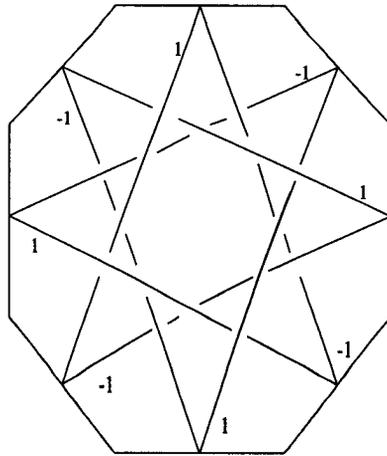
(ii) *Fig.6.8 describes the knot 8_{18} as a trajectory of a ball in a room with a regular octagonal floor. This pattern can be extended to obtain the knot (or link) which is the closure of the three braid $(\sigma_1\sigma_2^{-1})^{2k}$ in a regular $4k$ -gonal prism ($k > 1$).*



Graphics generated by M. Veve

The Listing (figure eight) knot as a trajectory of a ball in a room with a regular octagonal floor.

Fig. 6.7



The knot 8_{18} [Ro], realized as a trajectory of a ball in a room with a regular octagonal floor.

Fig. 6.8

In the example below we show that the cylinder $D^2 \times [-1, 1]$ support an infinite number of different knot types (in the case of a cube it was shown in [La]).

Example 6.6 (i) Any torus knot (or link) of type (n, k) , where $n \geq 2k + 1$, can be realized as a trajectory of a ball in the cylinder; compare Fig. 6.4(b), Fig. 6.5(b) and Fig. 6.6.

(ii) Every knot (or link) which is the closure of the three braid $(\sigma_1 \sigma_2^{-1})^{2k}$ can be realized as the trajectory of a ball in the cylinder. See Fig. 6.7(b) for the case of $k = 1$ (Listing knot) and Fig. 6.8 for the case of $k = 2$ and the general pattern.

Any type of knot can be obtained as a trajectory of a ball in some polyhedral billiard (possibly very complicated). To see this, consider a polygonal knot in R^3 and place “mirrors” (walls) at any vertex, in such a way that the polygon is a “light ray” (ball) trajectory.

Conjecture 6.7

Any knot type can be realized as the trajectory of a ball in a polytope.

Conjecture 6.8

Any polytope supports an infinite number of different knot types.

Problem 6.9

1. *Is there a convex polyhedral billiard in which any knot type can be realized as the trajectory of a ball?*
2. *Can any knot type be realized as the trajectory of a ball in a room with a regular polygonal floor?*
3. *Which knot types can be realized as trajectories of a ball in a cylinder $(D^2 \times [-1, 1])$?*

The partial answer to 6.9(3.) was given in [J-P] and [L-O]. In particular Lamm and Obermeyer have shown that not all knots are knots in a cylinder (e.g. 5_2 and 8_{10} are not cylinder knots). The new interesting feature of [L-O] is the use of ribbon condition.

Below we list some information on Lissajous knots (or equivalently billiard knots in a cube).

Theorem 6.10 ([BHJS]) *An even Lissajous knot is 2-periodic and an odd Lissajous knot is strongly + amphicheiral. A Lissajous knot is called odd if all, η_x, η_y and η_z are odd. Otherwise it is called an even Lissajous knot.*

Theorem 6.11 ([J-P]) *In the even case the linking number of the axis of the Z_2 -action with the knot is equal to ± 1 .*

Corollary 6.12 (i) *([BHJS].) The Arf invariant of the Lissajous knot is 0.*

(ii) *([J-P]) A nontrivial torus knot is not a Lissajous knot.*

(iii) *([J-P]) For $\eta_z = 2$ a Lissajous knot is a two bridge knot and its Alexander polynomial is congruent to 1 modulo 2.*

Theorem 6.13 ([La]) *Let $\eta_x, \eta_y > 1$ be relatively prime integers and K the Lissajous knot with $\eta_z = 2\eta_x\eta_y - \eta_x - \eta_y$ and $\phi_x = \frac{2\eta_x-1}{2\eta_z}\pi$, $\phi_y = \frac{\pi}{2\eta_z}$, $\phi_z = 0$.*

Then the Lissajous diagram of the projection on the $x - y$ plane is alternating. Above knots form an infinite family of different Lissajous knots.

Motivated by the case of 2-periodic knots we propose

Conjecture 6.14

Turks head knots, (e.g. the closure of the 3-string braids $(\sigma_1\sigma_2^{-1})^{2k+1}$), are not Lissajous. Observe that they are strongly + amphicheiral.

We do not think, as the above conjecture shows, that the converse to Theorem 6.11 holds. However for 2-periodic knots it may hold (the method sketched in Section 0.4 of [BHJS] may work).

Problem 6.15

Let K be a Z_2 -periodic knot, such that the linking number of the axis of the Z_2 -action with K is equal to ± 1 . Is K an even Lissajous knot?

The first prime knots (in the knot tables [Ro]) which may or may not be Lissajous are $8_3, 8_6$ (7_5 is constructed in [La]).

7 Applications and Speculations

A lot can be said about the structure of a manifold by studying its symmetries. The existence of Z_r action on a homology sphere is reflected in the Reshetikhin-Turaev-Witten invariants. In our description we follow [K-P]. Let M be a closed connected oriented 3-manifold represented as a surgery on a framed link $L \subset S^3$. Let $r \geq 3$, and set the variable A used in the Kauffman skein relation to be a primitive root of unity of order $2r$. In particular, $A^{2r} = 1$. Recall that the invariant $\mathcal{I}_r(M)$ is given by:

$$\mathcal{I}_r(M) = \kappa^{-3\sigma_L} \eta^{com(L)} [L(\Omega_r)] \quad (2)$$

In $L(\Omega)$ each component of L is decorated by an element Ω from the Kauffman bracket skein module of a solid torus (see Proposition 7.3):

$$\Omega_r = \sum_{i=0}^{[(r-3)/2]} [e_i]e_i$$

Elements e_i satisfy the recursive relation:

$$e_{i+1} = ze_i - e_{i-1}$$

where z can be represented by a longitude of the torus, and $e_0 = 1$, $e_1 = z$. The value of the Kauffman bracket skein module of the skein element e_i , when the solid torus is embedded in S^3 in a standard way, is given by $[e_i] = (-1)^i \frac{A^{2i+2} - A^{-2i-2}}{A^2 - A^{-2}}$. $[e_i]$ is the version of the Kauffman bracket, normalized in such a way that $[\emptyset] = 1$ and $[L] = (-A^2 - A^{-2})\langle L \rangle$. In the equation (2), η is a number which satisfies $\eta^2[\Omega_r] = 1$, and κ^3 is a root of unity such that

$$\kappa^6 = A^{-6-r(r+1)/2}.$$

Finally, σ_L denotes the signature of the linking matrix of L .

Theorem 7.1 (K-P)

Suppose that M is a homology sphere and r is an odd prime. If M is r -periodic then

$$\mathcal{I}_r(M)(A) = \kappa^{6j} \cdot \mathcal{I}_r(M)(A^{-1}) \pmod{r}$$

for some integer j .

Theorem 7.1 holds also for Z_r -homology spheres.

Skein modules can be thought as generalizations to 3-manifolds of polynomial invariants of links in S^3 . Our periodicity criteria (especially when link and its quotient under a group action are compared), can be thought as the first step toward understanding relations between skein modules of the base and covering space in a covering. We will consider here the Kauffman bracket skein module and the relation between a base and covering space in the case of the solid torus.

Definition 7.2 ([P-2, H-P-2])

Let M be an oriented 3-manifold, \mathcal{L}_{fr} the set of unoriented framed links in M (\emptyset allowed), $R = \mathbb{Z}[A^{\pm 1}]$, and $R\mathcal{L}_{fr}$ the free R module with basis \mathcal{L}_{fr} . Let $S_{2,\infty}$ be the submodule of $R\mathcal{L}_{fr}$ generated by skein expressions $L_+ - AL_0 - A^{-1}L_\infty$, where the triple L_+, L_0, L_∞ is presented in Fig. 7.1, and $L \sqcup T_1 + (A^2 + A^{-2})L$, where T_1 denotes the trivial framed knot. We define the Kauffman bracket skein module (KBSM), $S_{2,\infty}(M)$, as the quotient $\mathcal{S}_{2,\infty}(M) = R\mathcal{L}_{fr}/S_{2,\infty}$.

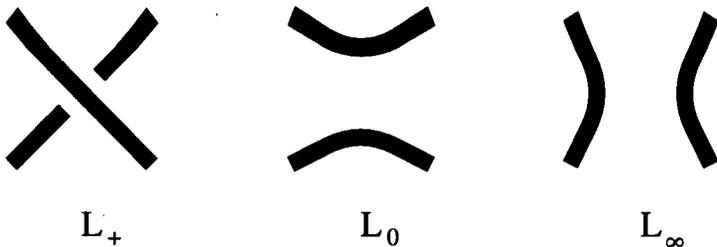


Fig. 7.1.

Notice that $L^{(1)} = -A^3L$ in $S_{2,\infty}(M)$; we call this the framing relation. In fact this relation can be used instead of $L \sqcup T_1 + (A^2 + A^{-2})L$ relation.

Proposition 7.3 ([H-P-1, P-2])

The KBSM of a solid torus (presented as an annulus times an interval), is an algebra generated by a longitude of the solid torus; it is $\mathbb{Z}[A^{\pm 1}]$ algebra isomorphic to $\mathbb{Z}[A^{\pm 1}][z]$ where z corresponds to the longitude.

Let the group Z_r acts on the solid torus $(S^1 \times [1/2, 1]) \times [0, 1]$ with the generator $\varphi(z, t) = (e^{2\pi i/r} z, t)$. where z represents an annulus point and t an interval point. Let p be the r -covering map determined by the action (see Fig. 7.2). We have the “transfer” map p_*^{-1} from the KBSM of the base to the KBSM of the covering space (modulo r) due to the generalization of Lemma 3.2. where z represent an annulus point and t an interval point. To present the generalization in the most natural setting we introduce the notion of the “moduli” equivalence of links and associated moduli skein modules.

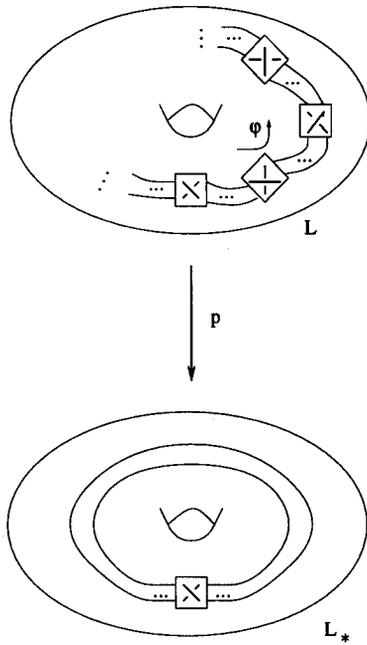


Fig. 7.2

Definition 7.4 (i) We say that two links in a manifold M are moduli equivalent if there is a preserving orientation homeomorphism of M sending one link to another.

(ii) Let G be a finite group action on M . We say that two links L and L' in M are G moduli equivalent if L' is ambient isotopic to $g(L)$ for some $g \in G$.

(ii) We define moduli KBSM (resp. G moduli KBSM) as KBSM divided by moduli (resp. G moduli) relation. We would use the notation $S_{2,\infty}^m(M)$ (resp. $S_{2,\infty}^G(M)$).

Lemma 7.5 Let the group Z_r , r prime, act on the oriented 3-manifold M and let $L = L_{sym(cr)}$ be a framed singular link in M which satisfies

$\varphi(L) = L$ where $\varphi : M \rightarrow M$ is the generator of the Z_r -action, and Z_r has no fixed points on L . Then in the skein module $\mathcal{S}_{2,\infty}^{Z_r}(M)$ one has the formula

$$L_{\text{sym}(\times)} \equiv A^r L_{\text{sym}(\smile)} + A^{-r} L_{\text{sym}(\frown)} \pmod r$$

where $L_{\text{sym}(\times)}$, $L_{\text{sym}(\smile)}$ and $L_{\text{sym}(\frown)}$ denote three φ -invariant diagrams of links which are the same outside of the Z_r -orbit of a neighborhood of a fixed singular crossing at which they differ by replacing $L_{\text{sym}(cr)}$ by \times or \smile , or \frown , respectively.

Notice that in the case of an action on the solid torus, φ is isotopic to identity, thus Z_r moduli KBSM is the same as KBSM.

We can use Lemma 7.5 to find the formula for a torus knot in a solid torus modulo r , for a prime r .

Theorem 7.6 *The torus knot $T_{r,k}$ satisfies:*

$$T(r, k) = A^{r(k-1)} \left(\frac{x^{k+1} - x^{-k-1} - A^{-4r}(x^{k-1} - x^{1-k})}{x - x^{-1}} \right)$$

where $z = x + x^{-1}$ is a longitude of the solid torus (an annulus times an interval).

One can prove Theorem 7.6 by rather involved induction (compare example 7.7), however modulo r Theorem 7.6 has very easy proof using Lemma 7.5. Namely $T_{r,k}$ is an r cover of a knot $T_{1,k}$ and for $T_{1,k}$ one has $T_{1,k+2} = A(x + x^{-1})T_{1,k+1} - A^2T_{1,k}$ in KBSM, thus $T_{1,k} = A^{(k-1)} \left(\frac{x^{k+1} - x^{-k-1} - A^{-4}(x^{k-1} - x^{1-k})}{x - x^{-1}} \right)$ Now one compares binary computational trees of $T_{1,k}$ and $T_{r,k}$ (modulo r) using lemma 7.5, to get mod r version of Theorem 7.6.

In fact for any regular r covering $p : M \rightarrow M_*$ we have the transfer map (mod r):

$$p_*^{-1} : \mathcal{S}_{2,\infty}(M_*) \rightarrow \mathcal{S}_{2,\infty}^{Z_r}(M) \pmod r$$

which is a Z homomorphism and $p_*^{-1}(w(A)L) \equiv w(A^r)p^{-1}(L) \pmod r$.

Example 7.7 We compute here the torus link $T_{2,k}$ in the KBSM of a solid torus. We will prove that for an odd k (that is when $T_{2,k} = K_{2,k}$ is a knot):

$$L(2, k) = A^{2(k-1)} \frac{x^{k+1} - x^{-k-1} - A^{-8}(x^{k-1} - x^{1-k})}{x - x^{-1}}$$

and for an even k :

$$L(2, k) = A^{2(k-1)} \frac{x^{k+1} - x^{-k-1} - A^{-8}(x^{k-1} - x^{1-k})}{x - x^{-1}} + 2A^{-6}$$

The formulas follow inductively from the following, easy to check, recurrence relation ($k > 2$), compare Fig 7.3.:

$$L(2, k) = A^2 L_{-, -}^{a, b} - A^{-2} L_{|, |}^{a, b} + A^{-1} L_{+, |}^{a, b} + A^{-1} L_{|, +}^{a, b} = A^2 z L_{2, k-1} - A^4 L_{2, k-2} - 2A^{-4} u_k$$

where $u_k = z$ for k odd and $u_k = -A^2 - A^{-2}$ for k even. If we denote $e_n = \frac{x^{n+1} - x^{-n-1}}{x - x^{-1}}$, Chebyshev polynomial in the variable $z = x + x^{-1}$, then in odd cases the formula can be written as:

$$L(2, k) = A^{2(k-1)}(e_k - A^{-8}e_{k-2})$$

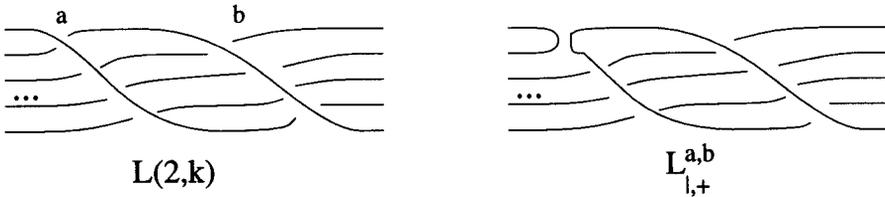


Fig. 7.3

Corollary 7.8 (Traczyk)

- (a) $A^3 e_3$ is represented in the solid torus by a 2-component link being a closure of the 3-braid $\Delta = \sigma_1 \sigma_2 \sigma_1$; Fig. 7.4.

- (b) In S^3 , the closed braid links $\hat{\Delta}^{2n+1}$ form an infinite family of (simple) links whose Jones polynomial differ only by an invertible element in $Z[t^{\pm\frac{1}{2}}]$.
- (c) There are infinite families of links sharing the same Jones polynomial (e.g. $t^2 + 2 + t^{-2}$).

$$L(2,3) = A^4 (e_3 - A^{-8} e_1) = \text{Diagram 1} = A \text{Diagram 2} \\ -A^{-1} \text{Diagram 3} = A \Delta - A^{-4} e_1$$

Fig. 7.4

Proof:

- (a) Using the formula of Example we have $L_{2,3} = A^4(e_3 - A^{-8}e_1)$. The necessary calculation is shown in Fig. 7.4.
- (b) e_3 is an eigenvector of the Dehn twist on the solid torus and any braid, γ , is changed by a Dehn twist to $\Delta^2\gamma$. Thus $\hat{\Delta}^{2n+1} = A^{6n}\hat{\Delta}$ in the KBSM of the solid torus.
Closed braids $\hat{\Delta}^{2n+1}$ form an infinite family of different 2-component links in S^3 (linking number equal to k).
- (c) Consider a connected sum of $\hat{\Delta}^{2n+1}$ and $\hat{\Delta}^{2m+1}$. For fixed $n + m$ one has an infinite family of different 3-component links with the same Jones polynomial. For $m = -n - 1$ one get the family of links with the Jones polynomial of the connected sum of right handed and left handed Hopf links (thus $(t + t^{-1})^2$).

□

We do not know which e_n can be realized by framed links in a solid torus, however we have:

Lemma 7.9 *If e_n is realized by a framed link then $n = 2^k - 1$ for some k .*

Proof: Consider the standard embedding of a solid torus in S^3 . Then $\langle e_n \rangle_{A=-1} = (-1)^{n-1} \frac{n+1}{2}$ on the other hand for a link L one has $\langle L \rangle_{-1} = (-2)^{\text{com}(L)-1}$. Thus $n = 2^k - 1 \square$

References

- [Bi] J.S.Birman, *Braids, Links and mapping class groups*, Ann.Math. Studies 82, Princeton N.J.: Princeton Univ. Press, 1974.
- [BGKT] M.Boshernitzan, G.Galperin, T.Kröger, S.Troubetzkoy, Some remarks on periodic billiard orbits in rational polygons, preprint, SUNY Stony Brook, 1994.
- [BHJS] M.G.V. Bogle, J.E. Hearst, V.F.R. Jones, L. Stoilov, Lissajous knots, *Journal of Knot Theory and Its Ramifications*, 3(2), 1994, 121-140.
- [B-L-M] R. D. Brandt, W. B. R.Lickorish, K. C. Millett, A polynomial invariant for unoriented knots and links, *Invent. Math.*, 84 (1986), 563-573.
- [Bu] W.Burau, Über Zopfgruppen und gleichsinnig verdrillte Verkettungen, *Abh. Math. Sem. Hansischen Univ.*, 11, 171-178, 1936.
- [D-T] C. H. Dowker, M. B. Thistlethwaite, Classification of knot projections, *Topology and its applications* 16 (1983), 19-31, and Tables of knots up to 13 crossings (computer printout).
- [Ho] C. F. Ho, A new polynomial for knots and links; preliminary report. *Abstracts AMS* 6(4) (1985), p 300.
- [H-K] R.Hartley, A.Kawauchi, Polynomials of amphicheiral knots, *Math. Ann.*, 243, 1979, 63-70.
- [H-P-1] J. Hoste, J.H. Przytycki, An invariant of dichromatic links, *Proc. Amer. Math.Soc.*, 105(4) 1989, 1003-1007.
- [H-P-2] J. Hoste, J.H. Przytycki, A survey of skein modules of 3-manifolds, in *Knots 90*, Proceedings of the International Conference on Knot Theory and Related Topics, Osaka (Japan), August

- 15-19, 1990, Editor A. Kawauchi, Walter de Gruyter 1992, 363-379.
- [Jo] V. Jones, Hecke algebra representations of braid groups and link polynomials, *Ann. of Math.* (1987) 335-388.
- [Ka] L.H. Kauffman, An invariant of regular isotopy, *Trans. Amer. Math. Soc.*, 318(2), 1990, 417-471.
- [J-P] V.F.R.Jones, J.H.Przytycki, Lissajous knots and billiard knots, Banach Center Publications, Vol. 42, *Knot Theory*, 1998, to appear.
- [K-P] J.Kania-Bartoszyńska, J.H.Przytycki, 3-manifold invariants and periodicity of homology spheres, preprint 1997.
- [La] C.Lamm, There are infinitely many Lissajous knots, *Manuscripta Mathematica*, 93, 1997, 29-37.
- [L-O] C.Lamm, D.Obermeyer, A note on billiard knots in a cylinder, preprint, October 1997.
- [Li] W.B.R.Lickorish. Polynomials for links. *Bull. London Math. Soc.* 20 (1998) 558-588.
- [L-M] W.B.R. Lickorish, K. Millett, A polynomial invariant of oriented links, *Topology* 26(1987), 107-141.
- [Mu-1] K. Murasugi, On periodic knots, *Comment Math. Helv.*, 46(1971), 162-174.
- [Mu-2] K. Murasugi, Jones polynomials of periodic links, *Pac. J. Math.*, 131 (1988), 319-329.
- [Na] S.Naik, New invariants of periodic knots, *Math. Proc. Cambridge Philos. Soc.* 122 (1997), no. 2, 281-290.
- [Per] K.A.Perko, Invariants of 11-crossing knots, *Publications Math. d'Orsay*, 1980.
- [P-1] J.H.Przytycki, Survey on recent invariants in classical knot theory, *Warsaw University Preprints* 6,8,9; 1986.

- [P-2] J.H.Przytycki, Skein modules of 3-manifolds, *Bull. Ac. Pol.: Math.*; 39(1-2), 1991, 91-100.
- [P-3] J.H.Przytycki, On Murasugi's and Traczyk's criteria for periodic links. *Math Ann.*, 283 (1989) 465-478.
- [P-4] J.H.Przytycki, Vassiliev-Gusarov skein modules of 3-manifolds and criteria for periodicity of knots, Low-dimensional topology (Knoxville, TN, 1992), 143-162, Conf. Proc. Lecture Notes Geom. Topology, III, Internat. Press, Cambridge, MA, 1994.
- [P-5] J.H.Przytycki, An elementary proof of the Traczyk-Yokota criteria for periodic knots, *Proc. Amer. Math. Soc.* 123 (1995), no. 5, 1607-1611.
- [Ro] D. Rolfsen, *Knots and links*. Publish or Perish, 1976.
- [Sm] *The Smith Conjecture* (J.W.Morgan, H.Bass). New York Academic Press 1984.
- [Ta] S.Tabachnikov, *Billiards* Panor. Synth. No. 1, 1995, vi+142 pp.
- [Thur] W.Thurston, The Geometry and Topology of 3-manifolds, Mimeographed notes, 1977-1979.
- [T-1] P.Traczyk, 10_{101} has no period 7: a criterion for periodic links, *Proc. Amer. Math. Soc.* 180 (1990) 845-846.
- [T-2] P.Traczyk, A criterion for knots of period 3, *Topology Appl.* 36 (1990), no. 3, 275-281.
- [T-3] P.Traczyk, Periodic knots and the skein polynomial, *Invent. Math.*, 106(1), (1991), 73-84.
- [Ve] D.L.Vertigan. On the computational complexity of Tutte, Jones, Homfly and Kauffman invariants. PhD. thesis, 1991.
- [We] Hermann Weyl, *Symmetry*, Princeton University Press, 1952.
- [Yo-1] Y.Yokota, The Jones polynomial of periodic knots, *Proc. Amer. Math. Soc.* 113 (1991), no. 3, 889-894.

- [Yo-2] Y.Yokota, The skein polynomial of periodic knots, *Math. Ann.*, 291 (1991), 281-291.
- [Yo-3] Y.Yokota, The Kauffman polynomial of periodic knots, *Topology* 32 (1993), no. 2, 309-324.
- [Yo-4] Y.Yokota, Polynomial invariants of periodic knots, *J. Knot Theory Ramifications* 5 (1996), no. 4, 553-567.

Address:

Department of Mathematics
George Washington University
Washington, DC 20052
e-mail: przytyck@math.gwu.edu

