

XVIth International Congress on
MATHEMATICAL PHYSICS

edited by *Pavel Exner*

 World Scientific



XVIth International Congress on
MATHEMATICAL PHYSICS

This page intentionally left blank



XVIth International Congress on **MATHEMATICAL PHYSICS**

Prague, Czech Republic 3 – 8 August 2009

edited by

Pavel Exner

Doppler Institute, Prague, Czech Republic

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

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.

Cover image © Seán Duggan (1997)

**XVITH INTERNATIONAL CONGRESS ON MATHEMATICAL PHYSICS
(With DVD-ROM)**

Copyright © 2010 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-13 978-981-4304-62-7

ISBN-10 981-4304-62-X

Printed in Singapore.

PREFACE

While mathematical physics has deep roots reaching back to the nineteenth century, congresses in this discipline have convened since the early 1970's. New challenges which appeared at that time led to a stronger community sense and a need to exchange ideas regularly. In the wake of the first three conferences the International Association of Mathematical Physicists was born which made them subsequently its flagship meetings. The name was changing from issue to issue at the beginning, but since the 1980's it became an International Congress on Mathematical Physics, and the intervals between the congresses stabilized to every three years.

As the field is truly global, the previous fifteen congresses met at different places on several continents. After the last one convening for the first time in Latin America, the present ICMP returned to Europe. This time the place was in the central part of the continent where the congress had never been held before even if the region has strong mathematical-physics traditions. We hope that they will find their continuation and that the present gathering of numerous bright mathematicians and physicists contributed to this effect.

The International Scientific Committee and the topical session organizers worked hard to find a selection of interesting talks which would cover the whole spectrum of the field. At the same time the organizers made effort to have a strong bottom-up component in the form of contributed talks selected from the participant proposals. Every congress has some topics which attract a lot attention as a result of a local appeal, hard to predict in advance. The Prague ICMP had particularly large numbers of contributions in dynamical systems, PDE's, and quantum mechanics.

This proceedings volume tries to reflect the congress programme as faithfully as possible. Following the tradition, the main part of the contents are papers based on the plenary talks – all but one – and invited session talks – about two thirds of them. The remaining invited talks, whose authors decided not to submit a paper, and the contributed ones are present through their abstracts and we also list the exhibited posters. It was our aim also to prepare the volume reasonably soon after the meeting before its record would have only a historical value. I am grateful to Petr Ambrož, Jaroslav Dittrich, Jean Downes, David Krejčířík, and Jiří Lipovský whose assistance helped me to meet the set deadline.

In order to give a more complete picture of the congress activities, we attach to this volume a disk where you can find a lot of additional information. In the first place it includes most of the presentations of the invited and contributed talks as they appeared on the screen. An overwhelming majority of these are *pdf* or *ppt* files

prepared by the speakers, in a few cases scanned transparencies. The disk space allows us, however, to show more of the congress programme than just the talks.

An important counterpart to the ICMP was in the last decade the Young Researcher Symposium, since 2003 held as a separate smaller meeting in the same location preceding the congress. We followed the tradition and organized the symposium on July 31 – August 1, 2009. Its programme consisted of six plenary lectures by distinguished mathematical physicists together with more than twenty seminar talks given by the young participants; most of the presentations are included.

The congresses are a natural opportunity to honor exceptional work in mathematical physics. The main IAMP distinction, the Henri Poincaré Prize sponsored by the Daniel Iagolnitzer Foundation was awarded for the fifth time; it went to Jürg Fröhlich, Robert Seiringer, Yakov Sinai, and Cédric Villani. This year it was accompanied by the newly instituted IAMP Early Career Award, won by Mihalis Dafermos, and at the Young Researcher Symposium the IUPAP Young Scientists Prizes in mathematical physics were given. The information together with laudatias for the Poincaré Prize laureates are given in the opening of the volume.

The introductory part also lists members of the congress committees, dealing with programme, prizes, and organizational matters. One has to add the session organizers named at the opening of each topical session chapter. The work of all these colleagues, often quite demanding, was crucial for the congress success. Equally sincere thanks belong to the congress sponsors listed in the opening; it was their support, financial and material, which made it possible to organize the meeting in a way accessible to more than six hundred members of our community.

The disk also reflects other parts of the programme. There was a round table discussion on future challenges in mathematical physics. Joel Lebowitz organized traditionally a session devoted to solidarity issues; given the place and time it was an opportunity to recall the twentieth anniversary of the sweeping changes which altered the lives of many of us. Not to forget, the IAMP history was recalled by means of a small exhibition featuring the fifteen previous congresses, and there was also a social and cultural programme.

All in all, I believe that the volume in your hands represents a good summary of the XVIth International Congress on Mathematical Physics in Prague, conveying the essential part of its scientific contents and also something of its atmosphere.

Prague, December 2009

Pavel Exner
ICMP Convener

CONGRESS COMMITTEES

INTERNATIONAL SCIENTIFIC COMMITTEE

Michael Aizenman
 Ingrid Daubechies
 Pavel Exner
 Jürg Fröhlich
 Giovanni Gallavotti, *chair*
 Yakov G. Sinai
 Jan Philip Solovej

INTERNATIONAL ADVISORY COMMITTEE

Yosi Avron
 David Brydges
 Marcus Büttiker
 E. Brian Davies
 Eduard Feireisl
 Svetlana Jitomirskaya
 Roman Kotecký
 Ari Laptev
 Bruno Nachtergaele
 Leonid Pastur
 Vincent Rivasseau
 Vladas Sidoravicius
 Herbert Spohn
 Pavel Štěda
 Kenji Yajima
 Jakob Yngvason
 Jean-Claude Zambrini

LOCAL ORGANIZING COMMITTEE

Petr Ambrož
 Jaroslav Dittrich
 Pavel Exner, *chair*
 Miloslav Havlíček
 Ladislav Hlavatý
 Jiří Horáček
 Jiří Hořejší
 Igor Jex
 Roman Kotecký
 Jan Kratochvíl
 David Krejčířík
 Olga Krupková
 Zuzana Masáková
 Josef Málek
 Edita Pelantová
 Václav Špička
 Miloš Tater
 Jiří Tolar
 Leopold Vrána
 Miloš Zahradník
 Miloslav Znojil

PRIZE COMMITTEES

HENRI POINCARÉ PRIZE COMMITTEE

Eric Carlen
Giovanni Gallavotti
Krzysztof Gawedzki
Bruno Nachtergaele, *chair*
Herbert Spohn

IAMP EARLY CAREER AWARD COMMITTEE

Jürg Fröhlich
Svetlana Jitomirskaya
Israel Michael Sigal
Jan Philip Solovej, *chair*
Horng-Tzer Yau

IUPAP YOUNG SCIENTIST PRIZE COMMITTEE

Peter Bouwknegt
Ana Bela Cruzeiro
Pavel Exner, *chair*
Antti Kupiainen
Michael Loss

SPRINGER BEST POSTER PRIZE COMMITTEE

Goce Chadzitaskos
Ira Herbst, *chair*
Olga Krupková
Herbert Spohn
Domokos Szász

ACKNOWLEDGMENTS

SPONSORSHIP

MAIN SPONSORS

Doppler Institute for Mathematical Physics and Applied Mathematics
 NSF – National Science Foundation
 DFG – Deutsche Forschungsgemeinschaft
 IUPAP – International Union of Pure and Applied Physics
 Foundation for Support of Theoretical Physics, 51601 Slemeno 33, Czech Republic
 ESF – European Science Foundation
 Centre of Particle Physics, Charles University and Czech Academy of Sciences
 Fondation Daniel Iagolnitzer

OTHER SPONSORS

Jindřich Nečas Centre of Mathematical Modeling
 Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University
 Annales Henri Poincaré
 ICIAM – International Council for Industrial and Applied Mathematics
 Czech Academy of Sciences – Council of Foreign Relations
 Institute of Theoretical Physics, Charles University
 IMU - International Mathematical Union
 JČMF – Union of Czech Mathematicians and Physicists
– as well as several anonymous sponsors

PUBLISHER BOOTHS

Springer
 Journal of Mathematical Physics
 IOP Publishing
 American Mathematical Society
 Birkhäuser
 Cambridge University Press
 European Mathematical Society Publishing House

ACADEMIC AUSPICES

President of the Academy of Sciences of the Czech Republic
Rector of the Charles University in Prague
Rector of the Czech Technical University in Prague

PROFESSIONAL PARTNER

Congress Business Travel, Ltd., Lidická 43/66, 15000 Prague, Czech Republic
<http://www.cbtttravel.cz/>

PAST CONGRESSES

Moscow 1972
Warsaw 1974
Kyoto 1975
Rome 1977
Lausanne 1979
Berlin 1981
Boulder 1983
Marseille 1986
Swansea 1988
Leipzig 1991
Paris 1994
Brisbane 1997
London 2000
Lisbon 2003
Rio de Janeiro 2006

CONTENTS

Preface	v
Congress Committees	vii
Prize Committees	viii
Acknowledgments	ix
Welcome addresses	1
Henri Poincaré Prize	5
IAMP Early Career Award and other prizes	17
 Part A Plenary Talks	 21
Quantum geometry of 3-dimensional lattices and tetrahedron equation <i>V.V. Bazhanov, V.V. Mangazeev, S.M. Sergeev</i>	23
The formation of black holes in general relativity <i>D. Christodoulou</i>	45
Liouville quantum gravity & the KPZ relation: a rigorous perspective <i>B. Duplantier</i>	56
Universality of Wigner random matrices <i>L. Erdős</i>	86
Uses of free probability in random matrix theory <i>A. Guionnet</i>	106
The physics of decision making: stochastic differential equations as models for neural dynamics and evidence accumulation in cortical circuits	123

*P. Holmes, P. Eckhoff, K.F. Wong-Lin, R. Bogacz,
M. Zacksenhouse, J.D. Cohen*

New technologies in the hunt for new physics	143
<i>D.A. Kosower</i>	

Operator algebras and noncommutative geometric aspects in conformal field theory	161
<i>R. Longo</i>	

Kinetic transport in crystals	162
<i>J. Marklof</i>	

Ising models, universality and the non renormalization of the quantum anomalies	180
<i>V. Mastropietro</i>	

The infrared problem in nonrelativistic QED	200
<i>A. Pizzo</i>	

Parking in the city: an example of limited resource sharing	222
<i>P. Šeba</i>	

Hot topics in cold gases	231
<i>R. Seiringer</i>	

Vortex patterns in Ginzburg-Landau minimizers	246
<i>S. Serfaty, E. Sandier</i>	

Quantization of integrable systems and four dimensional gauge theories	265
<i>N.A. Nekrasov, S.L. Shatashvili</i>	

Changing views of quantum field theory	290
<i>S. Weinberg</i>	

Part B Topical Sessions **311**

Dynamical Systems (including integrable systems and Hamiltonian stability)	313

Almost dense orbit on energy surface	314
<i>V. Kaloshin, K. Zhang, Y. Zheng</i>	

Dissipative perturbations of KdV <i>S.B. Kuksin</i>	323
Billiard models and energy transfer <i>Zs. Pajor-Gyuliai, D. Szász, I.P. Tóth</i>	328
Quantum Calogero-Moser systems: a view from infinity <i>A.N. Sergeev, A.P. Veselov</i>	333
Other talks	338
Posters	343
Equilibrium Statistical Mechanics (classical and quantum)	345
Renormalization group trajectories between two fixed points <i>A. Abdesselam</i>	346
Gradient models with non-convex interactions <i>S. Adams</i>	352
Supersymmetric Ward identities in quantum diffusion <i>M. Disertori</i>	357
Off-critical lattice models and massive SLEs <i>N. Makarov, S. Smirnov</i>	362
Disorder relevance for pinning/wetting models: a review <i>F.L. Toninelli</i>	372
Other talks	377
Posters	379
Non-Equilibrium Statistical Mechanics (including stochastic evolutions except SLE and Boltzmann equation)	380
Diffusion in coupled map lattices <i>A. Kupiainen</i>	381
Decay of equilibrium time correlations in a weakly nonlinear Schrödinger equation <i>J. Lukkarinen, H. Spohn</i>	386

Lieb-Robinson bounds and the existence of infinite system dynamics <i>B. Nachtergaele</i>	391
Heat equation from microscopic dynamics: a weak coupling approach <i>C. Liverani, S. Olla</i>	397
KPZ universality for KPZ <i>J. Quastel</i>	401
Derivation of effective evolution equations from many body quantum dynamics <i>B. Schlein</i>	406
Other talks and posters	417
PDE's (fluid dynamics, wave equation and NLSE, Boltzmann equation, general relativity, and elasticity)	420
A new physical-space approach to decay for the wave equation with applications to black hole spacetimes <i>M. Dafermos, I. Rodnianski</i>	421
Nonlinear regularizing effect for hyperbolic partial differential equations <i>F. Golse</i>	433
On the Maxwell-Klein-Gordon system in Lorenz gauge <i>S. Selberg, A. Tesfahun</i>	438
A regularity theorem for energy critical wave-maps <i>J. Sterbenz, D. Tataru</i>	444
Other talks	452
Posters	460
Probability Theory (including random matrices and SLE)	463
Glauber dynamics for spin systems at high and critical temperatures <i>E. Lubetzky</i>	464
On the scaling limits of random planar maps with large faces <i>J.-F. Le Gall, G. Miermont</i>	470

The scaling limit of the Minimal Spanning Tree — a preliminary report <i>G. Pete</i>	475
Scaling limits of Dyson's β -ensemble <i>B. Valkó</i>	481
Other talks and posters	486
Operator Algebras (including noncommutative geometry)	489
Deformations of operator algebras and the construction of quantum field theories <i>G. Lechner</i>	490
On superselection theory of quantum fields in low dimensions <i>M. Müger</i>	496
Quantizations of Poisson Lie groups as noncommutative manifolds <i>S. Neshveyev, L. Tuset</i>	504
Positivity in Rieffel's strict deformation quantization <i>S. Waldmann</i>	509
Other talks and posters	514
Nonrelativistic Quantum Mechanics (including quantum information)	517
Radiationless transitions through avoided crossings <i>V. Betz</i>	518
Equivalence of Sobolev inequalities and Lieb–Thirring inequalities <i>R.L. Frank, E.H. Lieb, R. Seiringer</i>	523
Entanglement spread and clean resource inequalities <i>A.W. Harrow</i>	536
Optimal decoupling <i>R. Renner</i>	541
Entanglement at finite temperature in the electronic two-particle interferometer <i>P. Samuelsson, I. Neder, M. Büttiker</i>	546

Spectral properties of discrete alloy-type models <i>M. Tautenhahn, I. Veselić</i>	551
Complete dynamical localization in disordered quantum multi-particle systems <i>M. Aizenman, S. Warzel</i>	556
Other talks	566
Posters	571
String Theory and Quantum Gravity	573
Extremal black holes and nilpotent orbits <i>G. Bossard</i>	574
Black hole spacetimes with Killing–Yano symmetries <i>D. Kubizňák</i>	580
Maximal supersymmetry and non-Abelian gauge groups <i>H. Samtleben</i>	588
Other talks and posters	593
Quantum Field Theory (including field theoretic methods in condensed matter)	595
Three études in QFT <i>D. Kreimer</i>	596
Graphene: relativistic transport in a nearly perfect quantum liquid <i>M. Müller, L. Fritz, S. Sachdev, J. Schmalian</i>	602
Defect lines, dualities and generalised orbifolds <i>J. Fröhlich, J. Fuchs, I. Runkel, Ch. Schweigert</i>	608
Algebraic Bethe Ansatz approach to correlation functions: from exact representations on the lattice to asymptotic behavior <i>V. Terras</i>	614
Other talks	619
Posters	622

Many-Body Quantum Mechanics (condensed matter, nanoscale, elasticity)	624
Strength of superconductivity close to critical magnetic field <i>A. Fournais, A. Kachmar</i>	625
The 2D Hubbard model on the honeycomb lattice <i>A. Giuliani</i>	630
Quantum phases of supersymmetric lattice models <i>L. Huijse, K. Schoutens</i>	635
Other talks and posters	640
New Frontiers in Mathematical Physics	644
Glimpses into the early Universe <i>D. Langlois</i>	645
Spatiotemporal dynamics in ecology: insights from physics <i>J.A. Sherratt, M.J. Smith, J.D.M. Rademacher</i>	651
Other talks	656
Posters	661
Part C Supplementary Programme	663
Young Researcher Symposium	665
YRS plenary talks	666
YRS seminar talks	669
Round Table “Mathematical Physics: Future Challenges”	671
Human Rights and Social Responsibility	674
Satellite Meetings	676
Registered participants	678
Speaker Index	689

This page intentionally left blank

WELCOME ADDRESSES

WELCOME ADDRESS BY JIŘÍ DRAHOŠ PRESIDENT OF THE ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Distinguished colleagues, ladies and gentlemen,

Let me welcome all of you in this country on behalf of all branches of the Czech scientific community. These include the Academy of Sciences representing the national system of research institutes, two major Prague universities whose rectors are present here, and numerous other scientific establishments.

The quest for understanding how things work, be it the universe, an elementary particle, or the interior of our body and mind, was always one of the main driving forces of the progress. Europe made many contributions to it since the days of Antique, and while we had our dark periods, science was the element which gave the dynamics to our part of the world since the Renaissance times. It is pleasant to remark that the city hosting you is a place where important physicists and mathematicians worked in different historical periods. This is the place where Kepler discovered the laws of planetary motion, Bolzano laid the “epsilon-delta” grounds of mathematical analysis, Doppler formulated his principle, and Einstein did an early work on the general theory of relativity.

The importance of serious scientific investigations is equally vital nowadays, or maybe even its importance increases in the world which is perceived as smaller and increasingly competitive. To use the full potential of scientific tools one has to respect structure and complexity of the scientific work. This is important to say in particular at a gathering of scientists belonging to a domain which is predominantly of theoretical nature like yours.

In a democratic society scientists have to make efforts to find means to conduct research being dependent on wisdom of politicians, or a lack of it. In many countries including the one you are visiting now, pressures are encountered to support only scientific activities which promise short-time benefits, at the expense of curiosity-driven research and establishing consistency of our model of the world from first principles. It has to be said aloud and repeated that such demands are deeply flawed: they forget that with its roots cut the flower of science will inevitably wither soon.

Your discipline belongs to those inhabiting a border region between two major science branches. Being myself a physical chemist I understand well that such a position is both advantageous and difficult, especially if the two elephants in question have a history which can be traced back to the dawn of civilization. Sitting on a ridge, one can find a lot of inspiration on both its sides but at the same time, it needs an effort to convince people from core areas of the two sciences about the importance of your undertaking. Looking at the programme of your congress I am sure that the community of mathematical physicists is doing well in this respect.

Major congresses like yours give an opportunity to see the global picture of activities in the field. As a representative of Czech science I am glad to see that Czech mathematical physicists are an active part of your community. Science is one of the activities of a truly worldwide character and local teams should strive to stand in good stead without regard to national or any other boundaries.

To conclude this opening address let me wish to all of you a successful congress, a lot of interesting lectures, fruitful exchanges and a lot of inspiration, as well as a pleasant stay in Prague.

OPENING SPEECH BY PAVEL EXNER PRESIDENT OF IAMP

Mr. President, Magnificences, distinguished colleagues, dear guests,

Let me welcome all of you on behalf of the Local Organizing Committee to the opening of 16th International Congress on Mathematical Physics. With a particular pleasure I welcome here members of the International Scientific Committee, former presidents of the International Association of Mathematical Physics and a number of its longtime members.

Before coming to the main message let me mention with a sorrow that since our last meeting three years ago in Rio some distinguished members of the mathematical-physics community passed away, in particular, Vladimir Geyler, Jan Lopuszanski, Oded Schramm and Mikhail Birman. I ask you to stand up and honor their memory by a few moments of silence.

Let me turn now to the matter which brought us together. If I may borrow a term from Gabriel García Márquez this is a congress in the time of cholera. The economy in almost any country of the world is in serious trouble and the consequences are felt. True, previous congresses also had their difficult moments. There was a political turmoil in Germany, broke national airline in Brazil, etc., but never before the trouble was of global character. It cost us a lot of participants – many colleagues told us with regrets they were unable to come due to financial restrictions imposed by their institutes. It also had a serious impact on the sponsor

support: despite various promises and our concentrated efforts we did not get finally a single penny from a private source unrelated to science.

This is the bad news, but there is also good news. Tough times are useful as a test of the community strength and there is no doubt that we can be pleased from what we see. Even if the numbers are lower than projected, more than six hundred people registered for the congress making it the third largest in history. Secondly, the sponsor support we obtained from academic and other science related bodies, both Czech and international, was substantial and in fact it surpassed our cautious estimates. It is encouraging indeed to see that the academic community functions, and it is a strong message to all those who give us lectures about how research institutions should be run as industrial corporations.

I want to express our sincere and profound gratitude to the sponsors the names of which will be displayed repeatedly during the congress. We list them in the order of their contribution size, however, our thanks go to all of them equally; we are well aware that the sums have to be measured against the sponsors own budget. We also received from colleagues support of various type which cannot be labeled as sponsorship but helped us a lot.

All this support was one of the main factors which allowed us not only to keep the promises about the registration we made three years ago when making in Rio a bid for this congress but in fact to do a bit better. It made it also possible to offer a support of various degree to many participants from economically weak countries, and we are happy that about sixty people were able to use this offer.

Since I have two roles here, apart from being responsible for local organization I have the honor to be the IAMP President – an intriguing combination I would say – allow me to say a few words in my other capacity.

When I prepared for this event I looked at what my predecessors said at the similar occasion and found that there is a pair of recurrent themes. One of them I would call a gathering-cloud claim: mathematical physics is doing well – or at least it became “socially acceptable” as it was put famously by Arthur Wightman twenty three years ago – but dangers are looming at the horizon, namely our material base is threatened by requirements to reorient the research on more practical things bringing an immediate effect.

Such dangers do exist and they are serious, an example was quoted just a while ago by Mr. President in his welcome address. More than that, they are not a temporary affair. We have to live with the fact that this push is more or less permanent even if it may be lighter or harder at different periods. Many people, including those wielding a political power influencing our fates, do not realize that, say, without the seemingly useless investigations of the atomic structure of matter two generations ago most gadgets they like so much simply would not be here. It is not enough to produce deep and interesting results. Whether we want it or not, a part of our *métier* is a permanent effort to convince the society outside science borders that what we are doing is important, even very important.

The second recurrent theme at the ICMP openings is the play with the Association logo: $M \cap \Phi$? Rather $M \cup \Phi$! As any good joke it contains a deep reflection. We should keep a mathematical-physics perspective while being able to roam through the wide territory of both mathematics and physics, to say nothing about forays into other disciplines. It is not easy, of course, because people inhabiting these areas speak different languages, which may come from the same root as various European languages come from Latin, but still they are loaded with different experience and they need an effort to be understood properly.

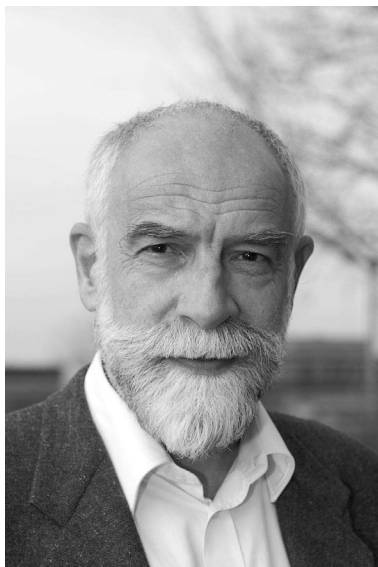
For a not very large discipline on the border between broad realms as ours an openness is a *conditio sine qua non*. The ability of detecting interesting new problems which allow to demonstrate the power of our instruments in a wider scientific environment, so to say, to our scientific cousins, is a guarantee that our field will remain lively and attractive.

Our triennial congresses represent an opportunity to get in contact with many colleagues and ideas we do not meet everyday. I wish you that such encounters bring you a lot of inspiration through fruitful discussions in a pleasant environment. With this let me declare the 16th International Congress on Mathematical Physics open.

**THE HENRI POINCARÉ PRIZE SPONSORED
BY THE DANIEL IAGOLNITZER FOUNDATION**

The prize has been awarded since 1997 to recognize outstanding contributions in mathematical physics, and contributions which lay the groundwork for novel developments in this broad field. The prize was also created to recognize and support young people of exceptional promise who have already made outstanding contributions to the field of mathematical physics.

THE 2009 LAUREATES



JÜRG FRÖHLICH

Citation: For his fundamental contributions to quantum field theory and statistical mechanics which led to major advances in our understanding of physical systems with many degrees of freedom.

LAUDATIO BY KLAUS HEPP

When I was asked to give the laudatio for Jürg Fröhlich for his Poincaré Prize 2009 I immediately accepted this honor without realizing my total incompetence, having worked as a neuroscientist for the second half of my life. How could I give to this expert audience nontrivial insights, for instance about the connection between ‘Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory’, about which topic Jürg is so proud that he and his collaborators have devoted to it an advanced monograph (Springer 1992), well known to many graduate students here? Nevertheless I will try to do my best for the proceedings of this congress, at least by correctly spelling most of the key words.

Among people in this audience I probably know Jürg for the longest time. With the help of G Felder and GM Graf, and the support of W Beiglböck, I was recently taken by Jürg on ‘A Journey through Statistical Physics’, in the task of editing an impressive *Selecta Volume* (Springer, 2009) which you should all admire at the Springer book exhibit.

Jürg Fröhlich was born on July 4, 1946, at Schaffhausen in northern Switzerland. From 1965 to 1972 he studied mathematics and physics at the ETH Zürich and obtained there with honors a Diploma and a PhD in physics, with R Schrader and myself as midwives and R Jost and M Fierz as grandparents. Already in this period Jürg showed an intriguing complexity: As a brilliant student he felt infinitely old, charged by a load problems to keep him busy till to the end of his life. Still, as you can see, he is incredibly young even now! Moreover, although he is internationally widely visible as a famous lecturer at congresses and summer schools, he always lived in the deep infrared, in his PhD thesis on the infrared problem in E Nelson’s model, later when deriving with B Simon and T Spencer the powerful infrared bounds in statistical mechanics and quantum field theories, and today when analysing nonrelativistic quantum electrodynamics, as you will hear in A Pizzo’s invited talk at this conference.

Jürg’s *Wanderjahre* from 1972 to 1982 were probably the most beautiful period of his life. He got married to Eva Schubert and has two daughters. His peregrinations took him first to J-P Eckmann in Geneva, then to A Jaffe at Harvard University, then to E Lieb, E Nelson, E Seiler, T Spencer, B Simon and A Wightman at Princeton, and finally as ‘professeur permanent’ to the IHES near Paris. In the fall of 1982 he started as the successor of R Jost in the Physics Department of ETH Zürich, to which he kept his fidelity.

Characteristic for Jürg’s style is the attention to mathematical precision combined with a solid physical motivation. While it is hopeless to deduce in a mathematically rigorous way the behavior of a real physical system from the basic laws, Jürg has obtained many interesting limiting laws using powerful mathematical theories, by introducing ideas from quantum field theory into statistical mechanics and condensed matter physics. Helpful was his insight that research in physics is most

fruitful and entertaining as a collective enterprise. This can be seen in the list of more than 144 co-authors in more than 284 publications, as counted by A Jaffe for his wonderful slide-show ‘Constructive Jürg’, a highly valuable Internet download.

When we edited the Selecta Volume we asked ourselves what of the Alpine landscape of Jürg’s oeuvre we should include and where we should locate the highest peaks. For me the papers ‘The Kosterlitz-Thouless Transition in Two-Dimensional Abelian Spin Systems and the Coulomb Gas’ and the ‘Absence of Diffusion in the Anderson Tight Binding Model for Large Disorder or Low Energy’ are definitely of ‘Matterhorn’ format. Both were climbed together with T Spencer ‘in winter and without gloves’, as A Connes would say. As a neurobiologist living on treacherous ground I also enjoyed the acrobatics leading from Chern-Simons gauge theory to quantum Hall systems in a collaboration with T Kerler.

Jürg’s work has been internationally recognized by many prizes: the ‘Latsis Prize’ of the SNF, the ‘Heineman Prize’ of the APS, the Swiss ‘M Benoist Prize’, the German ‘Max-Planck Medal’ and now the ‘Henri Poincaré Prize’. He has honorable affiliations with the IHES in Paris and the University of Zürich.

Since 1982 he ‘has been struggling with trying to fulfil the diverse obligations and duties of an ETH professor, with varying success’, as he modestly described it. This is a profound understatement: Jürg has been instrumental for creating a ‘Center for Theoretical Studies’ at the ETH Institute for Theoretical Physics, he has been a powerful Chairman of the ETH Physics Department for many years of institutional turmoil and, by combining insights from old age with the fervour of the 68-generation, he has aided the governors of ETH to permit us to do meaningful work, also in mathematical physics.

Jürg is carried by the principle of Confucius of the unity of knowledge and action, as formulated by Wang Yangming (1472-1529) in the early 16th century in China: ‘Knowing is the intent of acting and acting is the work of knowing. Knowing is the beginning of acting and acting is the completion of knowing’.

Congratulations and thank you, Jürg!



ROBERT SEIRINGER

Citation: For his major contributions to the mathematical analysis of low temperature condensed matter systems, in particular for his work on Bose condensation and the Gross-Pitaevskii equation.

LAUDATIO BY JAKOB YNGVASON

It is my privilege to present to you Robert Seiringer who receives the Henri Poincaré Prize for his ‘major contributions to the mathematical analysis of low temperature condensed matter systems’. Triggered by experimental advances, cold quantum gases and Bose-Einstein condensation came into the focus of attention of physicists about 15 years ago and this topic is currently one of the most active research areas in physics. Mathematical physics has here an important role to play because due to the complexity of the quantum many-body problem some of the most fundamental questions can only be settled by penetrating mathematical analysis. It is here that Robert has been a key player and contributed deep insights.

It is now ten years since Robert came to me inquiring about a subject for a Master thesis. It so happened that Elliott Lieb and I had a little earlier written a paper on the ground state energy of an interacting, dilute Bose gas and were thinking about the extension to inhomogeneous systems. Physicists usually describe such systems by the Gross-Pitaevskii equation that is a non-linear Schrödinger equation expected to provide a good approximation to the many-body problem in the case of dilute gases. But no rigorous derivation starting from the full many-body problem existed yet. As a first task I suggested to Robert to work on the energetic upper

bound. This was on a Friday. A normal, good student would probably have started by studying lots of references and come back after three weeks, asking for further guidance. But I realized that Robert was someone quite out of the ordinary when he presented to me next Monday a beautiful LaTeX document with the complete solution of the problem. Thus Robert was from the very outset a full-fledged research partner rather than a student or junior partner and this first cooperation between Elliott, Robert and me very soon resulted in a joint publication as well as Robert's Master degree. His PhD followed a year later, his thesis containing besides further contributions to the theory of Bose gases some beautiful results about quite a different topic, namely atoms in strong magnetic field, partly obtained in collaboration with Bernhard Baumgartner.

At the age of 33 Robert is at present author or co-author of more than 40 original publications in prestigious journals. These publications are devoted to many different topics but I shall here focus on a few that are particularly linked to the citation for the Henri Poincaré prize.

A notoriously difficult problem in quantum many-body theory is to prove the existence of Bose-Einstein condensation (BEC) for a system of interacting Bosons in the thermodynamic limit. This phase transition is accompanied by spontaneous breaking of a continuous symmetry (gauge symmetry), which explains at least partly the difficulty in proving it. In fact, for a continuous system with realistic interactions this is still an open problem. Experiments with cold atoms are, however, carried out in traps where the gas has a finite extension and a simpler, but still far from an easy problem is to prove Bose-Einstein condensation in an appropriate limit, called the Gross-Pitaevskii (GP) limit, for such systems. In fact, this had been an open problem for quite some time when Robert entered the scene. Robert, who in the meantime had moved from Vienna to Princeton with a scholarship of the Austrian Science Fund, realized that the energy estimates used in the derivation of the Gross-Pitaevskii equation could be generalized to prove that the kinetic energy of the ground state wave function is essentially localized in small regions in configuration space and that this would be the key to a proof of BEC in the GP limit. An additional ingredient, namely a generalization of classical Poincaré inequalities, was also needed, and all this was put together in a beautiful joint publication of Robert and Elliott Lieb in 2002. This was the first derivation of BEC for an interacting system of direct experimental relevance.

The theory of BEC in the GP limit was subsequently generalized to prove also superfluidity in this limit and somewhat later a further outstanding problem, posed six years earlier by Lev Pitaevskii, could be solved, namely the derivation of the GP equation and the proof of BEC for gases in a rotating container. This problem is much more complex than for the non-rotating case due to the occurrence of quantized vortices that can break rotational symmetry. Here the solution was obtained in several steps with two papers of Robert playing a key role. One is his analysis of the Gross-Pitaevskii equation for a rotating system, containing the first general proof of rotational symmetry breaking for this equation. Most work on the GP equation with

rotation is concerned with asymptotic analysis with respect to a small parameter, the ratio of the healing length to the size of the system, but Robert's paper is one of the few that has no such limitations. In a second paper Robert linked the symmetry breaking to another important difference between the rotating and non-rotating case: while in the non-rotating case the absolute ground state, i.e., the ground state without symmetry restrictions, is automatically the Bose ground state, this is in general no longer true for rotating systems. He was nevertheless able to prove that the GP equation provides an upper bound to the bosonic ground state, also in the case of symmetry breaking. The derivation of the GP equation in the rotating case was finally completed when the lower bound was proved in an impressive joint paper with Elliott Lieb. Further work of Robert on rotating Bose gases includes quite recent papers with Mathieu Lewin, and with Elliott and myself on gases that are in such rapid rotation that they become effectively two-dimensional and their state confined to the lowest Landau level.

Besides his work on the ground states, Robert has also been a pioneer in the rigorous study of quantum gases at nonzero temperatures. His work here includes the proof of the next to leading order term in the pressure of a dilute Fermi gas, generalizing previous joint work with Lieb and Solovej on the ground state, and a highly sophisticated derivation of a lower bound to the free energy of a Bose gas, containing next to leading order contributions. A beautiful recent paper with Daniel Ueltschi contains rigorous upper bounds for the transition temperatures for superfluidity in two-dimensions and to BEC in a three dimensions.

As already mentioned, Robert has worked on many topics besides cold quantum gases, including Coulomb systems, quantum electrodynamics and BCS theory, partly in collaboration with another former student of mine, Christian Hainzl. All his work is characterized by depth of the mathematical analysis and elegance and clarity of the presentation. His accomplishments make him already at this early stage of his career a most deserving recipient of the Henri Poincaré Prize and point to a bright future. I congratulate him heartily on this occasion.



YAKOV G. SINAI

Citation: For his ground-breaking works concerning dynamical entropy, ergodic theory, chaotic dynamical systems, microscopic theory of phase transitions, and time evolution in statistical mechanics.

LAUDATIO BY DOMOKOS SZÁSZ

The main fields of interest of Yasha Sinai are ergodic theory, the theory of dynamical systems, probability theory, mathematical, in particular, statistical physics. He has a unique awareness of and sensitivity to the fundamental problems of both mathematics and physics. The problems he raises and attacks with a tremendous proving power if necessary, always lie at the very depth and imply wide and flourishing consequences.

I am more than pleased to challenge here the impossible: to say briefly about the scientific achievements of Yasha, about his contributions to mathematical physics. Let me try nevertheless and start with some of his early accomplishments:

- 1959: Kolmogorov-Sinai entropy
- 1961: dynamical systems with countable Lebesgue spectrum
- 1967: a rigorous theory of phase separation by Minlos and Sinai
- 1968: Markov partitions for uniformly hyperbolic systems
- 1970: Sinai billiard, ergodicity and Kolmogorov-mixing of two hard disks
- 1972: Gibbs measures in ergodic theory

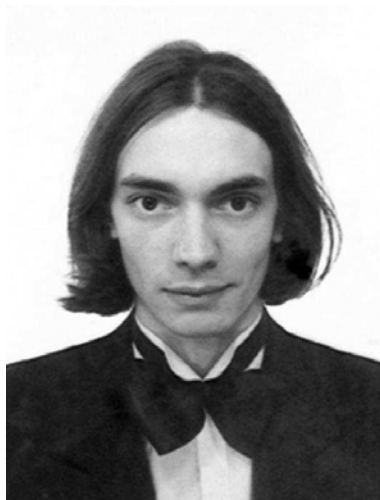
I interrupt this astonishing list here since I, personally, think that these works between 1968 and 1972: the foundations and the basic construction of the highly efficient tool of Markov-partitions for hyperbolic systems, the discovery of the Sinai-Ruelle-Bowen measure and the application of the Gibbs formalism to dynamical systems is the most brilliant example of the width, the depth and also the outflow of Yasha's mathematics. My duty is to continue and do this by only mentioning some of his later, most profound, original and highly influential results:

- proof of existence of Hamiltonian dynamics of infinite particle systems by introducing the idea of "cluster dynamics"
- the far-reaching generalization of Peierls' classical contour method (the well-known Pirogov-Sinai method for describing the phase-diagram in systems with a finite group of symmetry)
- the mathematically rigorous foundations of Kenneth Wilson's renormalization group method (for this method, Wilson was awarded the Nobel prize), this is applicable not only in statistical physics but also in dynamical systems (with Bleher)
- description of the spectrum of the quasi-periodic Schrödinger operator: establishing the localization of the eigenfunctions
- construction of Markov partitions for billiards and the Lorenz-map (with Bunimovich and Chernov)
- for 1-D random walks in a random environment a mathematically rigorous treatment of their subdiffusive behavior (Sinai-diffusion)
- the verification of asymptotic Poisson distribution of energy level gaps for a class of integrable dynamical systems via a very elegant way to treat the related classical and deep number theoretical problems
- the mathematical theory of coupled map lattices (known also as space-time chaos; with Bunimovich and with Pesin)
- hydrodynamics and his theory for the Navier-Stokes equation (with K. Khanin, J. Mattingly and D. Li).

This list obviously demonstrates that Sinai not only solved a number of extremely hard and deep classical problems and built theories. More than that, he is one of those absolutely rare unifying personalities who, by having an overview over a wide spectrum of areas, could initiate directions and theories that connect different fields within mathematics or between mathematics and physics.

Let me mention some personal experiences. Yasha completely revolutionized my interest at least two times. Originally a probabilist, I turned to statistical physics in 1975, and started to regularly visit the seminar of Yasha and that of Dobrushin-Sinai in Moscow. I also invited Yasha to us. His series of lectures held in Budapest in 1976, throughout a whole week, Saturday included, every day until our last breath, actually not his, were composed from the germ of his monograph: Theory of Phase Transitions. Rigorous Results. (He covered among others DLR-equations, theories of Bleher-Sinai, of Pirogov-Sinai and of Dobrushin-Shlosman.) For all of

us in Budapest: József Fritz, András Krámli and Péter Major, this became a twist in our interests. Let me also note in brackets that his four lectures held somewhat later on Gelfand's seminars in Moscow on the topics of this monograph eventually meant a turning point in the judgement of Gelfand's school on probability theory and on mathematical statistical physics. In 1979, before our celebrated Random Fields conference in Esztergom, we were taking sunbath with Yasha on the rooftop of Lukács swimming pool in Budapest, a favorite place of Yasha and Lena. He there told me about Markov partitions of the billiard and the CLT for the Lorentz process. Having been interested in dynamical theories of Brownian motion, exactly afterwards I started to study billiards. And I am most glad that quite recently, with Dolgopyat and Varju, we could finally solve a problem, he raised to me 28 years ago, on local perturbation of the Lorentz process. Talking about dynamical theory of Brownian motion, it was, indeed, most exciting to work parallelly with him on the Harris-Spitzer model. I was proud to hear from him that he had learnt much about this model from me. Characteristic to his style is what, during this work, he said to me in 1984 in Tashkent: if you hear the equation I have just found you will jump high up to the ceiling. I would have done if I had been able. All in all, Sinai is not only an exceptional scholar of our time, but the experience of our school in Budapest, and also that of further centers, among others, from Italy, from Poland, from Uzbekistan, etc. shows that his enthusiasm toward problems of science, his energy and his direct and friendly personality always attract many mathematicians from all over the world to beautiful, interesting and fundamental problems of mathematics and physics.



CÉDRIC VILLANI

Citation: For his innovative work on kinetic theory and optimal transport with applications to dissipative physical systems and Riemannian geometry.

LAUDATIO BY JOEL LEBOWITZ

It is a great pleasure to participate in this ceremony honoring some of our distinguished colleagues, both young and mature. The fact that these honorees are also good friends, indeed the mature ones very dear old friends, only adds to the pleasure of my task to give the laudatio for Cédric Villani. Cédric is not only young and brilliant he is also a prime example of the type of mathematician we want to join our fraternity of mathematical physicists. Here is an abbreviated c.v. taken from his web page.

Born October 5, 1973 in Brive-la-Gaillarde (France)
 PhD Thesis (1998; advisor P.-L. Lions); Habilitation dissertation (2000)
 Louis Armand Prize of the Academy of Sciences (2001)
 Peccot-Vimont Prize and Cours Peccot of the College de France (2003)
 Jacques Herbrand Prize of the Academy of Sciences (2007)
 Prize of the European Mathematical Society (2008)
 Director Institut Henri Poincaré, July 1, 2009
 hobbies: walking, music (piano)

Cédric started his research with studies of the Boltzmann equation (BE), an equation first “derived” by Ludwig Boltzmann in 1873. More precisely, instead of

derived I should say that it was “based” on a combination of deep physical insight and mathematical bravado. The equation describes the deterministic evolution of the smoothed or coarse grained empirical density in the six-dimensional space of position and velocity of an isolated, dilute, gas of particles. The dynamics are Newtonian with interactions idealized as succession of instantaneous binary collisions. It was and remains a paradigm of what mathematical-physics, at its best, can and should be. It is beautiful, challenging and it works. Cédric’s devotion to the study of the solution of the BE and more recently to that of the related Landau/Vlasov equation are therefore in the best tradition of mathematical physics.

Examination of implications of the BE started with Boltzmann who first thought that his H-theorem provided a microscopic derivation of the second law of thermodynamics – at least for gases. Of course this roused a veritable storm of objections by those who noted that, if the equation is taken literally to apply to every initial microscopic configuration of an isolated system, it violated the fundamental reversible nature of Newtonian mechanics (and for that matter of quantum mechanics). Furthermore, if assumed to hold for all times, it also violates, for finite systems, the Poincaré recurrence theorem. This led Boltzmann to clarify the meaning of his equation, as well as that of the second law, so that they are not only compatible with reversible microscopic laws but in fact can be expected to follow from them for “typical”, macroscopic initial states. This is however not the aspect of the BE on which Cédric has worked so far. What he has done is followed in the traditions of Hilbert, Chapman-Enskog, Grad, Cercignani, DiPerna-Lions, and others and looked for solutions of the equation. He has significantly advanced this line of research by proving theorems giving precise quantitative estimates on solutions, particularly concerning the rate of approach to equilibrium: a difficult problem, especially when one deals with spatially inhomogeneous systems.

Rather, than trying to give a summary of Cédric’s results which would go beyond my competence and time limits, I will quote from his own description of his work: My research activity lies between analysis, probability theory, statistical physics and more recently differential geometry. Here are some results:

- the first explicit estimates for convergence to equilibrium for very smooth solutions of the Boltzmann equation, without any a priori assumption of smallness of linearization.
- the discovery of tight links between certain concentration inequalities due to Talagrand, logarithmic Sobolev inequalities and the study of diffusion equations (in collaboration with Felix Otto).
- a synthetic definition of Ricci curvature lower bounds in measured metric spaces, the proof of stability of this definition, and its use to generalize to this setting various theorems of Riemannian geometry, thereby solving some open problems formulated by Gromov (works in collaboration with John Lott; closely related results have been independently obtained by Karl-Theodor Sturm).

- the discovery of a new geometric property of the tangent cut locus in Riemannian geometry: if the Ma-Trudinger-Wang curvature tensor is strictly positive, and in the absence of focalization, this cut locus is the boundary of a convex region. An outcome of this work was the proof that an almost round metric on the n -dimensional sphere has uniformly convex injectivity domains. This is part of a series of works by Alessio Figalli, Gregoire Loeper, Ludovic Rifford and myself.
- the first mathematical treatment of Landau damping in the nonlinear regime (with Clement Mouhot).

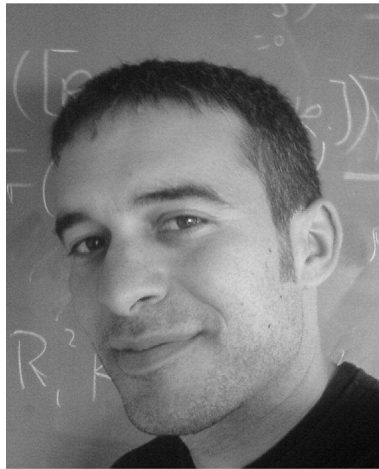
Let me conclude by noting the feat of Cédric's writing several books. One of them 1,000 pages long.

- A Review of Mathematical Topics in Collisional Kinetic Theory (230 pages)
- Topics in Optimal Transportation (360 pages)
- Optimal transport, old and new (1,000 pages)

This is amazing for a very young, very active scientist.
 Congratulations Cédric and Bon Chance.

IAMP EARLY CAREER AWARD

The prize was established by the IAMP in 2006 in recognition of a single achievement in Mathematical Physics. The prize is reserved for scientists whose age in years since birth on July 31 of the year of the Congress is less than 35, and it is awarded for the first time.



MIHALIS DAFERMOS

Citation: For his important contributions to the understanding of the strong cosmic censorship conjecture and Price law in general relativity.

OTHER PRIZES AWARDED AT THE CONGRESS

IUPAP YOUNG SCIENTIST PRIZE

The Young Scientist Prizes of the International Union of Pure and Applied Physics were established in 2007. The recipients of the awards in a given year should have a maximum of 8 years of research experience following their PhD, and should have performed original work of outstanding scientific quality. The IUPAP Commission C18 for Mathematical Physics awards its Young Scientist Prizes for the first time.



RUPERT FRANK

Citation: For outstanding results in analysis with application to quantum systems including solutions to some longstanding problems.



BENJAMIN SCHLEIN

Citation: For outstanding results in mathematical analysis of many-body quantum systems, in particular, Bose gases.



SIMONE WARZEL

Citation: For outstanding results in analysis of problems motivated by condensed matter physics, in particular, spectral and dynamical properties of random Schrödinger operators.

SPRINGER BEST POSTER PRIZE

- C. Cacciapuoti: THE MOTT'S CONJECTURE (with R. Carlone, G.F. Dell'Antonio, D. Finco, R. Figari, A. Teta)
- R. Nichols: THE SPECTRAL MINIMUM OF THE DISCRETE DISPLACEMENT MODEL (with G. Stolz)
- P.L. Ribeiro: LOCAL OFF-SHELL INTERTWINERS FOR CLASSICAL FIELD THEORIES (with R. Brunetti, K. Fredenhagen)

ANNALES HENRI POINCARÉ PRIZES

THE AHP PRIZE

- F. Vignes-Tourneret: RENORMALIZATION OF THE ORIENTABLE NON-COMMUTATIVE GROSS-NEVEU MODEL

THE AHP DISTINGUISHED PAPER AWARD

- N.Anantharaman and S.Zelditch: PATTERSON-SULLIVAN DISTRIBUTIONS AND QUANTUM ERGODICITY
- L. Erdős, M.Salmhofer and H.T.Yau: QUANTUM DIFFUSION FOR THE ANDERSON MODEL IN THE SCALING LIMIT
- M.Merkli, M.Mück and I.M.Sigal: THEORY OF NON-EQUILIBRIUM STATIONARY STATES AS A THEORY OF RESONANCES

PART A

Plenary Talks

This page intentionally left blank

QUANTUM GEOMETRY OF 3-DIMENSIONAL LATTICES AND TETRAHEDRON EQUATION

VLADIMIR V. BAZHANOV* and VLADIMIR V. MANGAZEEV

*Department of Theoretical Physics,
Research School of Physical Sciences and Engineering,
Australian National University, Canberra, ACT 0200, Australia
E-mail: Vladimir.Bazhanov@anu.edu.au

SERGEY M. SERGEEV

*Faculty of Information Sciences and Engineering,
University of Canberra, Bruce ACT 2601, Australia
E-mail: Sergey.Sergeev@canberra.edu.au*

We study geometric consistency relations between angles of 3-dimensional (3D) circular quadrilateral lattices — lattices whose faces are planar quadrilaterals inscribable into a circle. We show that these relations generate canonical transformations of a remarkable “ultra-local” Poisson bracket algebra defined on discrete 2D surfaces consisting of circular quadrilaterals. Quantization of this structure allowed us to obtain new solutions of the tetrahedron equation (the 3D analog of the Yang-Baxter equation) as well as reproduce all those that were previously known. These solutions generate an infinite number of non-trivial solutions of the Yang-Baxter equation and also define integrable 3D models of statistical mechanics and quantum field theory. The latter can be thought of as describing quantum fluctuations of lattice geometry.

Keywords: Quantum geometry; discrete differential geometry; integrable quantum systems; Yang-Baxter equation; tetrahedron equation; quadrilateral and circular 3D lattices.

1. Introduction

Quantum integrability is traditionally understood as a *purely algebraic* phenomenon. It stems from the Yang-Baxter equation [1, 2] and other algebraic structures such as the affine quantum groups [3, 4] (also called the quantized Kac-Moody algebras), the Virasoro algebra [5] and their representation theory. It is, therefore, quite interesting to learn that these algebraic structures also have remarkable geometric origins [6], which will be reviewed here.

Our approach [6] is based on connections between integrable *three-dimensional* (3D) quantum systems and integrable models of 3D discrete differential geometry. The analog of the Yang-Baxter equation for integrable quantum systems in 3D is called the *tetrahedron equation*. It was introduced by Zamolodchikov in [7, 8] (see also [9–16] for further important results in this field). Similarly to the Yang-Baxter equation the tetrahedron equation provides local integrability conditions which are

not related to the size of the lattice. Therefore the same solution of the tetrahedron equation defines different integrable models on lattices of different size, e.g., for finite periodic cubic lattices. Obviously, any such three-dimensional model can be viewed as a two-dimensional integrable model on a square lattice, where the additional third dimension is treated as an internal degree of freedom. Therefore every solution of the tetrahedron equation provides an infinite sequence of integrable 2D models differing by the size of this “hidden third dimension” [9]. Then a natural question arises whether known 2D integrable models can be obtained in this way. Although a complete answer to this question is unknown, a few non-trivial examples of such correspondence have already been constructed. The first example [11] reveals the 3D structure of the generalized chiral Potts model [17, 18]. Another example [16] reveals 3D structure of all two-dimensional solvable models associated with finite-dimensional highest weight representations for quantized affine algebras $U_q(\widehat{sl}_n)$, $n = 2, 3, \dots, \infty$ (where n coincides with the size of the hidden dimension).

Here we unravel yet another remarkable property of the same solutions of the tetrahedron equation (in addition to the hidden 3D structure of the Yang-Baxter equation and quantum groups). We show that these solutions can be obtained from quantization of geometric integrability conditions for the 3D *circular lattices* — lattices whose faces are planar quadrilaterals inscribable into a circle.

The 3D circular lattices were introduced [19] as a discretization of orthogonal coordinate systems, originating from classical works of Lamé [20] and Darboux [21]. In the continuous case such coordinate systems are described by integrable partial differential equations (they are connected with the classical soliton theory [22, 23]). Likewise, the quadrilateral and circular lattices are described by integrable difference equations. The key idea of the geometric approach [19, 24–31] to integrability of discrete classical systems is to utilize various consistency conditions [32] arising from geometric relations between elements of the lattice. It is quite remarkable that these conditions ultimately reduce to certain incidence theorems of elementary geometry. For instance, the integrability conditions for the quadrilateral lattices merely reflect the fact of existence of the 4D Euclidean cube [25]. In Sect.2 we present these conditions algebraically in a standard form of the *functional tetrahedron equation* [15]. The latter serves as the classical analog of the quantum tetrahedron equation, discussed above, and provides a connecting link to integrable quantum systems.

In Sect. 3.1 we study relations between edge angles on the 3D circular quadrilateral lattices and show that these relations describe symplectic transformations of a remarkable “ultra-local” Poisson algebra on quadrilateral surfaces (see Eq.(21)). Quantization of this structure allows one to obtain *all currently known solutions* of the tetrahedron equation. They are presented in Sect. 4, namely^a,

- (I) Zamolodchikov-Bazhanov-Baxter solution [8, 11, 13, 34]

^aThe other known solutions, previously found by Hietarinta [33] and Korepanov [14], were shown to be special cases of [34].

- (II) Bazhanov-Sergeev solution [16]
- (III) Bazhanov-Mangazeev-Sergeev solution [6]

including their interaction-round-a-cube and vertex forms. Additional details on the corresponding solvable 3D models, in particular, on their quasi-classical limit and connections with geometry, can be found in [6].

2. Discrete differential geometry: “Existence as integrability”

In this section we consider classical discrete integrable systems associated with the quadrilateral lattices. There are several ways to extract algebraic integrable systems from the geometry of these lattices. One approach, developed in [25, 27, 35–37], leads to discrete analogs of the Kadomtsev-Petviashvili integrable hierarchy. Here we present a different approach exploiting the angle geometry of the 3D quadrilateral lattices.

2.1. Quadrilateral lattices

Consider three-dimensional lattices, obtained by embeddings of the integer cubic lattice $\mathbb{Z}^3$ into the N -dimensional Euclidean space $\mathbb{R}^N$, with $N \geq 3$. Let $\mathbf{x}(m) \in \mathbb{R}^N$, denote coordinates of the lattice vertices, labeled by the 3-dimensional integer vector $m = m_1 e_1 + m_2 e_2 + m_3 e_3 \in \mathbb{Z}^3$, where $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$. Further, for any given lattice vertex $\mathbf{x}_0 = \mathbf{x}(m)$, the symbols $\mathbf{x}_i = \mathbf{x}(m + e_i)$, $\mathbf{x}_{ij} = \mathbf{x}(m + e_i + e_j)$, etc., will denote neighboring lattice vertices. The lattice is called *quadrilateral* if all its faces $(\mathbf{x}_0, \mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_{ij})$ are planar quadrilaterals. The existence of these lattices is based on the following elementary geometry fact (see Fig. 1) [25],

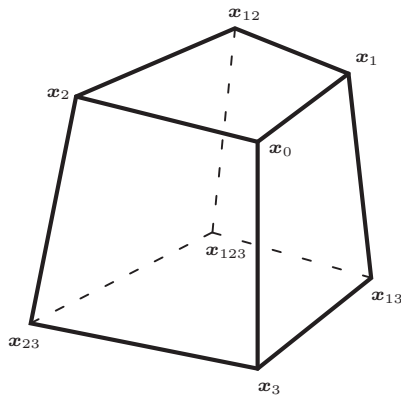


Fig. 1. An elementary hexahedron of a cubic quadrilateral lattice.

Consider four points $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ in general position in $\mathbb{R}^N$, $N \geq 3$. On each of the three planes $(\mathbf{x}_0, \mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i < j \leq 3$ choose an extra point

$\mathbf{x}_{ij}$ not lying on the lines $(\mathbf{x}_0, \mathbf{x}_i)$, $(\mathbf{x}_0, \mathbf{x}_j)$ and $(\mathbf{x}_i, \mathbf{x}_j)$. Then there exist a unique point $\mathbf{x}_{123}$ which simultaneously belongs to the three planes $(\mathbf{x}_1, \mathbf{x}_{12}, \mathbf{x}_{13})$, $(\mathbf{x}_2, \mathbf{x}_{12}, \mathbf{x}_{23})$ and $(\mathbf{x}_3, \mathbf{x}_{13}, \mathbf{x}_{23})$.

The six planes, referred to above, obviously lie in the same 3D subspace of the target space. They define a hexahedron with quadrilateral faces, shown in Fig. 1. It has the topology of the cube, so we will call it “cube”, for brevity. Let us study elementary geometry relations among the angles of this cube. Denote the angles

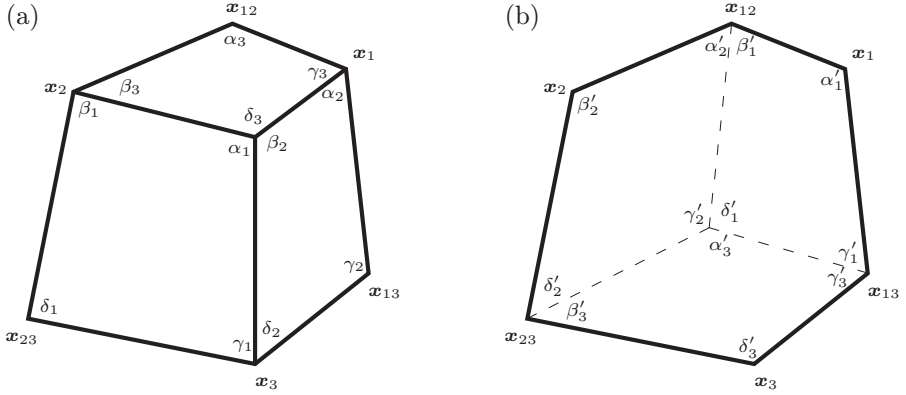


Fig. 2. The “front” (a) and “back” (b) faces of the cube in Fig. 1 and their angles.

between the edges as in Fig. 2. Altogether we have $6 \times 4 = 24$ angles, connected by six linear relations

$$\alpha_j + \beta_j + \gamma_j + \delta_j = 2\pi, \quad \alpha'_j + \beta'_j + \gamma'_j + \delta'_j = 2\pi, \quad j = 1, 2, 3, \quad (1)$$

which can be immediately solved for all “ δ ’s”. This leaves 18 angles, but only nine of them are independent. Indeed, a mutual arrangement (up to an overall rotation) of unit normal vectors to six planes in the 3D-space is determined by nine angles only. Once this arrangement is fixed all other angles can be calculated. Thus the nine independent angles of the three “front” faces of the cube, shown in Fig.2a, completely determine the angles on the three “back” faces, shown in Fig.2b, and vice versa. So the geometry of our cube provides an invertible map for three triples of independent variables

$$\mathcal{R}_{123} : \quad \{\alpha_j, \beta_j, \gamma_j\} \rightarrow \{\alpha'_j, \beta'_j, \gamma'_j\}, \quad j = 1, 2, 3. \quad (2)$$

Suppose now that all angles are known. To completely define the cube one also needs to specify lengths of its three edges. All the remaining edges can be then determined from simple linear relations. Indeed, the four sides of every quadrilateral are constrained by two relations, which can be conveniently presented in the matrix

form

$$\begin{pmatrix} \ell'_p \\ \ell'_q \end{pmatrix} = X \begin{pmatrix} \ell_p \\ \ell_q \end{pmatrix}, \quad X = \begin{pmatrix} A(\mathcal{A}) & B(\mathcal{A}) \\ C(\mathcal{A}) & D(\mathcal{A}) \end{pmatrix} = \begin{pmatrix} \frac{\sin \gamma}{\sin \delta} & \frac{\sin(\delta+\beta)}{\sin \delta} \\ \frac{\sin(\delta+\gamma)}{\sin \delta} & \frac{\sin \beta}{\sin \delta} \end{pmatrix} \quad (3)$$

where $\mathcal{A} = \{\alpha, \beta, \gamma, \delta\}$ denotes the set of angles and $\ell_p, \ell_q, \ell_{p'}, \ell_{q'}$ denote the edge lengths, arranged as in Fig.3. Note that due to (1) the entries of the two by two matrix in (3) satisfy the relation

$$AD - BC = (AB - CD)/(DB - AC). \quad (4)$$

Assume that the lengths ℓ_p, ℓ_q, ℓ_r , on one side of the two pictures in Fig.4 are given. Let us find the other three lengths $\ell_{p'}, \ell_{q'}, \ell_{r'}$ on their opposite side, by iterating the relation (3). Obviously, this can be done in two different ways: either using the front three faces, or the back ones — the results must be the same. This is exactly where the geometry gets into play. The results must be consistent due to the very existence of the cube in Fig. 1 as a geometric body. However, they will

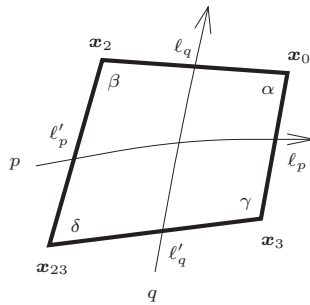


Fig. 3. The angles $\mathcal{A} = \{\alpha, \beta, \gamma, \delta\}$ and sides $\ell_p, \ell_q, \ell_{p'}, \ell_{q'}$ of a quadrilateral and the oriented rapidity lines.

be consistent only if all geometric relations between the two sets of angles in the front and back faces of the cube are taken into account. To write these relations in a convenient form we need to introduce additional notations. Note, that Fig.3 shows two thin lines, labeled by the symbols “ p ” and “ q ”. Each line crosses a pairs of opposite edges, which we call “corresponding” (in the sense that they correspond to the same thin line). Eq.(3) relates the lengths (ℓ_p, ℓ_q) of two adjacent edges with the corresponding lengths (ℓ'_p, ℓ'_q) on the opposite side of the quadrilateral.

Consider now Fig.4a which contains three directed thin lines connecting corresponding edges of the three quadrilateral faces. By the analogy with the 2D Yang-Baxter equation, where similar arrangements occur, we call them “rapidity” lines^b. We will now apply (3) three times starting from the top face and moving against

^bHowever, at the moment we do not assume any further meaning for these lines apart from using them as a convenient way of labeling to the corresponding (opposite) edges of quadrilaterals.

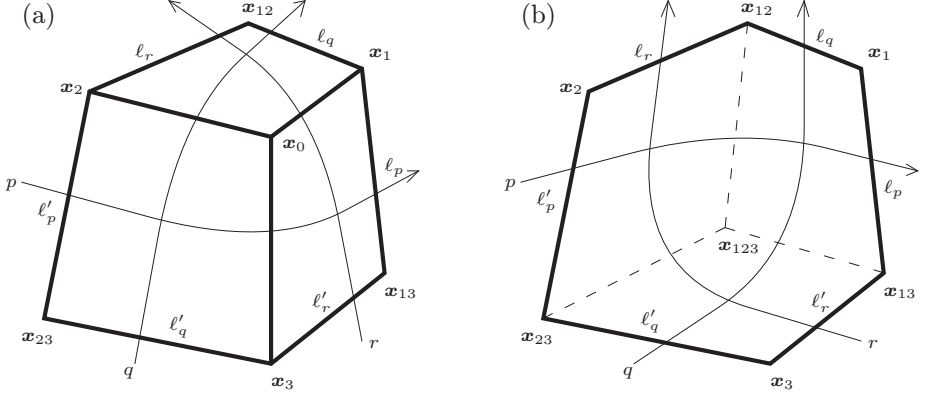


Fig. 4. The “front” (a) and “back” (b) faces of the cube in Fig. 1 and “rapidity” lines.

the directions of the arrows. Introduce the following three by three matrices

$$X_{pq}(\mathcal{A}) = \begin{pmatrix} A & B & 0 \\ C & D & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad X_{pr}(\mathcal{A}) = \begin{pmatrix} A & 0 & B \\ 0 & 1 & 0 \\ C & 0 & D \end{pmatrix}, \quad X_{qr}(\mathcal{A}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & A & B \\ 0 & C & D \end{pmatrix}, \quad (5)$$

where A, B, C, D are defined in (3) and their dependence on the angles $\mathcal{A} = \{\alpha, \beta, \gamma, \delta\}$ is implicitly understood. It follows that

$$(\ell'_p, \ell'_q, \ell'_r)^t = X_{pq}(\mathcal{A}_1) X_{pr}(\mathcal{A}_2) X_{qr}(\mathcal{A}_3) (\ell_p, \ell_q, \ell_r)^t \quad (6)$$

where

$$\mathcal{A}_j = \{\alpha_j, \beta_j, \gamma_j, \delta_j\}, \quad j = 1, 2, 3, \quad (7)$$

the lengths $\ell_p, \ell_q, \dots$ are defined as in Fig.4, and the superscript “ t ” denotes the matrix transposition. Performing similar calculations for the back faces in Fig.4b and equating the resulting three by three matrices, one obtains

$$X_{pq}(\mathcal{A}_1) X_{pr}(\mathcal{A}_2) X_{qr}(\mathcal{A}_3) = X_{qr}(\mathcal{A}'_3) X_{pr}(\mathcal{A}'_2) X_{pq}(\mathcal{A}'_1) \quad (8)$$

where

$$\mathcal{A}'_j = \{\alpha'_j, \beta'_j, \gamma'_j, \delta'_j\}, \quad j = 1, 2, 3. \quad (9)$$

This matrix relation contains exactly nine scalar equations where the LHS only depends on the front angles (7), while the RHS only depends on the back angles (9). Solving these equations one can obtain explicit form of the map (2). The resulting expressions are rather complicated and not particularly useful. However the mere fact that the map (2) satisfy a very special Eq.(8) is extremely important. Indeed, rewrite this equation as

$$X_{pq}(\mathcal{A}_1) X_{pr}(\mathcal{A}_2) X_{qr}(\mathcal{A}_3) = \mathcal{R}_{123} \left(X_{qr}(\mathcal{A}_3) X_{pr}(\mathcal{A}_2) X_{pq}(\mathcal{A}_1) \right) \quad (10)$$

where $\mathcal{R}_{123}$ is an operator acting as the substitution (2) for any function $F(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)$ of the angles,

$$\mathcal{R}_{123}\left(F(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)\right) = F(\mathcal{A}'_1, \mathcal{A}'_2, \mathcal{A}'_3) \quad (11)$$

Then, following the arguments of [14], one can show that the map (2) satisfies the *functional tetrahedron equation* [15]

$$\mathcal{R}_{123} \cdot \mathcal{R}_{145} \cdot \mathcal{R}_{246} \cdot \mathcal{R}_{356} = \mathcal{R}_{356} \cdot \mathcal{R}_{246} \cdot \mathcal{R}_{145} \cdot \mathcal{R}_{123} , \quad (12)$$

where both sides are compositions of the maps (2), involving six different sets of angles. Algebraically, this equation arises as an associativity condition for the cubic algebra (10). To discuss its geometric meaning we need to introduce *discrete evolution systems* associated with the map (2).

2.2. Discrete evolution systems: “Existence as integrability”

Consider a sub-lattice L of the 3D quadrilateral lattice, which only includes points $\mathbf{x}(m)$ with $m_1, m_2, m_3 \geq 0$. The boundary of this sub-lattice is a 2D discrete surface formed by quadrilaterals with the vertices $\mathbf{x}(m)$ having at least one of their integer coordinates m_1, m_2, m_3 equal to zero and the other two non-negative. Assume that all quadrilateral angles on this surface are known, and consider them as initial data. Then repeatedly applying the map (2) one can calculate angles on all faces of the sub-lattice L , defined above (one has to start from the corner $\mathbf{x}(0)$). The process can be visualized as an evolution of the initial data surface where every transformation (2) corresponds to a “flip” between the front and back faces (Fig. 2) of some cube adjacent to the surface. This makes the surface looking as a 3D “staircase” (or a pile of cubes) in the intersection corner of the three coordinate planes, see Fig. 5 showing two stages of this process. Note, that the corresponding evolution equations

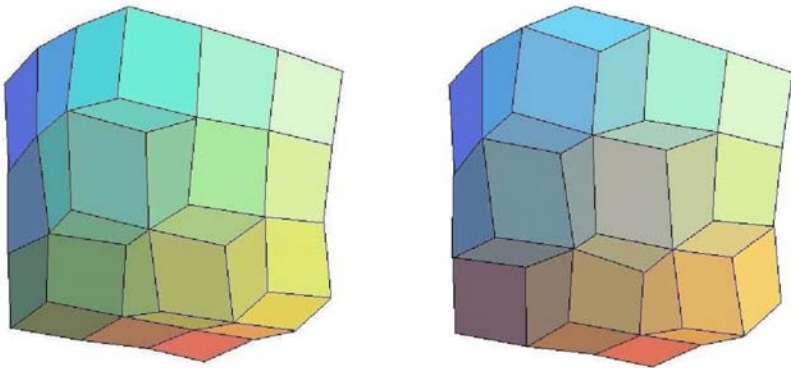


Fig. 5. Visualization of the 3D “staircase” evolution.

can be written in a covariant form for an arbitrary lattice cube (see Eq.(23) below

for an example). It is also useful to have in mind that the above evolution can be defined purely geometrically as a *ruler-and-compass* type construction. Indeed the construction of the point $\mathbf{x}_{123}$ in Fig. 1 from the points $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_{12}, \mathbf{x}_{13}, \mathbf{x}_{23}$ (and that is what is necessary for flipping a cube) only requires a *2D-ruler* which allows to draw planes through any three non-collinear points in the Euclidean space.

Similar evolution systems can be defined for other quadrilateral lattices instead of the 3D cubic lattice considered above. Since the evolution is local (only one cube is flipped at a time) one could consider finite lattices as well. For example, consider six adjacent quadrilateral faces covering the front surface of the rhombic dodecahedron^c shown in Fig. 6. Suppose that all angles on these faces are given and consider them as initial data. Now apply a sequence of four maps (2) and calculate angles on the back surface of the rhombic dodecahedron. This can be done in two alternative ways, corresponding to the two different dissection of the rhombic dodecahedron into four cubes shown in Fig.6. The functional tetrahedron equation

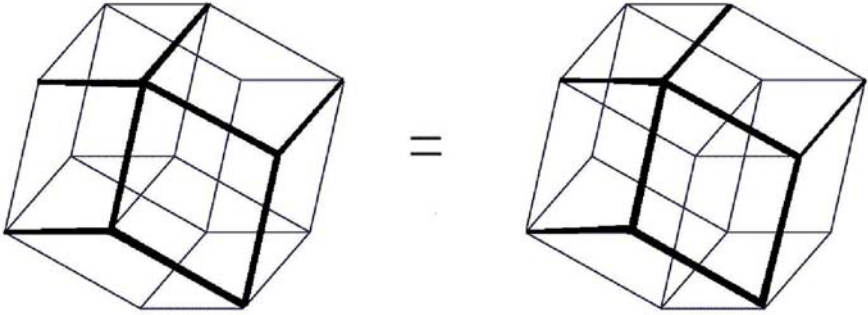


Fig. 6. Two dissections of the rhombic dodecahedron into four quadrilateral hexahedra.

(12) states that the results will be the same. Thereby it gives an algebraic proof for the equivalence of two “ruler-and-compass” type constructions of the back surface of the dodecahedron in Fig. 6. Can we also prove this equivalence geometrically? Although from the first sight this does not look trivial, it could be easily done from the point of view of the 4D geometry. The required statement follows just from the fact of existence of the quadrilateral lattice with the topology of the 4D cube [25]. The latter is defined by eight intersecting 3-planes in a general position in the 4-space. The two rhombic dodecahedra shown in Fig. 6 are obtained by a dissection of the 3-surface of this 4-cube, along its 2-faces, so these dodecahedra must have exactly the same quadrilateral 2-surface. Thus the functional tetrahedron equation (12), which plays the role of integrability condition for the discrete evolution system

^cIt is worth noting that the most general rhombic dodecahedron with quadrilateral faces can only be embedded into (at least) the 4D Euclidean space.

associated with the map (2), simply follows from the mere fact of existence of the 4-cube, which is the simplest 4D quadrilateral lattice. For a further discussion of a relationship between the geometric consistency and integrability see [32].

3. Quantization of the 3D circular lattices

3.1. *Poisson structure of circular lattices*

The 3D *circular lattice* [19, 28, 29] is a special 3D quadrilateral lattice where all faces are circular quadrilaterals (i.e., quadrilaterals which can be inscribed into a circle). The existence of these lattices is established by the following beautiful geometry theorem due to Miquel [38] (see Fig. 7)

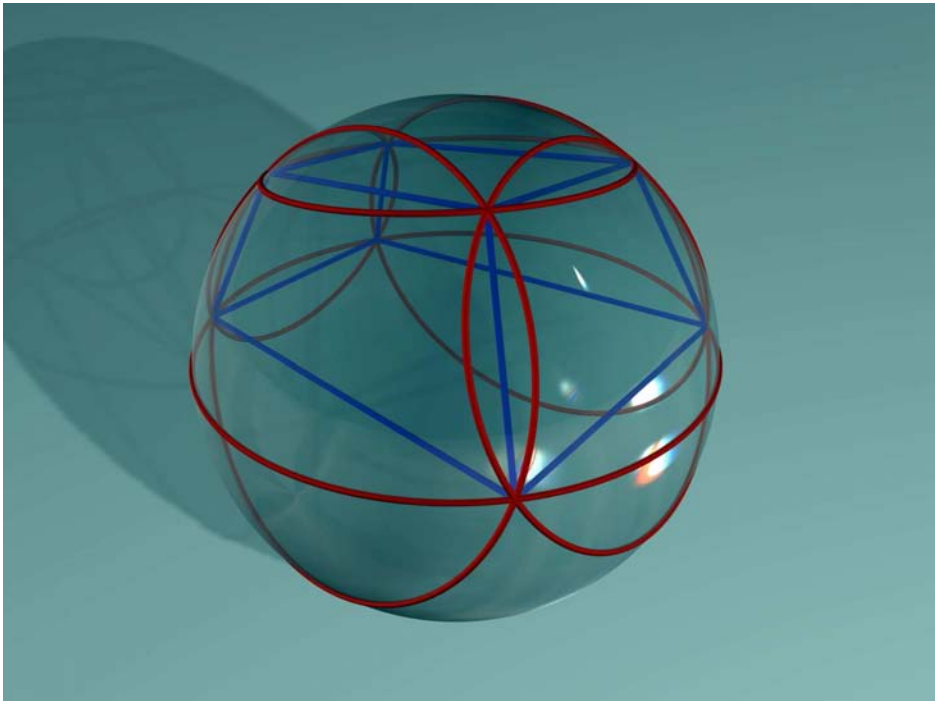


Fig. 7. Miquel configuration of circles in 3D space, an elementary hexahedron and its circumsphere.

Miquel theorem. Consider four points $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ in general position in $\mathbb{R}^N$, $N \geq 3$. On each of the three circles $c(\mathbf{x}_0, \mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i < j \leq 3$ choose an additional new point $\mathbf{x}_{ij}$. Then there exist a unique point $\mathbf{x}_{123}$ which simultaneously belongs to the three circles $c(\mathbf{x}_1, \mathbf{x}_{12}, \mathbf{x}_{13})$, $c(\mathbf{x}_2, \mathbf{x}_{12}, \mathbf{x}_{23})$ and $c(\mathbf{x}_3, \mathbf{x}_{13}, \mathbf{x}_{23})$.

It is easy to see that the above six circles lie on the same sphere. It follows then that every elementary “cube” on a circular lattice (whose vertices are at the circle intersection points) is inscribable into a sphere, see Fig. 7. The general formulae of the previous subsection can be readily specialized for the circular lattices. A circular quadrilateral has only two independent angles. In the notation of Fig. 3 one has

$$\gamma = \pi - \beta, \quad \delta = \pi - \alpha. \quad (13)$$

Due to the Miquel theorem we can simply impose these restrictions on all faces of the lattice without running to any contradictions. The two by two matrix in (3) takes the form

$$X = \begin{pmatrix} k & a^* \\ -a & k \end{pmatrix}, \quad \det X = 1, \quad (14)$$

where we have introduced new variables

$$k = (\csc \alpha) \sin \beta, \quad a = (\csc \alpha) \sin(\alpha + \beta), \quad a^* = (\csc \alpha) \sin(\alpha - \beta), \quad (15)$$

instead of the two angles $\{\alpha, \beta\}$. Note that the new variables are constrained by the relation

$$aa^* = 1 - k^2. \quad (16)$$

Conversely, one has

$$\cos \alpha = \frac{a - a^*}{2k}, \quad \cos \beta = \frac{a + a^*}{2}. \quad (17)$$

Let the variables $\{k_j, a_j, a_j^*\}$, $\{k'_j, a'_j, a'^*_j\}$, $j = 1, 2, 3$, correspond to the front and back faces of the cube. The map (2) then read explicitly

$$\mathcal{R}_{123} : \begin{cases} (k_2 a_1^*)' = k_3 a_1^* - \varepsilon k_1 a_2^* a_3, & (k_2 a_1)' = k_3 a_1 - \varepsilon k_1 a_2 a_3^*, \\ (a_2^*)' = a_1^* a_3^* + \varepsilon k_1 k_3 a_2^*, & (a_2)' = a_1 a_3 + \varepsilon k_1 k_3 a_2, \\ (k_2 a_3^*)' = k_1 a_3^* - \varepsilon k_3 a_1 a_2^*, & (k_2 a_3)' = k_1 a_3 - \varepsilon k_3 a_1^* a_2, \end{cases} \quad (18)$$

where $\varepsilon = +1$ and

$$k'_2 = \sqrt{1 - a'_2 a_2^*}. \quad (19)$$

At this point we note that exactly the same map together with the corresponding equations (8) and (12) were previously obtained in [16]. Moreover, it was discovered that this map is a canonical transformation preserving the Poisson algebra

$$\{a_i, a_j^*\} = 2 \delta_{ij} k_i^2, \quad \{k_i, a_j\} = \delta_{ij} k_i a_i, \quad \{k_i, a_j^*\} = -\delta_{ij} k_i a_i^*, \quad i, j = 1, 2, 3, \quad (20)$$

where $k_i^2 = 1 - a_i a_i^*$. Note that variables k, a, a^* on different quadrilaterals are in involution. The same Poisson algebra in terms of angle variables reads

$$\{\alpha_i, \beta_j\} = \delta_{ij}, \quad \{\alpha_i, \alpha_j\} = \{\beta_i, \beta_j\} = 0. \quad (21)$$

This “ultra-local” symplectic structure trivially extends to any circular quad-surface of initial data, discussed above. To resolve an apparent ambiguity in naming of the angles, this surface must be equipped with oriented rapidity lines, similar to those in Fig. 4^d. In addition, the angles for each quadrilateral should be arranged as in Fig. 3. Then one can assume that the indices i, j in (21) refer to all quadrilaterals on this surface.

Thus, the evolution defined by the map (18) is a symplectic transformation. The corresponding equations of motion for the whole lattice (the analog of the Hamilton-Jacobi equations) can be written in a “covariant” form. For every cube define

$$A_{32} = a_1, \quad A_{23} = a_1^*, \quad A_{31} = a_2', \quad A_{13} = a_2'^*, \quad A_{21} = a_3, \quad A_{12} = a_3^* \quad (22)$$

where A_{jk} stands for $A_{jk}(m)$, where m is such that $\mathbf{x}(m)$ coincides with the coordinates of the top front corner of the cube (vertex $\mathbf{x}_0$ in Fig.1). Let T_k be the shift operator $T_k A_{ij}(m) = A_{ij}(m + e_k)$. Then

$$\tilde{T}_k A_{ij} = \frac{A_{ij} - A_{ik} A_{kj}}{K_{ik} K_{kj}}, \quad K_{ij} = K_{ji} = \sqrt{1 - A_{ij} A_{ji}}, \quad (23)$$

where (i, j, k) is an arbitrary permutation of $(1, 2, 3)$ and

$$\tilde{T}_1 = T_1, \quad \tilde{T}_2 = T_2^{-1}, \quad \tilde{T}_3 = T_3. \quad (24)$$

Note that Eq.(23) also imply

$$(\tilde{T}_k K_{ij}) K_{kj} = (\tilde{T}_i K_{kj}) K_{ij}. \quad (25)$$

Remarks. The equations (23) have been previously obtained in [29], see Eq.(7.20) therein. The quantities A_{ij} in (23) should be identified with the *rotation coefficients* denoted as $\tilde{\beta}_{ij}$ in [29]. The same equations (23) are discussed in §3.1 of [32], where one can also find a detailed bibliography on the circular lattices (we are indebted to A.I.Bobenko for these important remarks).

3.2. Quantization and the tetrahedron equation

In the next section we consider different quantizations of the map (18) and obtain several solutions of the full quantum tetrahedron equation (see Eq.(33) below). In all cases we start with the canonical quantization of the Poisson algebra (21),

$$[\alpha_i, \beta_j] = \xi \hbar \delta_{ij}, \quad [\alpha_i, \alpha_j] = 0, \quad [\beta_i, \beta_j] = 0, \quad (26)$$

where $\hbar$ is the quantum parameter (the Planck constant) and ξ is a numerical coefficient, introduced for a further convenience. The indices i, j label the faces of the “surface of initial data” discussed above. Since the commutation relations (26)

^dWe refer the reader to our previous paper [39] where the relationship between the rapidity graphs and quadrilateral lattices is thoroughly discussed.

are ultra-local (in the sense that the angle variables on different faces commute with each other), let us concentrate on the local Heisenberg algebra,

$$H : \quad [\alpha, \beta] = \xi \hbar , \quad (27)$$

for a single lattice face (remind that the angles shown in Fig. 3 are related by (13)). The map (18) contains the quantities k, a, a^* , defined in (15), which now become operators. For definiteness, assume that the non-commuting factors in (15) are ordered exactly as written. Then the definitions (15) give

$$\begin{aligned} k &= (U - U^{-1})^{-1} (V - V^{-1}), \\ a &= q^{-\frac{1}{2}} (U - U^{-1})^{-1} (U V - U^{-1} V^{-1}), \\ a^* &= q^{+\frac{1}{2}} (U - U^{-1})^{-1} (U V^{-1} - U^{-1} V), \end{aligned} \quad (28)$$

where the elements U and V generate the Weyl algebra,

$$U V = q V U, \quad U = e^{i\alpha}, \quad V = e^{i\beta}, \quad q = e^{-\xi \hbar} . \quad (29)$$

The operators (28) obey the commutation relations of the q -oscillator algebra,

$$\text{Osc}_q : \quad \begin{cases} q a^* a - q^{-1} a a^* = q - q^{-1}, & k a^* = q a^* k, & k a = q^{-1} a k, \\ k^2 = q (1 - a^* a) = q^{-1} (1 - a a^*) \end{cases} \quad (30)$$

where the element k is assumed to be invertible. This algebra is, obviously, a quantum counterpart of the Poisson algebra (20). In the previous Section we have already mentioned the result of [16] that

- (i) the map (18) is an automorphism of the tensor cube of the Poisson algebra (20) (remind that the relation (16) should be taken into account in (18)).

In the same paper [16] it was also shown that

- (ii) there exists a quantum version of the map (18), which acts as an automorphism of the tensor cube of the q -oscillator algebra (30). The formulae (18) for the quantum map stay exactly the same, but the relation (16) should be replaced by either of the two relations on the second line of (30), for instance, $k^2 = q (1 - a^* a)$. In particular, (19) should be replaced with

$$(k'_2)^2 = q (1 - a^{*'}_2 a'_2) . \quad (31)$$

- (iii) the quantum version of the map (18), defined in (ii) above, satisfies the functional tetrahedron equation (12).

For any irreducible representations of the q -oscillator algebra (30) the formulae (18) and (31) uniquely determine the $\mathcal{R}_{123}$ as an internal automorphism,

$$\mathcal{R}_{123}(F) = R_{123} F R_{123}^{-1}, \quad F \in \text{Osc}_q \otimes \text{Osc}_q \otimes \text{Osc}_q . \quad (32)$$

It follows then from (12) that the linear operator R_{123} satisfies the quantum tetrahedron equation

$$R_{123} R_{145} R_{246} R_{356} = R_{356} R_{246} R_{145} R_{123} , \quad (33)$$

where each of the operators R_{123} , R_{145} , R_{246} and R_{356} act as (32) in the three factors (indicated by the subscripts) of a tensor product of six q -oscillator algebras and act as the unit operator in the remaining three factors.

4. Solutions of the tetrahedron equation

Here we show that all known solutions of the tetrahedron equation can be obtained by solving (32) for the linear operator R_{123} with different irreducible representations of the q -oscillator algebra (30).

4.1. Fock representation solution

In this subsection we set $\xi = 1$ in (26) and $q = e^{-\hbar}$. Define the Fock representation of a single q -oscillator algebra (30),

$$a|0\rangle = 0, \quad a|n+1\rangle = |n\rangle, \quad a^*|n\rangle = (1 - q^{2+2n})|n+1\rangle, \quad k|n\rangle = q^{n+1/2}|n\rangle, \quad (34)$$

spanned by the vectors $|n\rangle$, $n = 0, 1, \dots, \infty$. Then using (18), (30), (31) and (32) one can show that

$$\begin{aligned} \langle n_1, n_2, n_3 | R | n'_1, n'_2, n_3 \rangle &= \delta_{n_1+n_2, n'_1+n'_2} \delta_{n_2+n_3, n'_2+n'_3} (-1)^{n_2} q^{(n'_1-n_2)(n'_3-n_2)} \\ &\times \binom{n_3}{n'_2}_{q^2} {}_2\phi_1(q^{-2n'_2}, q^{2(1+n'_3)}, q^{2(1-n'_2+n_3)}; q^2, q^{2(1+n_1)}) , \end{aligned} \quad (35)$$

where

$$(x; q^2)_n = \prod_{j=0}^{n-1} (1 - q^{2j}x), \quad \binom{n}{k}_{q^2} = \frac{(q^2; q^2)_n}{(q^2; q^2)_k (q^2; q^2)_{n-k}}, \quad (36)$$

and

$${}_2\phi_1(a, b, c; q^2, z) = \sum_{n=0}^{\infty} \frac{(a; q^2)_n (b; q^2)_n}{(q^2; q^2)_n (c; q^2)_n} z^n \quad (37)$$

is the q -deformed Gauss hypergeometric series. This 3D R -matrix satisfies the constant tetrahedron equation (33). In matrix form this equation reads

$$\begin{aligned} \sum_{n'_j=0}^{\infty} R_{n_1, n_2, n_3}^{n'_1, n'_2, n'_3} R_{n'_1, n_4, n_5}^{n''_1, n'_4, n'_5} R_{n'_2, n'_4, n_6}^{n''_2, n''_4, n'_6} R_{n'_3, n'_5, n'_6}^{n''_3, n''_5, n''_6} \\ = \sum_{n'_j=0}^{\infty} R_{n_3, n_5, n_6}^{n'_3, n'_5, n'_6} R_{n_2, n_4, n_6}^{n'_2, n'_4, n''_6} R_{n_1, n'_4, n'_5}^{n'_1, n''_4, n''_5} R_{n'_1, n'_2, n'_3}^{n''_1, n''_2, n''_3} \end{aligned} \quad (38)$$

where the sum is taken over six indices $n'_1, n'_2, n'_3, n'_4, n'_5, n'_6$ and

$$R_{n_1, n_2, n_3}^{n'_1, n'_2, n'_3} = \langle n_1, n_2, n_3 | R | n'_1, n'_2, n'_3 \rangle . \quad (39)$$

Note that Eq.(38) does not contain any spectral parameters. Originally the R -matrix (35) was obtained in [16] in terms of a solution of some recurrence relation, which was subsequently reduced to the q -hypergeometric function in [6].

4.2. Modular double solution

In this subsection we set in (26)

$$\xi = -i, \quad \hbar = \pi b^2, \quad q = e^{i\pi b^2} \quad (40)$$

where b is a free parameter, $\text{Re } b \neq 0$. Here it will be more convenient to work with a slightly modified version^e of the map (18), with the value $\varepsilon = -1$.

Consider a non-compact representation [40] of the q -oscillator algebra (30) in the space of functions $f(\sigma) \in \mathbb{L}^2(\mathbb{R})$ on the real line admitting an analytical continuation into an appropriate horizontal strip, containing the real axis in the complex σ -plane (see [40] for further details). Such representation essentially reduces to that of the Weyl algebra

$$W_q : \quad k w = q w k, \quad q = e^{i\pi b^2}, \quad (41)$$

realized as multiplication and shift operators

$$k |\sigma\rangle = i e^{\pi \sigma b} |\sigma\rangle, \quad w |\sigma\rangle = |\sigma - i b\rangle. \quad (42)$$

The generators a, a^* in (30) are expressed as

$$a = w^{-1}, \quad a^* = (1 - q^{-1} k^2) w. \quad (43)$$

As explained in [41] the representation (42) is not, in general, irreducible. Therefore, the relation (32) alone does not unambiguously define the linear operator R_{123} in this case. Following the idea of [41] consider the modular dual of the algebra (41),

$$W_{\tilde{q}} : \quad \tilde{k} \tilde{w} = \tilde{q} \tilde{w} \tilde{k}, \quad \tilde{q} = e^{-i\pi b^{-2}}, \quad (44)$$

acting in the same representation space

$$\tilde{k} |\sigma\rangle = -i e^{\pi \sigma b^{-1}} |\sigma\rangle, \quad \tilde{w} |\sigma\rangle = |\sigma + i b^{-1}\rangle. \quad (45)$$

We found that if the relation (32) is complemented by its modular dual

$$\tilde{\mathcal{R}}_{123}(\tilde{F}) = R_{123} \tilde{F} R_{123}^{-1}, \quad \tilde{F} \in \text{Osc}_{\tilde{q}} \otimes \text{Osc}_{\tilde{q}} \otimes \text{Osc}_{\tilde{q}}, \quad (46)$$

then the pair of relations (32) and (46) determine the operator R_{123} uniquely^f. The dual q -oscillator algebra $\text{Osc}_{\tilde{q}}$ is realized through the dual Weyl pair (44) and the relations

$$\tilde{a} = \tilde{w}^{-1}, \quad \tilde{a}^* = (1 - \tilde{q}^{-1} \tilde{k}^2) \tilde{w}. \quad (47)$$

^eNote that Eq.(18) is a particular case a general map considered in [16]; see also Eq.(59) below.

^fIt is worth mentioning similar phenomena in the construction of the R -matrix [42] for the modular double of the quantum group $U_q(sl_2)$ and the representation theory of $U_q(sl_2, \mathbb{R})$ [43].

The dual version of the map $\tilde{\mathcal{R}}_{123}$ is defined by the same formulae (18), where quantities k_j, a_j, a_j^* , $j = 1, 2, 3$ are replaced by their “tilded” counterparts $\tilde{k}_j, \tilde{a}_j, \tilde{a}_j^*$. The value of q does not, actually, enter the map (18), but needs to be taken into account in the relations between the generators of the q -oscillator algebra. Thus, the linear operator R_{123} in this case simultaneously provides the two maps $\mathcal{R}_{123}$ and $\tilde{\mathcal{R}}_{123}$ (given by (18) with $\varepsilon = -1$). The explicit form of this operator is given below.

Denote

$$\eta = \frac{b + b^{-1}}{2}, \quad (48)$$

and define a special function

$${}_2\Psi_2 \left(\begin{matrix} c_1, c_2 \\ c_3, c_4 \end{matrix} \middle| c_0 \right) = \int_{\mathbb{R}} dz e^{2\pi iz(-c_0 - i\eta)} \frac{\varphi(z + \frac{c_1 + i\eta}{2}) \varphi(z + \frac{c_2 + i\eta}{2})}{\varphi(z + \frac{c_3 - i\eta}{2}) \varphi(z + \frac{c_4 - i\eta}{2})}, \quad (49)$$

where φ is the non-compact quantum dilogarithm [44]

$$\varphi(z) = \exp \left(\frac{1}{4} \int_{\mathbb{R} + i0} \frac{e^{-2izx}}{\sinh(xb) \sinh(x/b)} \frac{dx}{x} \right). \quad (50)$$

The values of c_1, c_2, c_3, c_4 are assumed to be such that poles of numerator in the integrand of (49) lie above the real axis, while the zeroes of the denominator lie below the real axis. For other values of c_j the integral (49) is defined by an analytic continuation. Note that for $\text{Im } b^2 > 0$ the integral ${}_2\Psi_2$ can be evaluated by closing the integration contour in the upper half plane (see eq.(78) in [6]), which is very convenient for numerical calculations.

The matrix elements of R -matrix solving the pair of the equations (32) and (46) are given by

$$\begin{aligned} \langle \sigma_1, \sigma_2, \sigma_3 | R | \sigma'_1, \sigma'_2, \sigma'_3 \rangle &= \delta_{\sigma_1 + \sigma_2, \sigma'_1 + \sigma'_2} \delta_{\sigma_2 + \sigma_3, \sigma'_2 + \sigma'_3} \\ &\times e^{i\pi(\sigma'_1 \sigma'_3 + i\eta(\sigma'_1 + \sigma'_3 - \sigma_2))} {}_2\Psi_2 \left(\begin{matrix} \sigma_1 - \sigma_3, -\sigma_1 + \sigma_3 \\ \sigma_1 + \sigma_3, -\sigma'_1 - \sigma'_3 \end{matrix} \middle| \sigma_2 \right). \end{aligned} \quad (51)$$

This R -matrix satisfies the constant tetrahedron equation (33). Its matrix form is given by (38) where $R_{n_i, n_j, n_k}^{n'_i, n'_j, n'_k}$ is substituted by $R_{\sigma_i, \sigma_j, \sigma_k}^{\sigma'_i, \sigma'_j, \sigma'_k} = \langle \sigma_i, \sigma_j, \sigma_k | R | \sigma'_i, \sigma'_j, \sigma'_k \rangle$, $i, j, k = 1, 2, \dots, 6$, and the sums are replaced by the integrals over $\sigma'_1, \sigma'_2, \sigma'_3, \sigma'_4, \sigma'_5, \sigma'_6$ along the real lines $-\infty < \sigma'_i < \infty$. One can verify that these integrals converge. The solution (51) was obtained in [6].

4.2.1. The “interaction-round-a-cube” formulation of the modular double solution

Note that due to the presence of two delta-functions in (51) the edge spins are constrained by two relations $\sigma_1 + \sigma_3 = \sigma'_1 + \sigma_3$ and $\sigma_2 + \sigma_3 = \sigma'_2 + \sigma'_3$ at each vertex of the lattice. Here we re-formulate this solution in terms of unconstrained *corner* spins, which also take continuous values on the real line. Figure 8 shows an

elementary cube of the lattice with the corner spins “ a, b, c, d, e, f, g, h ” arranged in the same way as in [10]. The corresponding Boltzmann weight reads [6]

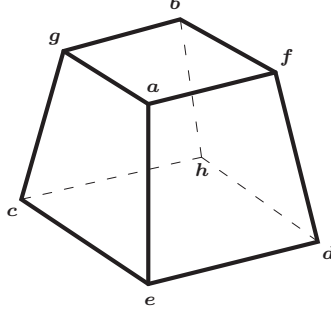


Fig. 8. The arrangement of corner spins around a cube.

$$W(a|e, f, g|b, c, d|h; \mathsf{T}_1, \mathsf{T}_2, \mathsf{T}_3) = e^{i\pi(\sigma'_1\sigma'_3 + i\eta(\sigma'_1 + \sigma'_3 - \sigma_2))} {}_2\Psi_2 \left(\begin{matrix} \sigma_1 - \sigma_3, -\sigma_1 + \sigma_3 \\ \sigma_1 + \sigma_3, -\sigma'_1 - \sigma'_3 \end{matrix} \middle| \sigma_2 \right), \quad (52)$$

where

$$\begin{aligned} \sigma_1 &= g + f - a - b - \mathsf{T}_1, \quad \sigma_2 = a + c - e - g + \mathsf{T}_2, \quad \sigma_3 = e + f - a - d - \mathsf{T}_3, \\ \sigma'_1 &= c + d - e - h - \mathsf{T}_1, \quad \sigma'_2 = f + h - b - d + \mathsf{T}_2, \quad \sigma'_3 = b + c - g - h - \mathsf{T}_3. \end{aligned} \quad (53)$$

Note, that the variables σ_i and σ'_i automatically satisfy the delta function constraints in (51). The weight functions (52) contains three arbitrary (spectral) parameters $\mathsf{T}_1, \mathsf{T}_2, \mathsf{T}_3$. It satisfies the tetrahedron equation in the “interaction-round-a-cube” form [10] (see Fig. 9),

$$\begin{aligned} & \int_{\mathbb{R}} dz W(a_4|c_1, c_3, c_2|b_3, b_2, b_1|z) W'(c_1|a_3, b_1, b_2|z, c_6, c_4|b_4) \\ & \quad \times W''(b_1|c_4, c_3, z|b_3, b_4, a_2|c_5) W'''(z|b_4, b_3, b_2|c_2, c_6, c_5|a_1) \\ & = \int_{\mathbb{R}} dz W'''(b_1|c_4, c_3, c_1, |a_4, a_3, a_2|z) W''(c_1|a_3, a_4, b_2|c_2, c_6, z|a_1) \\ & \quad \times W'(a_4|z, c_3, c_2|b_3, a_1, a_2|c_5) W(z|a_3, a_2, a_1|c_5, c_6, c_4|b_4), \end{aligned} \quad (54)$$

where the four sets of the spectral parameters are constrained as

$$\mathsf{T}'_1 = \mathsf{T}_1, \quad \mathsf{T}''_1 = -\mathsf{T}_2, \quad \mathsf{T}'''_1 = \mathsf{T}_3, \quad \mathsf{T}''_2 = \mathsf{T}'_2, \quad \mathsf{T}'''_2 = -\mathsf{T}'_3, \quad \mathsf{T}'''_3 = \mathsf{T}''_3. \quad (55)$$

Note that the parameters $\mathsf{T}_1, \mathsf{T}_2, \mathsf{T}_3$ are similar to those in the Zamolodchikov model [8] and its generalization for an arbitrary number of spin states [11], which is considered in Sect. 4.3 below. One can relate them as $\mathsf{T}_j = \log[\tan(\theta_j/2)]$ to the angles $\theta_1, \theta_2, \theta_3$ of the spherical triangle in Sect. 4.3.2. In total, there are twelve parameters $\mathsf{T}_j, \mathsf{T}'_j, \mathsf{T}''_j, \mathsf{T}'''_j$, $j = 1, 2, 3$ (three for each weight function) constrained

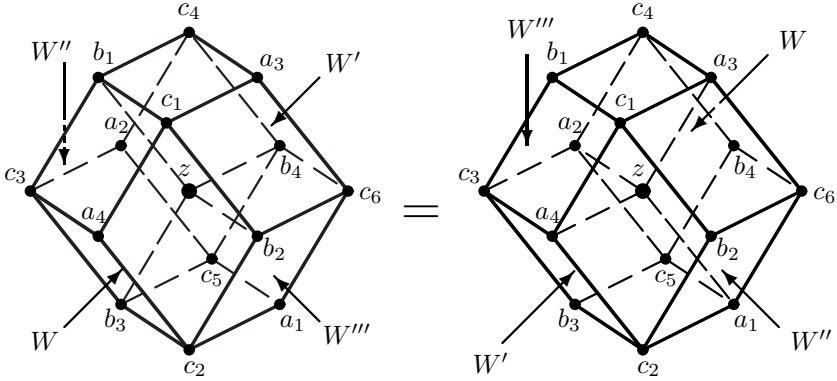


Fig. 9. Graphical representation for the tetrahedron equations for interaction-round-a-cube models.

by six relations (55). Thus, the tetrahedron equation (54) contains six independent parameters (in contrast to only five parameters in the cyclic case; see the text after Eq.(72) below).

The above solution can be simply generalized by multiplying the weight (52) by an “external field” factor

$$W(\dots) \rightarrow \exp \left[\sum_{j=1}^3 f_j(\sigma_j + \sigma'_j) \right] W(\dots) . \quad (56)$$

This substitution does not affect the validity of the tetrahedron equation (54), provided the field parameters f_j for different W ’s are constrained as

$$f_3'' = f_3' - f_3, \quad f_1''' = f_1'' - f_1', \quad f_2''' = f_2'' + f_1, \quad f_3''' = f_2 - f_2'. \quad (57)$$

Similar generalizations apply for the solutions (35) and (51). The solution (52) was previously obtained in [6].

4.3. Cyclic representation solution

4.3.1. Generalized form of the q -oscillator map

Here, we will solve equation (32) with a more general form of the map (18), considered in [16]. Let

$$\mathcal{L}(\mathcal{H}_j) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda_j \mathbf{k}_j & \mathbf{a}_j^+ & 0 \\ 0 & \lambda_j \mu_j \mathbf{a}_j^- & -\mu_j \mathbf{k}_j & 0 \\ 0 & 0 & 0 & \lambda_j \mu_j \end{pmatrix} \quad j = 1, 2, 3, \quad (58)$$

be an operator-valued matrix acting in the direct product of two vector spaces $\mathbb{C}^2 \otimes \mathbb{C}^2$, whose elements depends on a set of generators $\mathcal{H}_j = \{k_j, a_j, a_j^*\}$ of the q -oscillator algebra (30) and continuous parameters λ_j , and μ_j , where $j = 1, 2, 3$. The RHS of (58) is understood as a block matrix with two-dimensional blocks where matrix indices of the second space $\mathbb{C}^2$ numerate the blocks, while those for first space numerate the elements inside the blocks.

The required generalization of (32) reads [16]^g

$$L_{12}(\mathcal{H}_1) L_{13}(\mathcal{H}_2) L_{23}(\mathcal{H}_3) = R_{123} L_{23}(\mathcal{H}_3) L_{13}(\mathcal{H}_2) L_{12}(\mathcal{H}_1) R_{123}^{-1} \quad (59)$$

Here $\mathcal{L}_{12}$ denote the matrix which acts as (58) in the first and second component of the tensor product $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ and coincides with the unit operator in the third component ($\mathcal{L}_{13}$ and $\mathcal{L}_{23}$ are defined similarly). It is important to note that the additional parameters λ_j and μ_j enter (59) only in three combinations λ_2/λ_3 , $\lambda_1\mu_3$ and μ_1/μ_2 (explicit formulae for the corresponding generalization of the map (18) are given in Eqs.(24-26) of [16]).

4.3.2. Cyclic representations of the q -oscillator algebra

Let $N \geq 3$ be an odd integer,

$$q = -e^{i\pi/N}, \quad q^N = 1, \quad N = \text{odd}, \quad N \geq 3. \quad (60)$$

and V^N be an N -dimensional vector space, spanned by the vectors $|n\rangle$, $n \in \mathbb{Z}_N = \{0, 1, \dots, N-1\}$. Define N by N matrices (the index n is treated *modulo* N)

$$X|n\rangle = q^n |n\rangle, \quad Z|n\rangle = |n+1\rangle, \quad X^N = Z^N = 1, \quad (61)$$

and their embedding into a direct product $V^N \otimes V^N \otimes V^N$,

$$X_j = 1 \otimes \dots \otimes X_{j\text{-th}} \otimes \dots \otimes 1, \quad Z_j = 1 \otimes \dots \otimes Z_{j\text{-th}} \otimes \dots \otimes 1, \quad j = 1, 2, 3. \quad (62)$$

Equation (59) involves the direct product of three copies of the q -oscillator algebra (30), labeled by the subscript $j = 1, 2, 3$. The most general cyclic representation for this product

$$\mathbf{k}_j = \varkappa_j X_j, \quad \mathbf{a}_j^* = \frac{1}{\rho_j} (1 - q^{-1} \varkappa_j^2 X_j^2) Z_j, \quad \mathbf{a}_j = \rho_j Z_j^{-1}, \quad j = 1, 2, 3, \quad (63)$$

contains two continuous parameters $\varkappa_j, \rho_j$, for each factor, so there six arbitrary parameters altogether. Detailed inspection of (59) shows, however, that ρ_1, ρ_2, ρ_3 only enter through a ratio $\rho_1\rho_3/\rho_2$, so there are only four essential parameters.

Let $\theta_1, \theta_2, \theta_3$ and a_1, a_2, a_3 be angles and sides of a spherical triangle, respectively. Define

$$2\beta_1 = a_2 + a_3 - a_1, \quad 2\beta_2 = a_1 + a_3 - a_2, \quad 2\beta_3 = a_1 + a_2 - a_3, \quad \beta_0 = \pi - \beta_1 - \beta_2 - \beta_3. \quad (64)$$

^gEquation (59) reduces to (32) when $\lambda_1 = \lambda_2 = \lambda_3 = 1$ and $\mu_1 = \mu_2 = \mu_3 = -\varepsilon$, where ε enters the map (18).

and set

$$\varkappa_1 = i \sqrt[N]{\tan \frac{\theta_1}{2}}, \quad \varkappa_2 = i \sqrt[N]{\cot \frac{\theta_2}{2}}, \quad \varkappa_3 = i \sqrt[N]{\tan \frac{\theta_3}{2}}. \quad (65)$$

Then Eq.(59) implies

$$\rho_1 \rho_3 / \rho_2 = e^{-i\beta_2/N} \sqrt[N]{\sin a_2 / \sin \beta_0}, \quad (66)$$

and

$$\lambda_3 / \lambda_2 = e^{ia_1/N}, \quad \lambda_1 \mu_3 = e^{-ia_2/N}, \quad \mu_1 / \mu_2 = e^{ia_3/N}. \quad (67)$$

4.3.3. Solution of the tetrahedron equation

Let p denote a point on Fermat curve

$$p \stackrel{\text{def}}{=} (x, y \mid x^N + y^N = 1). \quad (68)$$

Introduce a function $\varphi_p(n)$, $n \in \mathbb{Z}_N$, (cyclic analog of the quantum dilogarithm) [11]

$$\varphi_p(0) = 1, \quad \frac{\varphi_p(n-1)}{\varphi_p(n)} = \frac{1 - xq^{2n}}{y}, \quad \varphi_p(n+N) = \varphi_p(n) \quad (69)$$

The R -matrix solving (59) for the cyclic representation (63) is given by

$$\begin{aligned} \langle n | R | n' \rangle &= \delta_{n_1+n_2, n'_1+n'_2} \delta_{n_2+n_3, n'_2+n'_3} \times \\ & q^{n_1 n_3 - n'_2(n_1+n_3)} \sum_{n \in \mathbb{Z}_N} q^{-2n n'_2} \frac{\varphi_{p_1}(n+n_1+n'_3) \varphi_{p_2}(n)}{\varphi_{p_3}(n+n_1) \varphi_{p_4}(n+n_3)}, \end{aligned} \quad (70)$$

where the points p_j on the curve (68) are defined by

$$\begin{aligned} x_1 &= e^{-ia_2/N} \sqrt[N]{\sin \beta_2 / \sin \beta_0}, & y_1 &= e^{i\beta_2/N} \sqrt[N]{\sin a_2 / \sin \beta_0}, \\ x_2 &= e^{-ia_2/N} \sqrt[N]{\sin \beta_0 / \sin \beta_2}, & y_2 &= e^{i\beta_0/N} \sqrt[N]{\sin a_2 / \sin \beta_2}, \\ x_3 &= e^{-i(a_2+\pi)/N} \sqrt[N]{\sin \beta_3 / \sin \beta_1}, & y_3 &= e^{-i\beta_3/N} \sqrt[N]{\sin a_2 / \sin \beta_1}, \\ x_4 &= e^{-i(a_2+\pi)/N} \sqrt[N]{\sin \beta_1 / \sin \beta_3}, & y_4 &= e^{-i\beta_1/N} \sqrt[N]{\sin a_2 / \sin \beta_3}. \end{aligned} \quad (71)$$

We write this R -matrix as $R_{123}(\theta_1, \theta_2, \theta_3)$ to indicate its dependence on the angles $\theta_1, \theta_2, \theta_3$. It satisfies the tetrahedron equation (33), provided the R -matrices therein are parametrized by four triples of dihedral angles at four vertices of an Euclidean tetrahedron

$$\begin{aligned} R_{123} &= R_{123}(\theta_1, \theta_2, \theta_3), & R_{145} &= R_{145}(\theta_1, \pi - \theta_4, \pi - \theta_5), \\ R_{246} &= R_{246}(\pi - \theta_2, \pi - \theta_4, \theta_6), & R_{356} &= R_{356}(\theta_3, \theta_5, \theta_4). \end{aligned} \quad (72)$$

Here θ_j , $j = 1, 2, \dots, 6$ are *inner* dihedral angles of the tetrahedron. These angles are constrained by one relation, therefore Eq.(33) in this case contains five arbitrary parameters. Finally, the substitution

$$\begin{aligned} n_1 &= g + f - a - b, & n_2 &= a + c - e - g, & n_3 &= e + f - a - d, \\ n'_1 &= c + d - e - h, & n'_2 &= f + h - b - d, & n'_3 &= b + c - g - h \end{aligned} \quad (73)$$

and some trivial equivalence transformations bring (70) into its “interaction-round-a-cube” form

$$W(a|e, f, g|b, c, d|h) = \sum_{n \in \mathbb{Z}_N} q^{2n(b+d-f-h)} \frac{\varphi_{p_1}(n-h+c)\varphi_{p_2}(n-f+a)}{\varphi_{p_3}(n-b+g)\varphi_{p_4}(n-d+e)}, \quad (74)$$

which satisfies (54) where the integration is replaced by the summation over $\mathbb{Z}_N$, while the angle parametrization remains the same as in (72). Although our considerations in the cyclic case were restricted to odd values of N in (60), the final result (74) is valid for even values of N as well. The solution (74) was previously obtained in [11]; for $N = 2$ it coincides with the Baxter’s form [10] for the Boltzmann weights the Zamolodchikov model [8].

5. Conclusion

In this paper we have exposed various connections between discrete differential geometry and statistical mechanics, displaying geometric origins of algebraic structures underlying integrability of quantum systems.

We have shown that the 3D circular lattices are associated with an integrable discrete Hamiltonian system and constructed three different quantizations of this system. In this way in Sect. 4 we have obtained all previously known solutions of the tetrahedron equation. They are given by Eqs. (35), (51), (52), (70). The resulting 3D integrable models can be thought of as describing quantum fluctuations of the lattice geometry. The classical geometry of the 3D circular lattices reveals itself [6] as a stationary configuration giving the leading contribution to the partition function of the quantum system in the quasi-classical limit.

The solutions of the tetrahedron equation discussed here possess a remarkable property: they can be used to obtain [11, 16] infinite number of two-dimensional solvable models related to various representations of quantized affine algebras $U_q(\widehat{sl}_n)$, $n = 2, 3, \dots, \infty$ (where n coincides with the size of the “hidden third dimension”). Apparently, a similar 3D interpretation, originating from other simple geometrical models, also exists for the trigonometric solutions of the Yang-Baxter equation, related with all other infinite series of quantized affine algebras [45, 46] and super-algebras [47] (see [48, 49] for recent results in this direction). Therefore, it might very well be that not only the phenomenon of quantum integrability but the quantized algebras themselves are deeply connected with geometry.

There are many important outstanding questions remain, in particular, the geometric meaning of the Poisson algebra (21) and connections of the 3D circular lattices with the 2D circle patterns [50] on the plane or the sphere. It would also be interesting to understand underlying reasons of a “persistent” appearance of the q -oscillator algebra (30) as a primary algebraic structure in many other important aspects of the theory of integrable systems, such as, for example, the construction of Baxter’s $\mathbf{Q}$ -operators [51] and the calculation of correlation functions of the XXZ model [52]. It would also be interesting to explore connections of our results with the invariants of the 3D manifolds [53–55], the link invariants [56–58], quantization

of the Teichmueller space [59, 60] and the representation theory of $U_q(sl(2|\mathbb{R}))$ [61]. So, there are many interesting questions about the quantum integrability still remain unanswered, but one thing is getting more and more clear: it is not just connected with geometry, it is geometry itself! (though the Quantum Geometry).

Acknowledgments

The authors thank R.J.Baxter, M.T.Batchelor, A.Doliwa, M.Jimbo, R.M.Kashaev, T.Miwa, F.A.Smirnov, Yu.B. Suris and P. Vassiliou for interesting discussions and remarks. One of us (VB) thanks M.Staudacher for his hospitality at the Albert Einstein Institute for Gravitational Physics in Golm, where some parts of this work have been done. Special thanks to A.I.Bobenko for numerous important comments and to D. Whitehouse at the ANU Supercomputer Facility for the professional graphics of the Miquel circles (Fig. 7).

References

- [1] C. N. Yang, *Phys. Rev. Lett.* **19**, 1312 (1967).
- [2] R. J. Baxter, *Ann. Physics* **70**, 193 (1972).
- [3] V. G. Drinfel'd, Quantum groups, in *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)*, (Amer. Math. Soc., Providence, RI, 1987).
- [4] M. Jimbo, *Lett. Math. Phys.* **10**, 63 (1985).
- [5] A. A. Belavin, A. M. Polyakov and A. B. Zamolodchikov, *Nucl. Phys. B* **241**, 333 (1984).
- [6] V. V. Bazhanov, V. V. Mangazeev and S. M. Sergeev, *J. Stat. Mech.* **7**, p. P07004 (2008).
- [7] A. B. Zamolodchikov, *Soviet Phys. JETP* **52**, 325 (1980).
- [8] A. B. Zamolodchikov, *Commun. Math. Phys.* **79**, 489 (1981).
- [9] V. V. Bazhanov and Y. G. Stroganov, *Theor. Math. Phys.* **52**, 685 (1982).
- [10] R. J. Baxter, *Physica* **18D**, 321 (1986).
- [11] V. V. Bazhanov and R. J. Baxter, *J. Statist. Phys.* **69**, 453 (1992).
- [12] V. V. Bazhanov and R. J. Baxter, *J. Statist. Phys.* **71**, 839 (1993).
- [13] R. M. Kashaev, V. V. Mangazeev and Y. G. Stroganov, *Int. J. Mod. Phys. A* **8**, p. 587 (1993).
- [14] I. G. Korepanov, *J. Stat. Phys.* **71**, 85 (1993).
- [15] R. M. Kashaev, I. G. Korepanov and S. M. Sergeev, *Teoret. Mat. Fiz.* **117**, 370 (1998).
- [16] V. V. Bazhanov and S. M. Sergeev, *J. Phys. A* **39**, 3295 (2006).
- [17] V. V. Bazhanov, R. M. Kashaev, V. V. Mangazeev and Y. G. Stroganov, *Comm. Math. Phys.* **138**, 393 (1991).
- [18] E. Date, M. Jimbo, K. Miki and T. Miwa, *Commun. Math. Phys.* **137**, 133 (1991).
- [19] A. I. Bobenko, Discrete conformal maps and surfaces, in *Symmetries and integrability of difference equations (Canterbury, 1996)*, , London Math. Soc. Lecture Note Ser. Vol. 255 (Cambridge Univ. Press, Cambridge, 1999) pp. 97–108.
- [20] G. Lamé, *Leçons sur la Théorie des coordonnées curvilignes et leurs diverses applications*. (Mallet-Bachelier, Paris, 1859).
- [21] G. Darboux, *Leçons sur les systèmes orthogonaux et les coordonnées curvilignes* (Gauthier-Villars, Paris, 1910).
- [22] V. E. Zakharov and S. V. Manakov, *Funct. Anal. Appl.* **19**, 89 (1985).

- [23] I. M. Krichever, *Funct. Anal. Appl.* **31**, 25 (1997).
- [24] A. Bobenko and U. Pinkall, *J. Reine Angew. Math.* **475**, 187 (1996).
- [25] A. Doliwa and P. M. Santini, *Phys. Lett. A* **233**, 365 (1997).
- [26] V. E. Adler, A. I. Bobenko and Y. B. Suris, *Comm. Math. Phys.* **233**, 513 (2003).
- [27] B. G. Konopelchenko and W. K. Schief, *J. Phys. A* **35**, 6125 (2002).
- [28] J. Cieslinski, A. Doliwa and P. M. Santini, *Phys. Lett. A* **235**, 480 (1997).
- [29] B. G. Konopelchenko and W. K. Schief, *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* **454**, 3075 (1998).
- [30] A. Doliwa, S. Manakov and P. Santini, *Commun. Math. Phys.* **196**, 1 (1998).
- [31] A. I. Bobenko and Y. B. Suris, *Uspekhi Mat. Nauk* **62**, 3 (2007).
- [32] A. I. Bobenko and Y. B. Suris, *Discrete differential geometry: Integrable structure*, Graduate Studies in Mathematics, Vol. 98 (American Mathematical Society, Providence, RI, 2008).
- [33] J. Hietarinta, *J. Phys. A* **27**, 5727 (1994).
- [34] S. M. Sergeev, V. V. Mangazeev and Y. G. Stroganov, *J. Statist. Phys.* **82**, 31 (1996).
- [35] R. Kashaev, *Letters in Mathematical Physics* **38**, 389 (1996).
- [36] L. V. Bogdanov and B. G. Konopelchenko, *J. Phys. A* **28**, L173 (1995).
- [37] A. Doliwa and P. M. Santini, *J. Geom. Phys.* **36**, 60 (2000).
- [38] A. Miquel, *J. Math. Pur. Appl. (Liouville J.)* **3**, 517 (1838).
- [39] V. V. Bazhanov, V. V. Mangazeev and S. M. Sergeev, *Nucl. Phys. B* **784**, 234 (2007).
- [40] K. Schmüdgen, *Comm. Math. Phys.* **159**, 217 (1994).
- [41] L. D. Faddeev, *Lett. Math. Phys.* **34**, 249 (1995).
- [42] L. Faddeev, Modular double of a quantum group, in *Conférence Moshé Flato 1999, Vol. I (Dijon)*, , Math. Phys. Stud Vol. 21 (Kluwer Acad. Publ., Dordrecht, 2000). math/9912078.
- [43] B. Ponsot and J. Teschner (1999), hep-th/9911110.
- [44] L. Faddeev, Currentlike variables in massive and massless integrable models, in *Quantum groups and their applications in physics (Varenna, 1994)*, , Proc. Internat. School Phys. Enrico Fermi Vol. 127 (IOS, Amsterdam, 1996).
- [45] V. V. Bazhanov, *Phys. Lett. B* **159**, 321 (1985).
- [46] M. Jimbo, *Comm. Math. Phys.* **102**, 537 (1986).
- [47] V. V. Bazhanov and A. G. Shadrnikov, *Theor. Math. Phys.* **73**, 1303 (1987).
- [48] S. M. Sergeev, *J. Math. Phys.* **50**, p. 083519 (2009).
- [49] S. M. Sergeev, *J. Phys. A: Math. Theor.* **42**, p. 082002 (2009).
- [50] A. I. Bobenko and B. A. Springborn, *Trans. Amer. Math. Soc.* **365**, p. 659-689 (2004).
- [51] V. V. Bazhanov, S. L. Lukyanov and A. B. Zamolodchikov, *Comm. Math. Phys.* **190**, 247 (1997), [hep-th/9604044].
- [52] H. Boos, M. Jimbo, T. Miwa, F. Smirnov and Y. Takeyama, *Comm. Math. Phys.* **272**, 263 (2007), hep-th/0606280.
- [53] E. Witten, *Comm. Math. Phys.* **121**, 351 (1989).
- [54] N. Reshetikhin and V. G. Turaev, *Invent. Math.* **103**, 547 (1991).
- [55] V. G. Turaev and O. Y. Viro, *Topology* **31**, 865 (1992).
- [56] R. M. Kashaev, *Modern Phys. Lett. A* **10**, 1409 (1995).
- [57] H. Murakami and J. Murakami, *Acta Mathematica* **186**, 85 (2001).
- [58] K. Hikami, *J. Geom. Phys.* **57**, 1895 (2007), arXiv.org:math.QA/0604094.
- [59] R. M. Kashaev, *Lett. Math. Phys.* **43**, 105 (1998).
- [60] V. V. Fok and L. O. Chekhov, *Teoret. Mat. Fiz.* **120**, 511 (1999).
- [61] B. Ponsot and J. Teschner, *Comm. Math. Phys.* **224**, 613 (2001).

THE FORMATION OF BLACK HOLES IN GENERAL RELATIVITY

DEMETRIOS CHRISTODOULOU

Department of Mathematics, ETH-Zurich, ETH-Zentrum, Rämistrasse 101, 8092 Zurich, Switzerland

In 1965 Penrose [1] introduced the fundamental concept of a trapped surface, on the basis of which he proved a theorem which asserts that a spacetime containing such a surface must come to an end. The presence of a trapped surface implies, moreover, that there is a region of spacetime, the black hole, which is inaccessible to observation from infinity. A major challenge since that time has been to find out how trapped surfaces actually form, by analyzing the dynamics of gravitational collapse. I recently published a monograph [2] which achieves this aim by establishing the formation of trapped surfaces in pure general relativity through the focusing of gravitational waves. The theorems proved in the monograph constitute the first foray into the long-time dynamics of general relativity in the large, that is, when the initial data are no longer confined to a suitable neighborhood of trivial data. The main new method, the short pulse method, applies to general systems of Euler-Lagrange equations of hyperbolic type, and provides the means to tackle problems which have hitherto seemed unapproachable.

Keywords: Short pulse method; Einstein vacuum equations; trapped surfaces.

In 1965 Penrose [1] introduced the concept of a closed *trapped surface*. He defined a trapped surface to be a spacelike surface in spacetime, such that an infinitesimal virtual displacement of the surface along either family of future-directed null geodesic normals to the surface leads to a pointwise decrease of the area element. On the basis of this concept, Penrose proved an *incompleteness* theorem. In the light of subsequent work, namely the uniqueness theorem of the maximal development of given initial data by Choquet-Bruhat and Geroch [3], and the work of Rendall [4] on the characteristic initial value problem, the incompleteness theorem of Penrose may be re-stated as follows:

Let us be given regular characteristic initial data on a complete null geodesic cone C_o of a point o . Let (M^*, g) be the maximal future development of the data on C_o . Suppose that M^* contains a closed trapped surface. Then (M^*, g) is future null geodesically incomplete.

An important remark here is that it is not *a priori* obvious that closed trapped surfaces are *evolutionary*. That is, it is not obvious whether closed trapped surfaces can form in evolution starting from initial conditions in which no such surfaces are present. What is more important, the physically interesting problem is the problem where the initial conditions are arbitrarily far from already containing

closed trapped surfaces, and we are asked to follow the long time evolution and show that, under suitable circumstances, closed trapped surfaces eventually form. Only an analysis of the dynamics of gravitational collapse can achieve this aim.

John Wheeler, my physics teacher, posed to me the following problem back in 1968: *to establish the formation of black holes in pure general relativity, by the focusing of incoming gravitational waves*. I have recently solved this problem. The solution is contained in a monograph [2] which was published in the series “Monographs in Mathematics” of the European Mathematical Society.

I shall now state the simplest version of the theorem on the formation of closed trapped surfaces in pure general relativity which this monograph establishes. This is the limiting version, where we have an asymptotic characteristic initial value problem with initial data at past null infinity. Denoting by $\underline{u}$ the “advanced time”, it is assumed that the initial data are trivial for $\underline{u} \leq 0$.

Let k, l be positive constants, $k > 1$, $l < 1$. Let us be given smooth asymptotic initial data at past null infinity which is trivial for advanced time $\underline{u} \leq 0$. Suppose that the incoming energy per unit solid angle in each direction in the advanced time interval $[0, \delta]$ is not less than $k/8\pi$. Then if δ is suitably small, the maximal development of the data contains a closed trapped surface S which is diffeomorphic to S^2 and has area

$$\text{Area}(S) \geq 4\pi l^2$$

We remark that by virtue of the scale invariance of the vacuum Einstein equations, the theorem holds with k , l , and δ , replaced by ak , al , and $a\delta$, respectively, for any positive constant a .

The above theorem is obtained through a theorem in which the initial data is given on a complete future null geodesic cone C_o . The generators of the cone are parametrized by an affine parameter s measured from the vertex o and defined so that the corresponding null geodesic vector field has projection T at o along a fixed unit future-directed timelike vector T at o . It is assumed that the initial data are trivial for $s \leq r_0$, for some $r_0 > 1$. The boundary of this trivial region is then a round sphere of radius r_0 .

The advanced time $\underline{u}$ is then defined along C_o by

$$\underline{u} = s - r_0 \tag{1}$$

The formation of closed trapped surfaces theorem is similar in this case, the only difference being that the “incoming energy per unit solid angle in each direction in the advanced time interval $[0, \delta]$ ”, a notion defined only at past null infinity, is replaced by the integral

$$\frac{r_0^2}{8\pi} \int_0^\delta e d\underline{u} \tag{2}$$

on the affine parameter segment $[r_0, r_0 + \delta]$ of each generator of C_o .

The function e is an invariant of the conformal intrinsic geometry of C_o , given by:

$$e = \frac{1}{2} |\hat{\chi}|_{\hat{g}}^2 \quad (3)$$

where $\hat{g}$ is the induced metric on the sections of C_o corresponding to constant values of the affine parameter, and $\hat{\chi}$ is the *shear* of these sections, the trace-free part of their 2nd fundamental form relative to C_o .

The theorem for a cone C_o is established for any $r_0 > 1$ and the smallness condition on δ is independent of r_0 . The domain of dependence, in the maximal development, of the trivial region in C_o is a domain in Minkowski spacetime bounded in the past by the trivial part of C_o and in the future by $\underline{C}_e$, the past null geodesic cone of a point e at arc length $2r_0$ along the timelike geodesic Γ_0 from o with tangent vector T at o . Considering then the corresponding complete timelike geodesic in Minkowski spacetime, fixing the origin on this geodesic to be the point e , the limiting form of the theorem is obtained by letting $r_0 \rightarrow \infty$, keeping the origin fixed, so that o tends to the infinite past along the timelike geodesic.

In the monograph [2], almost all the work goes into establishing an *existence theorem* for a development of the initial data which extends far enough into the future so that trapped spheres have eventually chance to form within this development. On the other hand, there is a wealth of information in this existence theorem, which gives us full knowledge of the geometry of spacetime when closed trapped surfaces begin to form.

I shall now give a brief discussion of the mathematical methods employed. The monograph relies on three methods, two of which originated in my work with Klainerman [5] on the stability of the Minkowski spacetime, and the third method is new. I shall first summarize the first two methods.

The first method is peculiar to Einstein's equations, while the second has wider application, and can, in principle, be extended to all Euler-Lagrange systems of partial differential equations of hyperbolic type.

The first method is a way of looking at Einstein's equations which allows estimates for the spacetime curvature to be obtained. Instead of considering the Einstein equations themselves, we considered the Bianchi identities in the form which they assume by virtue of the Einstein equations. We then introduced the general concept of a *Weyl field* W on a 4-dimensional Lorentzian manifold (M, g) to be a 4-covariant tensorfield with the algebraic properties of the Weyl or *conformal* curvature tensor. Given a Weyl field W one can define a left dual ${}^*W$ as well as a right dual W^* , but as a consequence of the algebraic properties of a Weyl field the two duals coincide.

Moreover, ${}^*W = W^*$ is also a Weyl field. A Weyl field is subject to equations which are analogues of Maxwell's equations for the electromagnetic field. These are linear equations, in general inhomogeneous, which we call *Bianchi equations*. They are of the form:

$$\nabla^\alpha W_{\alpha\beta\gamma\delta} = J_{\beta\gamma\delta} \quad (4)$$

the right hand side J , or more generally any 3-covariant tensorfield with the algebraic properties of the right hand side, we call a *Weyl current*.

These equations seem at first sight to be the analogues of only half of Maxwell's equations, but it turns out that they are equivalent to the equations

$$\nabla_{[\alpha} W_{\beta\gamma]\delta\epsilon} = \epsilon_{\mu\alpha\beta\gamma} J^{*\mu}_{\delta\epsilon}, \quad J^*_{\beta\gamma\delta} = \frac{1}{2} J_{\beta}{}^{\mu\nu} \epsilon_{\mu\nu\gamma\delta} \quad (5)$$

which are analogues of the other half of Maxwell's equations. Here ϵ is the volume 4-form of (M, g) . The fundamental Weyl field is the Riemann curvature tensor of (M, g) , (M, g) being a solution of the vacuum Einstein equations, and in this case the corresponding Weyl current vanishes, the Bianchi equations reducing to the Bianchi identities.

Given a vector field Y and a Weyl field W or Weyl current J there is a "variation" of W and J with respect to Y , a modified Lie derivative $\tilde{\mathcal{L}}_Y W$, $\tilde{\mathcal{L}}_Y J$, which is also a Weyl field or Weyl current respectively. The modified Lie derivative commutes with duality. The Bianchi equations have certain conformal covariance properties which imply the following. If J is the Weyl current associated to the Weyl field W according to the Bianchi equations, then the Weyl current associated to $\tilde{\mathcal{L}}_Y W$ is the sum of $\tilde{\mathcal{L}}_Y J$ and a bilinear expression which is on one hand linear in ${}^{(Y)}\tilde{\pi}$ and its first covariant derivative and other the other hand in W and its first covariant derivative.

Here we denote by ${}^{(Y)}\tilde{\pi}$ the *deformation tensor* of Y , namely the trace-free part of the Lie derivative of the metric g with respect to Y . This measures the rate of change of the conformal geometry of (M, g) under the flow generated by Y . From the fundamental Weyl field, the Riemann curvature tensor of (M, g) , and a set of vector fields $Y_1, \dots, Y_n$ which we call *commutation fields*, derived Weyl fields of up to any given order m are generated by the repeated application of the operators $\tilde{\mathcal{L}}_{Y_i} : i = 1, \dots, n$.

A basic requirement on the set of commutation fields is that it spans the tangent space to M at each point. The Weyl currents associated to these derived Weyl fields are then determined by the deformation tensors of the commutation fields.

Given a Weyl field W there is a 4-covariant tensorfield $Q(W)$ associated to W , which is symmetric and trace-free in any pair of indices. It is a quadratic expression in W , analogous to the Maxwell energy-momentum-stress tensor for the electromagnetic field.

We call $Q(W)$ the *Bel-Robinson tensor* associated to W , because it coincides with the tensor discovered by Bel and Robinson in the case of the fundamental Weyl field, the Riemann curvature tensor of a solution of the vacuum Einstein equations.

The Bel-Robinson tensor has a remarkable positivity property:

$Q(W)(X_1, X_2, X_3, X_4)$ is non-negative for any tetrad X_1, X_2, X_3, X_4 of future directed non-spacelike vectors at a point. Moreover, the divergence of $Q(W)$ is a bilinear expression which is linear in W and in the associated Weyl current J .

Given a Weyl field W and a triplet of future directed non-spacelike vectorfields X_1, X_2, X_3 , which we call *multiplier fields* we define the *energy-momentum density* vectorfield $P(W; X_1, X_2, X_3)$ associated to W and to the triplet X_1, X_2, X_3 by:

$$P(W; X_1, X_2, X_3)^\alpha = -Q(W)^\alpha_{\beta\gamma\delta} X_1^\beta X_2^\gamma X_3^\delta \quad (6)$$

Then the divergence of $P(W; X_1, X_2, X_3)$ is the sum of $-(\text{div} Q(W))(X_1, X_2, X_3)$ and a bilinear expression which is linear in $Q(W)$ and in the deformation tensors of X_1, X_2, X_3 .

The divergence theorem in spacetime, applied to a domain which is a development of part of the initial hypersurface, then expresses the integral of the 3-form dual to $P(W; X_1, X_2, X_3)$ on the future boundary of this domain, in terms of the integral of the same 3-form on the past boundary of the domain, namely on the part of the initial hypersurface, and the spacetime integral of the divergence. The boundaries being *achronal* – that is, no pair of points on each boundary can be joined by a timelike curve – the integrals are integrals of non-negative functions, by virtue of the positivity property of $Q(W)$.

For the set of Weyl fields of order up to m which are derived from the fundamental Weyl field, the Riemann curvature tensor of (M, g) , the divergences are determined by the deformation tensors of the commutation fields and their derivatives up to order m , and by the deformation tensors of the multiplier fields. And the integrals on the future boundary give control of all the derivatives of the curvature up to order m . This is how estimates for the spacetime curvature are obtained, once a suitable set of multiplier fields and a suitable set of commutation fields have been provided.

This is precisely where the second method comes in. This method constructs the required sets of vectorfields by using the geometry of the two parameter foliation of spacetime by the level sets of two functions. These two functions, in the first realization of this method, where the *time function* t , the level sets of which are maximal spacelike hypersurfaces H_t of vanishing total linear momentum, and the *optical function* u , which we may think of as “retarded time”, the level sets of which are outgoing null hypersurfaces C_u . These were chosen so that density of the foliation of each H_t by the traces of the C_u , that is, by the surfaces of intersection $S_{t,u} = H_t \cap C_u$, which are diffeomorphic to S^2 , tends to 1 as $t \rightarrow \infty$.

It was clear that the two functions did not enter the problem on equal footing. The optical function u played a much more important role. This is due to the fact that the problem involved outgoing waves reaching future null infinity, and it is the outgoing family of null hypersurfaces C_u which follow these waves. The role of the family of maximal spacelike hypersurfaces H_t was to obtain a suitable family of sections of each C_u , the family $S_{t,u}$ corresponding to a given u , and to serve as a means by which, in the proof of the existence theorem, the method of continuity can be applied.

The geometric entities describing the two parameter foliation of spacetime by the $S_{t,u}$ are estimated in terms of the spacetime curvature. This yields estimates

for the deformation tensors of the multiplier fields and the commutation fields in terms of the spacetime curvature, thus connecting with the first method.

A variant of the second method is obtained if we place in the role of the time function t another *optical function* $\underline{u}$, which we may think of as “advanced time”, the level sets of which are incoming null hypersurfaces. A two parameter family of surfaces diffeomorphic to S^2 , the “wave fronts”, are then obtained, namely the intersections of this incoming family with the outgoing family of null hypersurfaces.

In the present work, the roles of the two optical functions are reversed, because we are considering incoming rather than outgoing waves, and it is the incoming null hypersurfaces $\underline{C}_{\underline{u}}$, the level sets of $\underline{u}$, which follow these waves. However, in the present work, taking the other function to be the conjugate optical function u is not merely a matter of convenience, but it is essential for what we wish to achieve. This is because the C_u , the level sets of u , are here, like the initial hypersurface C_o itself, future null geodesic cones with vertices on the timelike geodesic Γ_0 , and the trapped spheres which eventually form are sections $S_{\underline{u},u} = \underline{C}_{\underline{u}} \cap C_u$ of “late” C_u , everywhere along which those C_u have negative expansion.

We now come to the new method. This method is a method of treating the focusing of incoming waves, and like the second method it is of wider application. A suitable name for this method is *short pulse method*. Its point of departure resembles that of the short wavelength or geometric optics approximation, in so far as it depends on the presence of a small length, but thereafter the two approaches diverge.

The short pulse method is a method which, in the context of Euler-Lagrange systems of partial differential equations of hyperbolic type, allows us to establish an existence theorem for a development of the initial data which is large enough so that interesting things have chance to occur within this development, if a nonlinear system is involved. One may ask at this point: what does it mean for a length to be small in the context of the vacuum Einstein equations? For, the equations are scale invariant. Here *small* means *by comparison to the area radius of the trapped sphere to be formed*.

With initial data on a complete future null geodesic cone C_o , as explained above, which are trivial for $s \leq r_0$, we consider the restriction of the initial data to $s \leq r_0 + \delta$. In terms of the advanced time $\underline{u}$, we restrict attention to the interval $[0, \delta]$, the data being trivial for $\underline{u} \leq 0$. The retarded time u is set equal to $u_0 = -r_0$ at o and therefore on C_o , which is then also denoted C_{u_0} . Also, $u - u_0$ is defined along Γ_0 to be one half the arc length from o . This determines u everywhere. The development whose existence we want to establish is that bounded in the future by the spacelike hypersurface H_{-1} where $\underline{u} + u = -1$ and by the incoming null hypersurface $\underline{C}_{\delta}$. We denote this development M_{-1} .

We define L and $\underline{L}$ to be the future directed null vectorfields the integral curves of which are the generators of the C_u and $\underline{C}_{\underline{u}}$, parametrized by $\underline{u}$ and u respectively, so that

$$Lu = \underline{L}\underline{u} = 0, \quad L\underline{u} = \underline{L}u = 1 \quad (7)$$

The flow Φ_τ generated by L defines a diffeomorphism of $S_{\underline{u},u}$ onto $S_{\underline{u}+\tau,u}$, while the flow $\underline{\Phi}_\tau$ generated by $\underline{L}$ defines a diffeomorphism of $S_{\underline{u},u}$ onto $S_{\underline{u},u+\tau}$.

The positive function Ω defined by

$$g(L, \underline{L}) = -2\Omega^2 \quad (8)$$

may be thought of as the inverse density of the double null foliation. We denote by $\hat{L}$ and $\hat{\underline{L}}$ the corresponding normalized future directed null vectorfields

$$\hat{L} = \Omega^{-1}L, \quad \hat{\underline{L}} = \Omega^{-1}\underline{L}, \quad \text{so that } g(\hat{L}, \hat{\underline{L}}) = -2 \quad (9)$$

The first step is the analysis of the equations along the initial hypersurface C_{u_0} . The analysis is particularly clear and simple because of the fact that C_{u_0} is a null hypersurface, so we are dealing with the characteristic initial value problem and there is a way of formulating the problem in terms of free data which are not subject to any constraints. The full set of data which includes all the curvature components and their transversal derivatives, up to any given order, along C_{u_0} , is then determined by integrating ordinary differential equations along the generators of C_{u_0} . We show that the free data may be described as a 2-covariant symmetric positive definite tensor density m , of weight -1 and unit determinant, on S^2 , depending on $\underline{u}$.

This is of the form:

$$m = \exp \psi \quad (10)$$

where ψ is a 2-dimensional symmetric trace-free matrix valued “function” on S^2 , depending on $\underline{u} \in [0, \delta]$, and transforming under change of charts on S^2 in such a way so as to make m a 2-covariant tensor density of weight -1. The transformation rule is particularly simple if stereographic charts on S^2 are used.

Then there is a function O defined on the intersection of the domains of the north and south polar stereographic charts on S^2 , with values in the 2-dimensional symmetric orthogonal matrices of determinant -1 such that in going from the north polar chart to the south polar chart or vice-versa, $\psi \mapsto \tilde{O}\psi O$ and $m \mapsto \tilde{O}mO$.

The crucial Ansatz of the short pulse method is the following. We consider an arbitrary smooth 2-dimensional symmetric trace-free matrix valued “function” ψ_0 on S^2 , depending on $s \in [0, 1]$, which extends smoothly by 0 to $s \leq 0$, and we set:

$$\psi(\underline{u}, \vartheta) = \frac{\delta^{1/2}}{|u_0|} \psi_0 \left(\frac{\underline{u}}{\delta}, \vartheta \right), \quad (\underline{u}, \vartheta) \in [0, \delta] \times S^2 \quad (11)$$

The analysis of the equations along C_{u_0} then gives, for the components of the spacetime curvature along C_{u_0} :

$$\begin{aligned}
\sup_{C_{u_0}} |\alpha| &\leq O_2(\delta^{-3/2}|u_0|^{-1}) \\
\sup_{C_{u_0}} |\beta| &\leq O_2(\delta^{-1/2}|u_0|^{-2}) \\
\sup_{C_{u_0}} |\rho|, \sup_{C_{u_0}} |\sigma| &\leq O_3(|u_0|^{-3}) \\
\sup_{C_{u_0}} |\underline{\beta}| &\leq O_4(\delta|u_0|^{-4}) \\
\sup_{C_{u_0}} |\underline{\alpha}| &\leq O_5(\delta^{3/2}|u_0|^{-5})
\end{aligned} \tag{12}$$

Here $\alpha, \underline{\alpha}$ are the trace-free symmetric 2-covariant tensorfields on each $S_{\underline{u}, u}$ defined by:

$$\alpha(X, Y) = R(X, \hat{L}, Y, \hat{L}), \quad \underline{\alpha}(X, Y) = R(X, \underline{\hat{L}}, Y, \underline{\hat{L}}) \tag{13}$$

for any pair of vectors X, Y tangent to $S_{\underline{u}, u}$ at a point, $\beta, \underline{\beta}$ are the 1-forms on each $S_{\underline{u}, u}$ defined by:

$$\beta(X) = \frac{1}{2}R(X, \hat{L}, \underline{\hat{L}}, \hat{L}), \quad \underline{\beta}(X) = \frac{1}{2}R(X, \underline{\hat{L}}, \underline{\hat{L}}, \hat{L}) \tag{14}$$

and ρ, σ are the functions on each $S_{\underline{u}, u}$ defined by:

$$\rho = \frac{1}{4}R(\underline{\hat{L}}, \hat{L}, \underline{\hat{L}}, \hat{L}), \quad \frac{1}{2}R(X, Y, \underline{\hat{L}}, \hat{L}) = \sigma \not\epsilon(X, Y) \tag{15}$$

for any pair of vectors X, Y tangent to $S_{\underline{u}, u}$ at a point, $\not\epsilon$ being the area form of $S_{\underline{u}, u}$. The symbol $O_k(\delta^p|u_0|^r)$ means the product of $\delta^p|u_0|^r$ with a non-negative non-decreasing continuous function of the C^k norm of ψ_0 on $[0, 1] \times S^2$. The pointwise magnitudes of tensors on $S_{\underline{u}, u}$ are with respect to the induced metric $\not{g}$, which is positive definite, the surfaces being spacelike.

One should focus on the dependence on δ of the right hand sides of 12. This displays what we may call the *short pulse hierarchy*. And this hierarchy is *nonlinear*. For, if only the linearized form of the equations was considered, a different hierarchy would be obtained: the exponents of δ in the first two of 12 would be the same, but the exponents of δ in the last three of 12 would instead be $1/2, 3/2, 5/2$, respectively.

A question that immediately comes up when one ponders the Ansatz 11, is why is the “amplitude” of the pulse proportional to the square root of the “length” of the pulse? (the factor $|u_0|^{-1}$ is the standard decay factor in 3 spatial dimensions, the square root of the area of the wave fronts). Where does this relationship come from? Obviously, there is no such relationship in a linear theory.

The answer is that it comes from our desire to form trapped surfaces in the development M_{-1} . If a problem involving the focusing of incoming waves in a different context was the problem under study, for example the formation of electromagnetic shocks by the focusing of incoming electromagnetic waves in a nonlinear medium,

the relationship between length and amplitude would be dictated by the desire to form such shocks within our development.

The short pulse hierarchy is the key to the existence theorem as well as to the trapped surface formation theorem. We must still outline however in what way do we establish that the short pulse hierarchy is preserved in evolution. This is of course the main step of the short pulse method. What we do is to reconsider the first two methods previously outlined in the light of the short pulse hierarchy.

Let us revisit the first method. We take as multiplier fields the vectorfields L and K , where

$$K = u^2 \underline{L} \quad (16)$$

As already mentioned above, we take the initial data to be trivial for $\underline{u} \leq 0$ and as a consequence the spacetime region corresponding to $\underline{u} \leq 0$ is a domain in Minkowski spacetime. We may thus confine attention to the nontrivial region $\underline{u} \geq 0$. We denote by M'_{-1} this non-trivial region in M_{-1} .

For each of the Weyl fields to be specified below, we define the energy-momentum density vectorfields

$$^{(n)}P(W) : n = 0, 1, 2, 3 \quad (17)$$

where:

$$\begin{aligned} ^{(0)}P(W) &= P(W; L, L, L) \\ ^{(1)}P(W) &= P(W; K, L, L) \\ ^{(2)}P(W) &= P(W; K, K, L) \\ ^{(3)}P(W) &= P(W; K, K, K) \end{aligned} \quad (18)$$

As commutation fields we take L , S , defined by:

$$S = u \underline{L} + \underline{u} L, \quad (19)$$

and the three rotation fields $O_i : i = 1, 2, 3$. The latter are defined according to the second method as follows. In the Minkowskian region we introduce rectangular coordinates $x^\mu : \mu = 0, 1, 2, 3$, taking the x^0 axis to be the timelike geodesic Γ_0 . In the Minkowskian region, in particular on the sphere S_{0,u_0} , the O_i are the generators of rotations about the $x^i : i = 1, 2, 3$ spatial coordinate axes. The O_i are then first defined on C_{u_0} by conjugation with the flow of L and then in spacetime by conjugation with the flow of $\underline{L}$.

The Weyl fields which we consider are, besides the fundamental Weyl field R , the Riemann curvature tensor, the following derived Weyl fields

$$\begin{aligned} \text{1st order: } & \tilde{\mathcal{L}}_L R, \tilde{\mathcal{L}}_{O_i} R : i = 1, 2, 3, \tilde{\mathcal{L}}_S R \\ \text{2nd order: } & \tilde{\mathcal{L}}_L \tilde{\mathcal{L}}_L R, \tilde{\mathcal{L}}_{O_i} \tilde{\mathcal{L}}_L R : i = 1, 2, 3, \\ & \tilde{\mathcal{L}}_{O_j} \tilde{\mathcal{L}}_{O_i} R : i, j = 1, 2, 3, \\ & \tilde{\mathcal{L}}_{O_i} \tilde{\mathcal{L}}_S R : i = 1, 2, 3, \tilde{\mathcal{L}}_S \tilde{\mathcal{L}}_S R \end{aligned} \quad (20)$$

We assign to each Weyl field the index l according to the number of $\tilde{\mathcal{L}}_L$ operators in the definition of W in terms of R . We then define total 2nd order energy-momentum densities

$$P_2^{(n)} : n = 0, 1, 2, 3 \quad (21)$$

as the sum of $\delta^{2l} P^{(n)}(W)$ over all the above Weyl fields in the case $n = 3$, all the above Weyl fields except those whose definition involves the operator $\tilde{\mathcal{L}}_S$ in the cases $n = 0, 1, 2$.

We then define the total 2nd order energies $E_2^{(n)}(u)$ as the integrals on the C_u and the total 2nd order fluxes $F_2^{(n)}(\underline{u})$ as the integrals on the $\underline{C}_{\underline{u}}$, of the 3-forms dual to the $P_2^{(n)}$. Of the fluxes only $F_2^{(3)}(\underline{u})$ plays a role in the problem. Finally, with the exponents $q_n : n = 0, 1, 2, 3$ defined by:

$$q_0 = 1, \quad q_1 = 0, \quad q_2 = -\frac{1}{2}, \quad q_3 = -\frac{3}{2}, \quad (22)$$

according to the short pulse hierarchy, we define the quantities

$$\begin{aligned} \mathcal{E}_2^{(n)} &= \sup_u \left(\delta^{2q_n} E_2^{(n)}(u) \right) : n = 0, 1, 2, 3, \\ \mathcal{F}_2^{(3)} &= \sup_{\underline{u}} \left(\delta^{2q_3} F_2^{(3)}(\underline{u}) \right) \end{aligned} \quad (23)$$

The objective then is to obtain bounds for these quantities in terms of the initial data.

This requires properly estimating the deformation tensor of K , as well as the deformations tensors of L, S and the $O_i : i = 1, 2, 3$ and their derivatives of up to 2nd order. In doing this, the short pulse method meshes with the second method previously described.

The estimates of the error integrals, namely the integrals of the absolute values of the divergences of the $P_2^{(n)}$, then yield inequalities for the quantities 23. These inequalities contain, besides the initial data terms

$$D^{(n)} = \delta^{2q_n} E_2^{(n)}(u_0) : n = 0, 1, 2, 3, \quad (24)$$

terms of $O(\delta^p)$ for some $p > 0$, which are innocuous, as they can be made less than or equal to 1 by subjecting δ to a suitable smallness condition, *but they also contain terms of $O(1)$ which are nonlinear in the quantities 23.*

From such a nonlinear system of inequalities, no bounds can in general be deduced, because here, in contrast with the work on the stability of Minkowski space-time, the initial data quantities are allowed to be arbitrarily large. However a miracle occurs: our system of inequalities is *reductive*. That is, the inequalities, taken in proper sequence, reduce to a sequence of sublinear inequalities, thus allowing us to obtain the sought for bounds.

References

- [1] R. Penrose, “Gravitational collapse and space-time singularities”, *Phys. Rev. Lett.* **14**, 57 - 59 (1965).
- [2] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, EMS Publishing House, 2009.
- [3] Y. Choquet-Bruhat and R. P. Geroch, “Global aspects of the Cauchy problem in general relativity”, *Commun. Math. Phys.* **14**, 329 - 335 (1969).
- [4] A. D. Rendall, “Reduction of the characteristic initial value problem to the Cauchy problem and its applications to the Einstein equations”, *Proc. Roy. Soc. Lond. A* **427**, 221 - 239 (1990).
- [5] D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton Mathematical Series **41**, Princeton University Press, 1993.

LIOUVILLE QUANTUM GRAVITY & THE KPZ RELATION: *A Rigorous Perspective*

BERTRAND DUPLANTIER

Institut de Physique Théorique, CNRS/URA 2306

Bât. 774, Orme des Merisiers, CEA/Saclay, F-91191 Gif-sur-Yvette Cedex, France

E-mail: bertrand.duplantier@cea.fr

This Proceeding describes joint work [1, 2] with SCOTT SHEFFIELD, and presents a (mathematically rigorous) probabilistic and geometrical proof of the Knizhnik–Polyakov–Zamolodchikov (KPZ) relation between scaling exponents in a Euclidean planar domain D and in Liouville quantum gravity. The Liouville quantum gravity measure on D is the weak limit as $\varepsilon \rightarrow 0$ of $d\mu_{\gamma, \varepsilon} = \varepsilon^{\gamma^2/2} e^{\gamma h_\varepsilon(z)} dz$, where dz is the standard 2D Euclidean (i.e., Lebesgue) measure on D , γ is a real parameter, $0 \leq \gamma < 2$, and $h_\varepsilon(z)$ denotes the mean value on the circle of radius ε centered at z of an instance h of the Gaussian free field on D . The proof can be extended to the boundary geometry. The singular case $\gamma > 2$ is related to the quantum measure $d\mu_{\gamma'}$, $\gamma' < 2$, by the fundamental duality $\gamma\gamma' = 4$.

Keywords: 2D Liouville quantum gravity; Gaussian free field; random measure; Knizhnik–Polyakov–Zamolodchikov relation.

1. Introduction

1.1. *Historical perspective*

One of the major theoretical advances in physics over the past thirty years has been the realization in gauge theory or string theory that transition amplitudes require summing over random surfaces, which replaces traditional sums over random paths, i.e., the celebrated Feynman path integrals of quantum mechanics and quantum field theory, as Polyakov announced in his 1981 seminal article introducing Liouville quantum gravity [3]. ^a

The study of certain natural probability measures on the space of two dimensional Riemannian manifolds (and singular limits of these manifolds) is often called

^aPolyakov writes in the introduction to [3]: “There are methods and formulae in science, which serve as master-keys to many apparently different problems. The resources of such things have to be refilled from time to time. In my opinion at the present time we have to develop an art of handling sums over random surfaces. These sums replace the old-fashioned (and extremely useful) sums over random paths. The replacement is necessary, because today gauge invariance plays the central role in physics. Elementary excitations in gauge theories are formed by the flux lines (closed in the absence of charges) and the time development of these lines forms the world surfaces. All transition amplitudes are given by the sums over all possible surfaces with fixed boundary.”

“two-dimensional quantum gravity.” These models have been very thoroughly studied in the physics literature, in part because of connections to string theory and conformal field theory [3–22], and to random matrix theory and discrete geometrical models; see, e.g., the references [23–62]. More recently, a purely combinatorial approach to discretized quantum gravity has been successful [63–94]. The topological expansion in powers of the genus of the surface has also witnessed significant progress; see Refs. [95–100].

Polyakov [3] first understood that the summation over random Riemannian metrics involved could be represented mathematically by the now celebrated *Liouville theory of quantum gravity*. The latter can be simply described as follows: Consider a bounded planar domain $D \subset \mathbb{C}$ as the parameter domain of the random Riemannian surface, and an instance h of the Gaussian free field (GFF) on D , with Dirichlet energy

$$(h, h)_{\nabla} := \frac{1}{2\pi} \int_D \nabla h(z) \cdot \nabla h(z) dz. \quad (1)$$

The quantum area is then (formally) defined by $\mathcal{A} = \int_D e^{\gamma h(z)} dz$, where dz is the standard 2D Euclidean (i.e., Lebesgue) measure and $e^{\gamma h(z)}$ the *random* conformal factor of the Riemannian metric, with a constant $0 \leq \gamma < 2$. The quantum Liouville action is then

$$S(h) = \frac{1}{2}(h, h)_{\nabla} + \lambda \mathcal{A}, \quad (2)$$

where $\lambda \geq 0$ is the so-called “cosmological constant”. The corresponding Boltzmann-Gibbs weight is thus

$$\exp[-S(h)] = \exp\left[-\frac{1}{2}(h, h)_{\nabla}\right] \exp(-\lambda \mathcal{A}), \quad (3)$$

to be integrated over with a “flat” functional measure $\mathcal{D}h$ on h , defined heuristically as a “uniform measure on the space of all functions.” (Of course, the latter makes perfect sense if one considers only a finite dimensional vector space of functions, such as real-valued functions defined on the vertices of a lattice, or functions whose Fourier coefficients beyond a certain frequency threshold are identically zero — in this case $\mathcal{D}h$ would be the Lebesgue measure on the vector space.)

Discrete quantum gravity models (a.k.a. random planar map models, random triangulation or quadrangulation models, etc.) have natural interpretations as continuum random metric spaces as well. For example, a random planar quadrangulation M_n on the sphere, or on the disk (Fig. 1) — chosen uniformly from the set of all simply connected planar quadrangulations with n quadrilaterals — can be viewed as a manifold by endowing each quadrilateral with the metric of a unit square. (The resulting manifold will have singularities: negative curvature point masses at vertices where more than four unit squares coincide and positive curvature point masses at vertices where fewer than four unit squares coincide.) One may then choose a uniform square from among this set. This random surface can be conformally mapped to the plane $\mathbb{C}$, or the unit disk $\mathbb{D}$, in such a way that the center of the distinguished

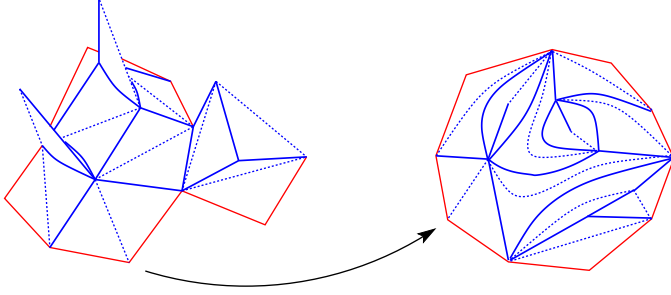


Fig. 1. A random quadrangulation made of uniform unit squares, with a disk topology, and the equivalent topological planar map. [To facilitate the (arbitrary) embedding in ambient space, the quadrangles have been folded along some diagonals (dotted lines).]

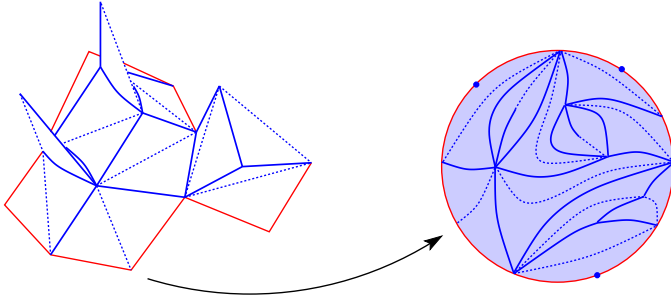


Fig. 2. Conformal map of the same manifold to the unit disk $\mathbb{D}$, with distortion of the unit squares.

square is mapped to the origin. The images of the unit squares of M_n form a tiling of $\mathbb{C}$ or $\mathbb{D}$ by “conformally distorted” unit squares, different squares having different sizes with respect to the Euclidean metric on the plane or on the disk (Fig. 2).

Taking an “infinite volume limit” (as $n \rightarrow \infty$) one obtains an infinite random quadrangulation M_∞ with a distinguished square. (See, e.g., [69] for a precise description of this construction for triangulations.) An alternative is to give a weight $e^{-\lambda n}$ to manifolds M_n of discrete area n . The pullback of the random area measure of the manifold M_∞ , or of the weighted set of M_n ’s, to the plane or the disk via the conformal map is expected to essentially take the Liouville form $e^{\gamma h} dz$, with a random conformal factor which is the exponential of some multiple of the GFF h ; the exponential factor $e^{-\lambda \mathcal{A}}$ in Equation (3) then corresponds to the similar exponential weight on n . Note that this equivalence of Liouville quantum gravity and discrete quantum gravity is taken as an Ansatz throughout much of the literature.

Kazakov introduced in 1986 the key idea of placing (critical) statistical models on a random planar lattice, when exactly solving the Ising model on a random triangulation [29] (Fig. 3). In this case, the probability of a given map (e.g., a triangulation or a quadrangulation) is proportional to the partition function of the

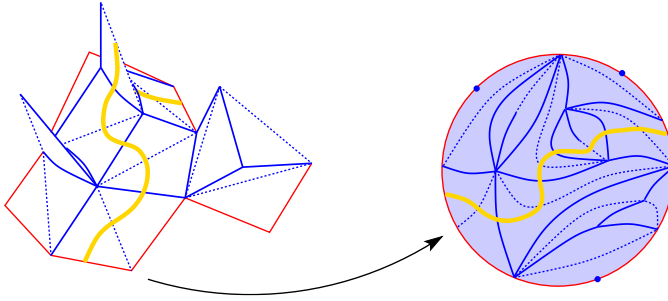


Fig. 3. Geometrical statistical model (here represented by a random path) born by the random quadrangulation, and the conformal map to $\mathbb{D}$.

statistical physics model on that map. Various values of γ are expected by choosing a random map together with a statistical physical model on that map (FK cluster model, percolation, $O(N)$ model, uniform spanning tree). (See the references on random matrix theory and geometrical models cited above for much more detail.) By the usual conformal invariance Ansatz, it is natural to expect that if one conditions on the map, and then samples the loops or clusters in these models (as mapped into the plane), their law (in the scaling limit) will be *independent* of the random metric.

One of the most influential papers in this field is a 1988 work of Knizhnik, Polyakov, and Zamolodchikov [7]. Building on a 1987 work of Polyakov [5], the authors derive a relationship (the **KPZ formula**) between scaling dimensions (i.e., conformal weights x) of fields defined using Euclidean geometry and analogous dimensions (Δ) defined via Liouville quantum gravity (as described earlier in Refs. [3, 4]):

$$x = \frac{\gamma^2}{4} \Delta^2 + \left(1 - \frac{\gamma^2}{4}\right) \Delta. \quad (4)$$

The positive inverse to relation (4) is

$$\Delta_\gamma := \frac{1}{\gamma} \left(\sqrt{4x + \left(\frac{2}{\gamma} - \frac{\gamma}{2}\right)^2} - \left(\frac{2}{\gamma} - \frac{\gamma}{2}\right) \right). \quad (5)$$

In the critical continuum limit, the statistical system born by the random lattice is described by a conformal field theory (CFT) with central charge $c \leq 1$, which *fixes* the value of γ as a function of c [7]

$$\gamma = \frac{1}{\sqrt{6}} (\sqrt{25 - c} - \sqrt{1 - c}) \leq 2, \quad c \leq 1. \quad (6)$$

With this identification, one finds the usual form of the KPZ relation

$$\Delta = \frac{\sqrt{24x + 1 - c} - \sqrt{1 - c}}{\sqrt{25 - c} - \sqrt{1 - c}}. \quad (7)$$

The first check of the KPZ relation was provided by the earlier result by Kazakov on the Ising model on a random lattice. This fact strongly suggested that discrete random lattice or matrix models could provide a regularization of continuum Liouville quantum gravity, as stressed in the recent historical note [101].^b Another confirmation soon followed with the solution to the multiple polymer spectrum on a randomly triangulated lattice [30]. The original derivation by KPZ used the so-called light-cone gauge. An alternative heuristic derivation using Liouville field theory in the so-called conformal gauge was proposed shortly after [8, 9] (see also Ref. [102]).

The KPZ relation was checked in multiple ways using explicit calculations in geometrical models on random planar lattices using the random matrix theory approach (see, e.g., Refs. [29–31, 33, 37, 45, 46, 49, 50, 61, 62, 103, 104]). Some of these calculations could in particular be compared (see, e.g., Refs. [103, 104]) to the results of the elaborate techniques developed for calculating correlation functions in Liouville field theory [105–114] (for reviews, see Ref. [115], and the Lecture Notes [22] and [62].) Let us add that 2D quantum gravity and the KPZ relation were also used to successfully predict the intersection exponents of planar Brownian paths [50], as well as the harmonic measure and rotation multifractal spectra for conformally invariant curves in the plane [21, 51, 53, 57, 59, 60]; these results were later proven rigorously using the celebrated Schramm–Loewner Evolution (SLE) [116–118].

Liouville quantum gravity provides the core continuous model of “2D quantum gravity”, whose deep and manifold connections to string theory, conformal field theory, random planar lattice models, random matrix theory and SLE are often still conjectural from a mathematical perspective. In particular, and despite its great importance for conformal field theory, the KPZ relation (4) had never been proven rigorously, nor its range of validity properly defined, and not even its geometrical meaning fully understood. The first proof appeared in joint work with Scott Sheffield [1]. The aim of this contribution is to present such a proof in a minimal, yet rigorous way, which is adapted from Refs. [1, 2]. Several follow-up works exist, either at the rigorous level [119, 120] or heuristic one [121].

In this geometrical and probabilistic approach, we start from the *critical* Liouville gravity, with action S (2) taken at $\lambda = 0$, i.e., a *free-field* action. The rôle of the cosmological constant λ in weight (3) is indeed to control the expectation of the whole quantum area; it thus acts as an “infrared” regulator, while the KPZ relation is a local scaling property of the quantum measure, hence an “ultraviolet” phenomenon, therefore independent of the value of λ .

We define a properly regularized quantum area measure, which allows for a transparent probabilistic understanding of KPZ relation (4) for any scaling fractal

^bIn [101], Polyakov writes: “It was hard to be certain that these models really have a continuous limit described by the Liouville theory, there were no proofs of this conjecture. To our surprise we found that the anomalous dimensions coming from our theory coincide with the ones computed from the matrix model. That left no doubts that in the case of the minimal models the Liouville description is equivalent to the matrix one. This relation received a lot of attention.”

set in D , as a direct consequence of the underlying *Brownian stochastic properties* of the two-dimensional GFF. The boundary analog of KPZ for fractal subsets of the boundary ∂D can be studied in the same way [1, 2].

One striking and important consequence of our perspective is that KPZ appears to hold in a much broader context than the original CFT realm which relates γ to c , i.e., for *any fractal structure as measured with the quantum random measure* $e^{\gamma h(z)} dz$, and for any $0 \leq \gamma < 2$. For instance, it predicts that the set of Euclidean exponents x of a random or a self-avoiding walk (a $c = 0$ CFT) obeys (4) with $\gamma = \sqrt{8/3}$ in pure gravity ($c = 0$), but also with $\gamma = \sqrt{3}$ on a random lattice equilibrated with Ising spins ($c = 1/2$). This *central charge mixing* yields new KPZ exponents Δ_γ 's in (5) or (7), settling theoretically an issue raised earlier but inconclusively in numerical simulations [122, 123].

Note that Eq. (6) gives only values of parameter γ in the range $\gamma \leq 2$. Our probabilistic approach also allows us to explain the *duality* property of Liouville quantum gravity: for $\gamma > 2$, the *singular* quantum measure can be properly defined [2] in terms of the *regular* γ' -quantum measure, for the dual value $\gamma' = 4/\gamma < 2$, establishing the existence of the so-called “other branch” of the γ -KPZ relation and its correspondence to standard γ' -KPZ for $\gamma' < 2$, as argued long ago in Refs. [18, 19, 47].

1.2. Quantum measure

For concreteness, let h be an instance of a centered GFF on a bounded simply connected domain D with zero boundary conditions. This means that $h = \sum_n \alpha_n f_n$ where the α_n are i.i.d. zero mean unit variance normal random variables and the f_n are an orthonormal basis, with respect to the Dirichlet inner product

$$(f_1, f_2)_\nabla := \frac{1}{2\pi} \int_D \nabla f_1(z) \cdot \nabla f_2(z) dz,$$

of the Hilbert space closure $H(D)$ of the space $H_s(D)$ of C^∞ real-valued functions compactly supported on D . Although this sum diverges pointwise almost surely, it does converge almost surely in the space of distributions on D , and one can also make sense of the mean value of h on various sets. (See Ref. [124] for a detailed account of this construction of the GFF.)

Given an instance h of the Gaussian free field on D , let now $h_\varepsilon(z)$ denote the mean value of h on the circle of radius ε centered at z (where $h(z)$ is defined to be zero for $z \in \mathbb{C} \setminus D$). As we shall see in Section 3, this provides a convenient regularization of the Liouville quantum measure, defined as

$$d\mu_{\gamma,\varepsilon} := \varepsilon^{\gamma^2/2} e^{\gamma h_\varepsilon(z)} dz, \quad (8)$$

where the extra power of ε is required for the limit measure to exist when $\varepsilon \rightarrow 0$. We then have the

Proposition 1.1. *Fix $\gamma \in [0, 2)$ and define h and D as above. Then it is (almost surely) the case that as $\varepsilon \rightarrow 0$ (along powers of two), the measures*

$d\mu_{\gamma,\varepsilon} = \varepsilon^{\gamma^2/2} e^{\gamma h_\varepsilon(z)} dz$ converge weakly to a limiting measure, which we denote by $d\mu_\gamma := e^{\gamma h(z)} dz$.

This proposition, rigorously proven in Ref. [1], mathematically defines Liouville quantum gravity.

For each $z \in D$, denote now by $C(z; D)$ the *conformal radius* of D viewed from z . That is, $C(z; D) = |\phi'(z)|^{-1}$ where $\phi : D \rightarrow \mathbb{D}$ is the conformal map to the unit disc with $\phi(z) = 0$. A simple geometrical analysis of the GFF properties (see Section 3) gives the variance of the circular average $h_\varepsilon(z)$:

$$\text{Var} h_\varepsilon(z) = -\log \varepsilon + \log C(z; D).$$

This gives the following equivalent definition of the (regularized) quantum measure $d\mu_{\gamma,\varepsilon}$ (8):

$$d\mu_{\gamma,\varepsilon} = \exp \left(\gamma h_\varepsilon(z) - \frac{\gamma^2}{2} \text{Var} h_\varepsilon(z) + \frac{\gamma^2}{2} \log C(z; D) \right) dz, \quad (9)$$

and the continuum quantum measure $d\mu_\gamma$ is the (weak) limit for $\varepsilon \rightarrow 0$ of these measures.

A standard property of the expectation of the exponential of a Gaussian random variable Y is

$$\mathbb{E} \exp Y = \exp \left(\mathbb{E} Y + \frac{1}{2} \text{Var} Y \right), \quad (10)$$

with here, $Y := h_\varepsilon(z)$, and $\mathbb{E} Y = \mathbb{E} h_\varepsilon(z) = 0$, from the zero boundary conditions. The *expectation* of the quantum measure (9) reads therefore

$$\mathbb{E} \mu_{\gamma,\varepsilon}(A) = \int_A C(z; D)^{\frac{\gamma^2}{2}} dz =: \mathbb{E} \mu_\gamma(A),$$

independently of ε , and for each measurable subset $A \subset D$.

Intuitively, we interpret the pair (D, μ_γ) as describing a “random surface” $\mathcal{M}$ parameterized conformally by D , with area measure given by μ_γ . The term “random metric” is often used as well; however, we stress that, since we have not endowed D with a two point distance function, “random metric” in the Liouville quantum gravity context does not mean “random metric space,” but “random measure”. The *fluctuations* of this random measure are at the heart of the KPZ relation to which we turn now.

1.3. Euclidean and quantum scaling exponents

Definition 1.1. For any fixed measure μ_γ on D (which we call the “quantum” measure), we let $B^\delta(z)$ be the Euclidean ball centered at z whose radius is chosen so that $\mu_\gamma(B^\delta(z)) = \delta$. (If there does not exist a unique δ with this property, take the radius to be $\sup\{\varepsilon : \mu_\gamma(B_\varepsilon(z)) \leq \delta\}$.) We refer to $B^\delta(z)$ as the **quantum ball** of area δ centered at z . In particular, if $\gamma = 0$ then μ_0 is Lebesgue measure and $B^\delta(z)$ is $B_\varepsilon(z)$ where $\delta = \pi\varepsilon^2$.

Given a subset $X \subset D$, we denote the ε neighborhood of X by

$$B_\varepsilon(X) = \{z : B_\varepsilon(z) \cap X \neq \emptyset\}.$$

We also define the **quantum δ neighborhood** of X by

$$B^\delta(X) = \{z : B^\delta(z) \cap X \neq \emptyset\}.$$

Translated into probability language, the **KPZ formula** is a quadratic relationship between the expectation fractal dimension of a random subset of D defined in terms of Euclidean measure (which is the Liouville gravity measure with $\gamma = 0$) and the corresponding expectation fractal dimension of X defined in terms of Liouville gravity with $\gamma \neq 0$.

We say that a (deterministic or random) fractal subset X of D has **Euclidean expectation dimension** $2 - 2x$ and **Euclidean scaling exponent** x if the expected area of $B_\varepsilon(X)$ decays like $\varepsilon^{2x} = (\varepsilon^2)^x$, i.e.,

$$\lim_{\varepsilon \rightarrow 0} \frac{\log \mathbb{E} \mu_0(B_\varepsilon(X))}{\log \varepsilon^2} = x.$$

In other words, if z is chosen uniformly in D , and independently of X , then the probability of intersection of the Euclidean ball $B_\varepsilon(z)$ and of X (as illustrated in Figure 4) scales as $\mathbb{P}\{B_\varepsilon(z) \cap X \neq \emptyset\} \asymp \varepsilon^{2x}$, in the sense that

$$\lim_{\varepsilon \rightarrow 0} \frac{\log \mathbb{P}\{B_\varepsilon(z) \cap X \neq \emptyset\}}{\log \varepsilon} = 2x.$$

Fix $\gamma \in [0, 2)$. We say that X has **quantum scaling exponent** Δ if when X and μ_γ (as defined above) are chosen independently we have

$$\lim_{\delta \rightarrow 0} \frac{\log \mathbb{E} \mu_\gamma(B^\delta(X))}{\log \delta} = \Delta.$$

For z chosen uniformly in D , and independently of X and μ_γ , the joint probability of intersection of the quantum ball $B^\delta(z)$ and of X then scales as

$$\mathbb{P}\{B^\delta(z) \cap X \neq \emptyset\} \asymp \delta^\Delta,$$

in the sense that

$$\lim_{\delta \rightarrow 0} \frac{\log \mathbb{P}\{B^\delta(z) \cap X \neq \emptyset\}}{\log \delta} = \Delta.$$

1.4. Box formulation of Liouville quantum gravity

One can alternatively define quantum scaling dimensions using boxes instead of balls. Define a **diadic square** to be a closed square (including its interior) of one of the grids $2^{-k}\mathbb{Z}^2$ for some integer k . Let μ be any measure on $\mathbb{C}$. For $\delta > 0$, we define a (μ, δ) **box** S to be a diadic square S with $\mu(S) < \delta$ and $\mu(S') \geq \delta$ where S' is the diadic parent of S . Clearly, if a point $z \in \mathbb{C}$ does not lie on a boundary of a diadic square—and it satisfies $\mu(\{z\}) < \delta < \mu(\mathbb{C})$ —then there is a unique (μ, δ)

SAW in half plane - 1,000,000 steps

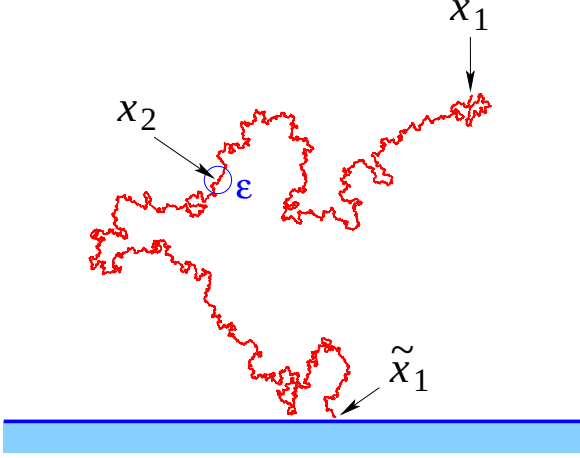


Fig. 4. Fractal path X (here a self-avoiding walk) intersecting a Euclidean ball B_ε of radius ε . The intersection probability scales as $\mathbb{P}\{B_\varepsilon(z) \cap X \neq \emptyset\} \asymp \varepsilon^{2x}$, with a fractal exponent $x = x_2$ for typical intersections in the bulk of path X , and $x = x_1$ at the tip of X . A similar boundary exponent $\tilde{x}_1$ corresponds to a path extremity anchored at a Dirichlet boundary. [Courtesy of T. Kennedy (University of Arizona).]

box containing z , which we denote by $S^\delta(z)$. Let $\mathcal{S}_\mu^\delta$ be the set of all (μ, δ) boxes. The boxes in $\mathcal{S}_\mu^\delta$ do not overlap one another except at their boundaries. Thus, they form a tiling of $\mathbb{R}^2$ (see Figures 5 and 6 for an illustration of this construction for the Liouville quantum gravity measure on a discrete torus).

When ε is a power of 2, analogously define $S_\varepsilon(z)$ to be the dyadic square containing z with edge length ε , i.e., Euclidean area ε^2 . Likewise, define

$$S_\varepsilon(X) = \{z : S_\varepsilon(z) \cap X \neq \emptyset\},$$

$$S^\delta(X) = \{z : S^\delta(z) \cap X \neq \emptyset\},$$

i.e., respectively as the unions of (μ_0, ε^2) **Euclidean** boxes or (μ_γ, δ) **quantum** boxes intersected by X (Fig. 7). The following gives the equivalence of the scaling dimension definition when boxes are used instead of balls.

Proposition 1.2. Fix $\gamma \in [0, 2)$ and let X be a random subset of a deterministic compact subset of D . Let $N_\gamma(\delta, X)$ be the number of (μ_γ, δ) boxes intersected by X and $N_0(\varepsilon^2, X)$ the number of dyadic squares intersecting X that have edge length ε (a power of 2). Then X has Euclidean scaling exponent $x \geq 0$ if and only if

$$\lim_{\varepsilon \rightarrow 0} \frac{\log \mathbb{E}[\mu_0(S_\varepsilon(X))]}{\log \varepsilon^2} = \lim_{\varepsilon \rightarrow 0} \frac{\log \mathbb{E}[\varepsilon^2 N_0(\varepsilon^2, X)]}{\log \varepsilon^2} = x,$$

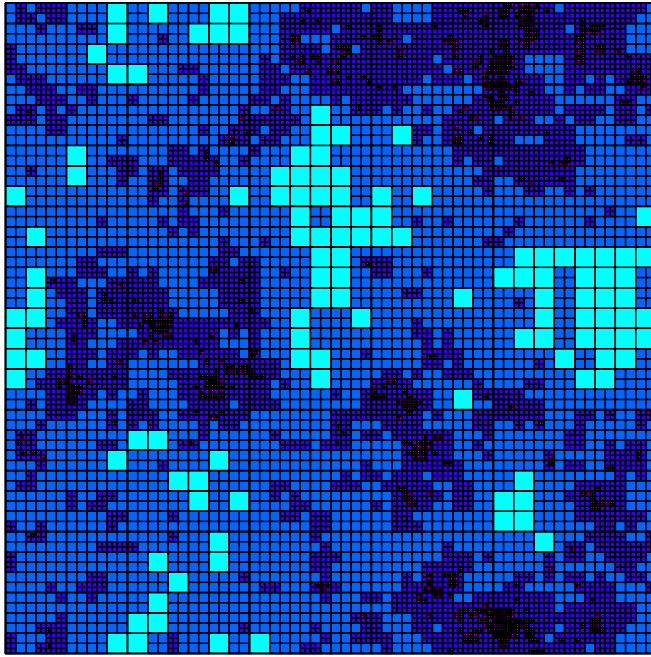


Fig. 5. (μ_γ, δ) boxes of the random measure $\mu_\gamma = e^{\gamma h} dz$, where $\gamma = 1$ and h is the (discrete) Gaussian free field on a very fine (1024×1024) grid on the torus, dz is counting measure on the vertices of that grid, and δ is 2^{-12} times the total mass of μ_γ . (We view μ_γ as a discrete approximation of the continuum Liouville quantum gravity measure.) One way to construct this figure is to view the entire torus as a square; then subdivide each square whose μ_γ measure is at least δ into four smaller squares, and repeat until all squares have μ_γ measure less than δ . The squares shown have roughly the same μ_γ size — in the sense that each square has μ_γ measure less than δ but each square's diadic parent has μ_γ measure greater than δ .

or equivalently,

$$\lim_{\varepsilon \rightarrow 0} \frac{\log \mathbb{E}[N_0(\varepsilon^2, X)]}{\log \varepsilon^2} = x - 1.$$

Similarly, X has **quantum scaling exponent** Δ if and only if when X and μ_γ (as defined above) are chosen independently we have

$$\lim_{\delta \rightarrow 0} \frac{\log \mathbb{E}[\mu_\gamma(S^\delta(X))]}{\log \delta} = \lim_{\delta \rightarrow 0} \frac{\log \mathbb{E}[\delta N_\gamma(\delta, X)]}{\log \delta} = \Delta, \quad (11)$$

or equivalently,

$$\lim_{\delta \rightarrow 0} \frac{\log \mathbb{E}[N_\gamma(\delta, X)]}{\log \delta} = \Delta - 1. \quad (12)$$

Equivalently, for z chosen uniformly in D , and independently of X and μ_γ , the joint probability of intersection of the quantum square $S^\delta(z)$ and of X scales as

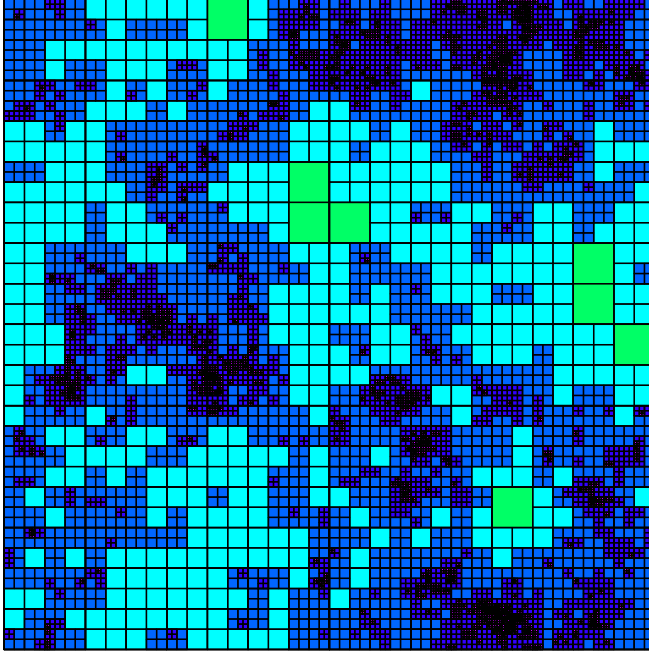


Fig. 6. Analog of Figure 5 with $\gamma = 3/2$, using the same instance h of the GFF.

$\mathbb{P}\{S^\delta(z) \cap X \neq \emptyset\} \asymp \delta^\Delta$, in the sense that

$$\lim_{\delta \rightarrow 0} \frac{\log \mathbb{P}\{S^\delta(z) \cap X \neq \emptyset\}}{\log \delta} = \Delta.$$

1.5. Statement of KPZ

The following is the KPZ scaling exponent relation [1]. To avoid boundary technicalities, we restrict attention here to a compact subset of D .

Theorem 1.1. *Fix $\gamma \in [0, 2)$ and a compact subset $\tilde{D}$ of D . If $X \cap \tilde{D}$ has Euclidean scaling exponent $x \geq 0$ then it has quantum scaling exponent $\Delta = \Delta_\gamma$, where Δ_γ is the non-negative solution (5) to*

$$x = \frac{\gamma^2}{4} \Delta_\gamma^2 + \left(1 - \frac{\gamma^2}{4}\right) \Delta_\gamma. \quad (13)$$

Note that expectation in the above is with respect to both random variables, X and μ_γ .

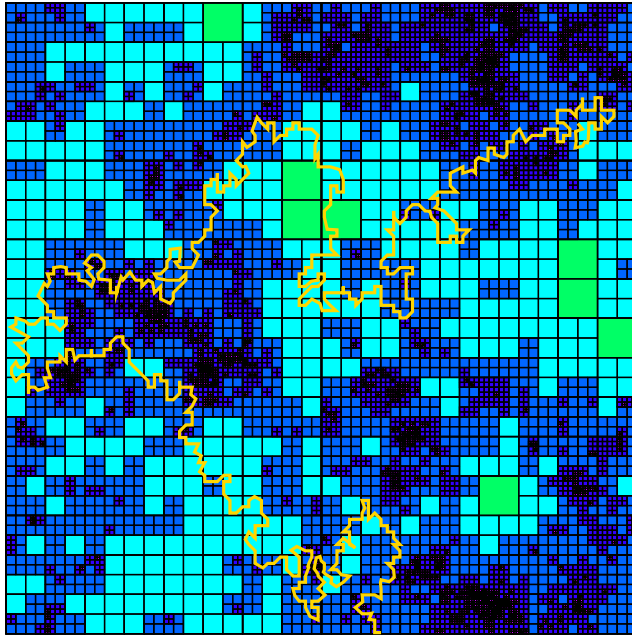


Fig. 7. Fractal set X (here a self-avoiding walk) intersecting the quantum grid $S_{\mu_\gamma}^\delta$ made of (μ_γ, δ) boxes for $\gamma = 3/2$ (Fig. 6). (Recall also Fig. 3.) The number $N_\gamma(\delta, X)$ of (μ_γ, δ) boxes intersected by X scales as $N_\gamma(\delta, X) \asymp \delta^{\Delta-1}$, where Δ is the quantum scaling exponent of X . Equivalently, the joint probability of intersection of a dyadic quantum box $S^\delta(z)$ and X scales as $\mathbb{P}\{S^\delta(z) \cap X \neq \emptyset\} \asymp \delta^\Delta$.

2. GFF regularization

2.1. GFF circular average

In this section, we present a mathematically convenient regularization of the Gaussian free field in the continuum [1], hence of the Liouville measure, which addresses a question left opened by Polyakov in 1987 [5].^c

Let h be a centered Gaussian free field on a bounded simply connected domain D with Dirichlet zero boundary conditions. Special care is required to make sense of the quantum gravity measure, since the GFF is a distribution and not a function (it typically oscillates between $\pm\infty$) [see e.g., Ref. 124 and Figure 8].

For each $z \in D$, define the ball $B_\varepsilon(z) := \{w : |w - z| < \varepsilon\}$, and let ε_0^z be the largest ε such that $B_\varepsilon(z) \subset D$ (Fig. 9). When $\varepsilon \leq \varepsilon_0^z$, write $h_\varepsilon(z)$ for the average value of h on the circle $\partial B_\varepsilon(z) := \{w : |w - z| = \varepsilon\}$. (Similar averages were

^cThe latter writes in Ref. 5: “The most simple form this formula takes is in the conformal gauge, where $g_{ab} = e^\varphi \delta_{ab}$ where it becomes a free field action. Unfortunately this simplicity is an illusion. We have to set a cut-off in quantizing this theory, such that it is compatible with general covariance. Generally, it is not clear how to do this. For that reason, we take a different approach.”

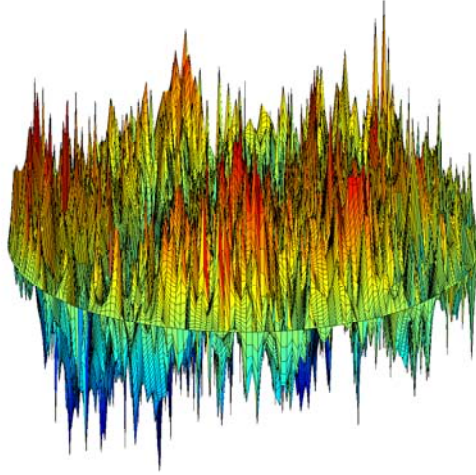


Fig. 8. GFF with Dirichlet boundary conditions on a disk. A very fine regularization leads to a proliferation of spikes. [Courtesy of N.-G. Kang (Caltech).]

considered in Ref. 125.) Denote by $\rho_\varepsilon^z(y)$ the uniform (Dirac-like) density (of total mass one) localized on the circle $\partial B_\varepsilon(z)$,

$$\rho_\varepsilon^z(y) := \frac{1}{2\pi} \int_0^{2\pi} d\theta \delta(y - z - \varepsilon e^{i\theta}), \quad (14)$$

such that one can write $h_\varepsilon(z)$ as the scalar product on D

$$h_\varepsilon(z) = (h, \rho_\varepsilon^z) := \int_D h(y) \rho_\varepsilon^z(y) dy. \quad (15)$$

To the density ρ_ε^z is naturally associated a Newtonian potential f_ε^z . We define the function $f_\varepsilon^z(y)$, for $y \in D$,

$$f_\varepsilon^z(y) := -\log(\varepsilon \vee |y - z|) - \tilde{G}_z(y), \quad (16)$$

where $\tilde{G}_z(y)$ is the *harmonic extension* of $-\log|y - z|$ to D , i.e., the harmonic function of $y \in D$ with boundary value equal to the restriction of $-\log|y - z|$ to $y \in \partial D$. By construction this $f_\varepsilon^z(y)$ satisfies Dirichlet boundary conditions and the Poisson equation

$$-\Delta f_\varepsilon^z = 2\pi \rho_\varepsilon^z. \quad (17)$$

This (regular) potential function is represented in Fig. 10.

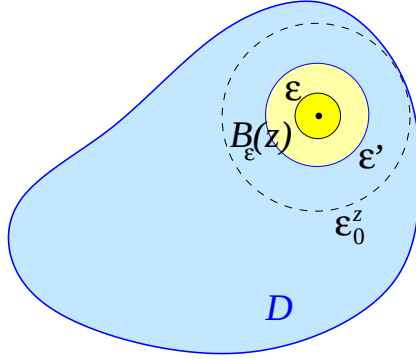


Fig. 9. Nested balls $B_\varepsilon(z)$ centered at z . The value ε_0^z is the largest radius ε such that $B_\varepsilon(z) \subset D$. A uniform (Dirac-like) density (of total mass one) is localized on the circle $\partial B_\varepsilon(z)$.

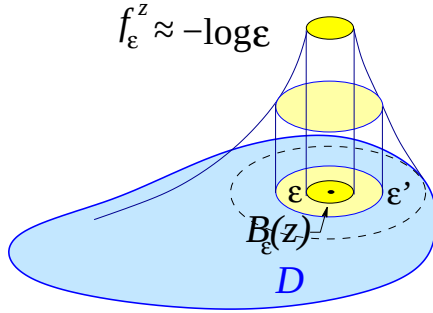


Fig. 10. Potential f_ε^z (16) created by a uniform mass distribution localized on the circle $\partial B_\varepsilon(z)$; its leading constant value inside the disc $B_\varepsilon(z)$ is $-\log \varepsilon$.

Substituting (17) into (15), integrating by parts, and using Dirichlet boundary conditions for h , we immediately have the following:

$$h_\varepsilon(z) = (h, \rho_\varepsilon^z) = \frac{1}{2\pi} (h, -\Delta f_\varepsilon^z) = \frac{1}{2\pi} \int_D \nabla h(y) \cdot \nabla f_\varepsilon^z(y) dy,$$

which we write

$$h_\varepsilon(z) = (h, f_\varepsilon^z)_\nabla, \quad (18)$$

in terms of the Dirichlet inner product

$$(f_1, f_2)_\nabla := \frac{1}{2\pi} \int_D \nabla f_1(y) \cdot \nabla f_2(y) dy, \quad (19)$$

i.e., the *interaction energy of gradient fields associated with potentials* $f_{i=1,2}$.

2.2. GFF covariance

The random variables $(h, f)_\nabla$ are zero mean Gaussian random variables for each f , with the fundamental covariance property:

$$\text{Cov}((h, f_1)_\nabla, (h, f_2)_\nabla) = (f_1, f_2)_\nabla. \quad (20)$$

From this and (18), one obtains the covariance of the averaged $h_\varepsilon(z)$ fields on two

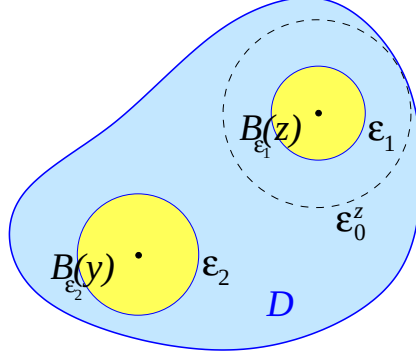


Fig. 11. Balls $B_{\varepsilon_1}(z)$ and $B_{\varepsilon_2}(y)$ with unit masses localized on their respective boundaries $\partial B_{\varepsilon_1}(z)$ and $\partial B_{\varepsilon_2}(y)$.

arbitrary circles (Fig. 11):

$$\text{Cov}(h_{\varepsilon_1}(z), h_{\varepsilon_2}(y)) = (f_{\varepsilon_1}^z, f_{\varepsilon_2}^y)_\nabla,$$

which is precisely the *Newtonian interaction energy* of the unit mass distributions localized on the two circles $\partial B_{\varepsilon_1}(z)$ and $\partial B_{\varepsilon_2}(y)$.

2.3. GFF circular average and Brownian motion

Consider *nested circles* centered at z (Fig. 9). We obtain the covariance of the averaged $h_\varepsilon(z)$ fields (18) on those circles

$$\text{Cov}(h_\varepsilon(z), h_{\varepsilon'}(z)) = (f_\varepsilon^z, f_{\varepsilon'}^z)_\nabla,$$

an interaction energy which is easily evaluated by using Gauss' theorem. By integration by parts, and use of (17) with the explicit potential (16), the Newtonian interaction energy of two nested circles can be written as

$$\begin{aligned} (f_\varepsilon^z, f_{\varepsilon'}^z)_\nabla &= (2\pi)^{-1} (f_\varepsilon^z, -\Delta f_{\varepsilon'}^z) = (f_\varepsilon^z, \rho_{\varepsilon'}^z) \\ &= f_{\varepsilon \vee \varepsilon'}^z(z) = -\log(\varepsilon \vee \varepsilon') - \tilde{G}_z(z), \end{aligned} \quad (21)$$

where we have used the fact that the function $\tilde{G}_z$ is harmonic, hence $(\tilde{G}_z, \rho_{\varepsilon'}^z) = \tilde{G}_z(z)$. This gives the covariances

$$\text{Cov}(h_\varepsilon(z), h_{\varepsilon'}(z)) = \mathbb{E}(h_\varepsilon(z)h_{\varepsilon'}(z)) = -\log(\varepsilon \vee \varepsilon') - \tilde{G}_z(z), \quad (22)$$

with here $\mathbb{E} h_\varepsilon(z) = 0$ for Dirichlet boundary conditions.

Specifying the above results to $\varepsilon' = \varepsilon$, we get in particular the *variance* of field $h_\varepsilon(z)$

$$\text{Var } h_\varepsilon(z) = (f_\varepsilon^z, f_\varepsilon^z)_\nabla = f_\varepsilon^z(z) = -\log \varepsilon - \tilde{G}_z(z). \quad (23)$$

The harmonic extension $\tilde{G}_z(y)$, taken at point $y = z$, is $\tilde{G}_z(z) = -\log C(z; D)$, in terms of the conformal radius C of D viewed from z , a smooth function of z .

2.3.1. Brownian motion

From (22), and (23) we thus get the two important variances

$$\text{Var } h_\varepsilon(z) = -\log \varepsilon + \log C(z; D), \quad (24)$$

$$\text{Var}[h_\varepsilon(z) - h_{\varepsilon'}(z)] = |\log \varepsilon - \log \varepsilon'|. \quad (25)$$

Note that (24) mathematically quantifies the unboundedness of the GFF fluctuations in the continuum as $\varepsilon \rightarrow 0$, as illustrated in Fig. 8. Moreover, Equation (25) shows that $h_\varepsilon(z)$ does not converge to a limit as $\varepsilon \rightarrow 0$ (which would be $h(z)$ if h were a proper continuous function, and not a distribution), but that it indefinitely wanders like *Brownian motion* in time parameter $t = -\log \varepsilon$. Indeed, define the reference radius $\varepsilon_0 := \varepsilon_0^z$. The interpretation of (24) and (25) is immediate: For fixed z , the Gaussian random variable $h_\varepsilon(z) - h_{\varepsilon_0}(z)$ is *one-dimensional standard Brownian motion* $\mathcal{B}_t$ when parameterized by time $t := -\log(\varepsilon/\varepsilon_0)$:

$$h_\varepsilon(z) - h_{\varepsilon_0}(z) \stackrel{(\text{in law})}{=} \mathcal{B}_t \quad (26)$$

$$-\log(\varepsilon/\varepsilon_0) = t \in [0, \infty), \quad \varepsilon \leq \varepsilon_0 \quad (27)$$

$$\mathcal{B}_0 = 0. \quad (28)$$

This Brownian property of the circular average $h_\varepsilon(z)$ is the key to the KPZ relation [1].

3. Random measure and Liouville quantum gravity

3.1. Gaussian exponential averages

Applying Eq. (10) to the Gaussian variable $h_\varepsilon(z)$, and since $\mathbb{E} h_\varepsilon(z) = 0$, we have from (24) the exponential expectation:

$$\mathbb{E} e^{\gamma h_\varepsilon(z)} = e^{\text{Var}[\gamma h_\varepsilon(z)]/2} = [C(z; D)/\varepsilon]^{\gamma^2/2}. \quad (29)$$

Since (29) ultimately *diverges* for $\varepsilon \rightarrow 0$, we are led to *regularize* Liouville quantum gravity by defining the random measure, as in (8) and (9)

$$d\mu_{\gamma, \varepsilon} = M_{\gamma, \varepsilon}(z) dz, \quad M_{\gamma, \varepsilon}(z) := \varepsilon^{\gamma^2/2} e^{\gamma h_\varepsilon(z)}, \quad (30)$$

in a way similar to the so-called Wick normal ordering (see e.g., Ref. 126). In Ref. 1, it is shown that for $\gamma \in [0, 2)$ the limit $d\mu_\gamma$ of this regularized measure exists as $\varepsilon \rightarrow 0$, which mathematically defines Liouville quantum gravity (see also Refs. [127–129] for earlier work).

3.2. Liouville weighted GFF measure

We now consider a measure on pairs (z, h) where h is the Gaussian free field, and given h the point z is chosen from the regularized quantum area measure $e^{\gamma h_\varepsilon(z)} dz$. Such a measure has the form

$$e^{\gamma h_\varepsilon(z)} dh dz = \exp \left[-\frac{1}{2}(h, h)_\nabla + \gamma h_\varepsilon(z) \right] \mathcal{D}h dz, \quad (31)$$

where dh represents the whole GFF measure and $\mathcal{D}h$ the flat functional measure. The total action of the GFF *Liouville weighted measure* is thus the quadratic combination

$$S_\gamma(h) := \frac{1}{2}(h, h)_\nabla - \gamma h_\varepsilon(z). \quad (32)$$

Owing to (18), the latter can be rewritten as

$$\begin{aligned} S_\gamma(h) &= \frac{1}{2}(h, h)_\nabla - \gamma(h, f_\varepsilon^z)_\nabla = \frac{1}{2}(h - \gamma f_\varepsilon^z, h - \gamma f_\varepsilon^z)_\nabla - \frac{\gamma^2}{2}(f_\varepsilon^z, f_\varepsilon^z)_\nabla \\ &= \frac{1}{2}(\underline{h}, \underline{h})_\nabla - \frac{\gamma^2}{2} \text{Var } h_\varepsilon(z), \quad \underline{h} := h - \gamma f_\varepsilon^z, \end{aligned} \quad (33)$$

where we used (23). In the substitution $\underline{h} := h - \gamma f_\varepsilon^z$, $\underline{h}$ is *standard* GFF. The probability weight involved in our random measure (31) and (32) can finally be written, thanks to (29), as

$$\begin{aligned} \exp[-S_\gamma(h)] &= \exp \left[-\frac{1}{2}(h, h)_\nabla + \gamma h_\varepsilon(z) \right] \\ &= \exp \left[-\frac{1}{2}(\underline{h}, \underline{h})_\nabla \right] \mathbb{E} e^{\gamma h_\varepsilon(z)}, \end{aligned} \quad (34)$$

where the second factor on the right-hand side is the *marginal distribution* of z . The meaning of (34) is that, after sampling z from its marginal distribution, the law of h *weighted* by $e^{\gamma h_\varepsilon(z)}$ is identical to that of the original GFF $\underline{h}$ *plus* the deterministic function γf_ε^z (16):

$$h \stackrel{(\text{in law})}{=} \underline{h} + \gamma f_\varepsilon^z. \quad (35)$$

3.3. Random measure and Brownian motion with drift

For fixed z , the identity in law (35) can be averaged on the circle $\partial B_\varepsilon(z)$ with the density ρ_ε^z (14) to yield

$$\begin{aligned} h_\varepsilon(z) &\stackrel{(\text{in law})}{=} \underline{h}_\varepsilon(z) + \gamma(f_\varepsilon^z, \rho_\varepsilon^z) \\ &= \underline{h}_\varepsilon(z) + \gamma f_\varepsilon^z(z) \\ &= \underline{h}_\varepsilon(z) - \gamma \log[\varepsilon/C(z; D)], \end{aligned} \quad (36)$$

where use was made of result (21). Since $\underline{h}$ is standard Gaussian free field, one can apply to its circular average $\underline{h}_\varepsilon(z)$ the identity in law (26) to standard Brownian

motion $\mathcal{B}_t$, $t = -\log(\varepsilon/\varepsilon_0)$, and rewrite (36) as

$$\begin{aligned} h_\varepsilon(z) - h_{\varepsilon_0}(z) &\stackrel{(\text{in law})}{=} \underline{h}_\varepsilon(z) - \underline{h}_{\varepsilon_0}(z) - \gamma \log(\varepsilon/\varepsilon_0) \\ &\stackrel{(\text{in law})}{=} \mathcal{B}_t + \gamma t, \quad \mathcal{B}_0 = 0. \end{aligned} \quad (37)$$

The statement of (37) is the following [1]

Proposition 3.1. *For fixed z , when the law of h is weighted by the Liouville conformal factor $e^{\gamma h_\varepsilon(z)}$ as in (31) or (34), the law of Gaussian random variable $h_\varepsilon(z) - h_{\varepsilon_0}(z)$ is identical to that of one-dimensional standard Brownian motion $\mathcal{B}_t$ with drift γt , when parameterized by time $t = -\log(\varepsilon/\varepsilon_0)$.*

4. KPZ Proof

4.1. Quantum balls

It can be shown [1] that when ε is small, the regularized stochastic quantum measure $\mu_{\gamma,\varepsilon}(B_\varepsilon(z))$ (8) or (30) of the Euclidean ball $B_\varepsilon(z)$, as well as the limit quantum measure $\mu_\gamma(B_\varepsilon(z))$ of proposition (1.1), are very well approximated by the simple form $\mu_{\gamma\odot}(B_\varepsilon(z)) := \pi\varepsilon^2\varepsilon^{\gamma^2/2}e^{\gamma h_\varepsilon(z)}$, i.e., Lebesgue measure of the ball times the regularized conformal factor at the ball's center z . Let us then write

$$\mu_{\gamma\odot}(B_\varepsilon(z)) = \pi \exp[\gamma h_\varepsilon(z) + \gamma Q_\gamma \log \varepsilon], \quad (38)$$

$$Q_\gamma := 2/\gamma + \gamma/2, \quad (39)$$

and in this simplified perspective, we view $\mu_{\gamma\odot}$ as a function on balls of the form $B_\varepsilon(z)$, defined by (38), rather than the fully defined measure on D . It proves convenient to consider the normalized quantity

$$\mu_{\gamma\odot}(B_\varepsilon(z))/\mu_{\gamma\odot}(B_{\varepsilon_0}(z)) = \exp[\gamma(h_\varepsilon(z) - h_{\varepsilon_0}(z)) + \gamma Q_\gamma \log(\varepsilon/\varepsilon_0)]. \quad (40)$$

In accordance with definition 1.1 in Section 1.3, define then the *quantum ball* $\tilde{B}^\delta(z)$ of area δ centered at z as the (largest) Euclidean ball $B_\varepsilon(z)$ whose radius ε is chosen so that (40) equals δ

$$\mu_{\gamma\odot}(\tilde{B}^\delta(z))/\mu_{\gamma\odot}(B_{\varepsilon_0}(z)) = \delta. \quad (41)$$

4.2. Quantum scaling

Recall that a (deterministic or random) fractal subset X of D has *Euclidean scaling exponent* x (and Euclidean dimension $2 - 2x$) if, for z chosen uniformly in D and independently of X , the probability

$$\mathbb{P}\{B_\varepsilon(z) \cap X \neq \emptyset\} \asymp \varepsilon^{2x}, \quad (42)$$

in the sense that $\lim_{\varepsilon \rightarrow 0} \log \mathbb{P} / \log \varepsilon = 2x$.

In analogy to the definitions given in Section 1.3, X has *quantum scaling exponent* Δ if when X and (z, h) , sampled with weight (31) (34), are chosen independently we have

$$\mathbb{P}\{\tilde{B}^\delta(z) \cap X \neq \emptyset\} \asymp \delta^\Delta. \quad (43)$$

4.3. Brownian stopping time

According to proposition 3.1, $h_\varepsilon(z) - h_{\varepsilon_0}(z)$, when sampled with (31) and (34), has the same law (37) as Brownian motion with drift. Substituting $t = -\log(\varepsilon/\varepsilon_0)$ we can rewrite (40) under the form

$$\mu_{\gamma\odot}(B_\varepsilon(z))/\mu_{\gamma\odot}(B_{\varepsilon_0}(z)) \stackrel{(\text{in law})}{=} \exp[\gamma\mathcal{B}_t + \gamma(\gamma - Q_\gamma)t]. \quad (44)$$

Equality of (40) to (41) then relates *stochastically* the Euclidean radius ε to the given quantum area δ , and, thanks to (44), characterizes the (smallest) time $t = -\log(\varepsilon/\varepsilon_0)$ such that

$$\exp[\gamma(\mathcal{B}_t - a_\gamma t)] = \delta, \quad (45)$$

where we introduced the parameter

$$a_\gamma := Q_\gamma - \gamma = 2/\gamma - \gamma/2 > 0; \quad \gamma \in (0, 2). \quad (46)$$

The Euclidean radius is thus given in terms of the *stopping time* (Fig. 12)

$$-\log(\varepsilon_A/\varepsilon_0) := T_A := \inf\{t : -\mathcal{B}_t + a_\gamma t = A\} \quad (47)$$

$$A := -(\log \delta)/\gamma > 0. \quad (48)$$

At time $t = 0$, Brownian motion $\mathcal{B}_t$ is started at $\mathcal{B}_0 = 0$, in accordance with (37).

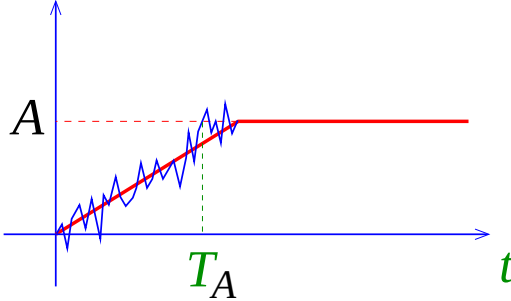


Fig. 12. Illustration of definition (47) of stopping time T_A . The drift term $a_\gamma t$ gives a constant positive slope, on which Brownian motion $\mathcal{B}_t$ is superimposed.

The probability (42) that the Euclidean ball $B_{\varepsilon_A}(z)$ intersects X scales as $\varepsilon_A^{2x} = \varepsilon_0^{2x} e^{-2xT_A}$. Computing the expectation

$$\mathbb{E} \varepsilon_A^{2x} = \varepsilon_0^{2x} \mathbb{E} [\exp(-2xT_A)] \quad (49)$$

with respect to the random stopping time T_A will give the quantum probability (43)

$$\mathbb{P}\{\tilde{B}^\delta(z) \cap X \neq \emptyset\} \asymp \mathbb{E} [\exp(-2xT_A)], \quad (50)$$

in the sense that the ratio of the logarithms of these quantities tends to 1 as $\delta \rightarrow 0$, $A \rightarrow \infty$.

4.4. Brownian martingale

Consider then for any β the standard *Brownian exponential martingale*

$$\mathbb{E} [\exp(-\beta \mathcal{B}_t - \beta^2 t/2)] = \mathbb{E} [\exp(-\beta \mathcal{B}_0)] = 1, \quad (51)$$

valid for $0 \leq t < \infty$ and $\mathcal{B}_0 = 0$.

When $0 \leq t \leq T_A$, we have $-\mathcal{B}_t + a_\gamma t - A \leq 0$ by definition of T_A , so the argument of martingale (51) stays bounded, for $\beta > 0$, by $\exp(\beta A - (\beta a_\gamma + \beta^2/2)t) \leq \exp(\beta A)$, hence by a fixed constant. We can then apply the martingale (51) at the stopping time T_A , when $T_A < \infty$ and where $\mathcal{B}_{T_A} = a_\gamma T_A - A$. We thus get

$$\mathbb{E} [\exp[\beta A - (\beta a_\gamma + \beta^2/2)T_A] 1_{T_A < \infty}] = 1,$$

which can be written as

$$\mathbb{E}[\exp(-2xT_A) 1_{T_A < \infty}] = \exp(-\beta A) \quad (52)$$

$$2x := \beta a_\gamma + \beta^2/2. \quad (53)$$

This is valid for $\beta \geq 0$, which is thus equal to the positive root to (53)

$$\beta = \beta_\gamma(x) := (a_\gamma^2 + 4x)^{1/2} - a_\gamma. \quad (54)$$

Thanks to definition (48) of A , we thus obtain the scaling behavior

$$\mathbb{E}[\exp(-2xT_A) 1_{T_A < \infty}] = \exp(-\beta_\gamma(x)A) = \delta^{\Delta_\gamma} \quad (55)$$

$$\Delta_\gamma := \beta_\gamma(x)/\gamma = [(a_\gamma^2 + 4x)^{1/2} - a_\gamma]/\gamma. \quad (56)$$

For $x = 0$ one finds in particular

$$\mathbb{P}(T_A < \infty) = \mathbb{E}[1_{T_A < \infty}] = 1, \quad (57)$$

since $\beta_\gamma(0) = (|a_\gamma| - a_\gamma)/\gamma = 0$ for $a_\gamma > 0, \gamma \in (0, 2)$, so that the conditioning on $T_A < \infty$ can actually be *omitted* for $\gamma \in (0, 2)$.

We thus obtain with (50) and (55) the expected quantum scaling behavior (43), together with $\Delta = \Delta_\gamma$ (56), which is, thanks to (46),

$$\Delta_\gamma = \gamma^{-1} \left(\sqrt{4x + (2/\gamma - \gamma/2)^2} - (2/\gamma - \gamma/2) \right), \quad (58)$$

i.e., the expected **inverse KPZ** function (5). Note also that substituting $\beta_\gamma = \gamma \Delta_\gamma$ into (53) gives **KPZ** relation (13), **QED**.

It is interesting to note that in (53), which is in essence KPZ, the term linear in β corresponds in the Brownian martingale (52) to the intersection time $T'_A := A/a_\gamma$ of the linear drift term $a_\gamma t$ with level A in Fig. 12. It explains the origin of the term linear in Δ_γ in (13). The second term $\beta^2/2$ in (53) corresponds to pure Brownian fluctuations in martingale (52) and in Fig. 12. These fluctuations thus generate the distinctive quadratic term in KPZ formula (13).

Notice finally that we used the function $\mu_{\gamma\odot}(B_\varepsilon(z))$ (38), together with its Brownian representation (44). The complete proof of KPZ [1] requires using instead the

full Liouville measure $\mu_\gamma(B_\varepsilon(z))$. One defines a stopping time $\overline{T}_A := -\log(\overline{\varepsilon}_A/\varepsilon_0)$, where $\overline{\varepsilon}_A$ is the Euclidean radius such that $\mu_\gamma(B_{\overline{\varepsilon}_A}(z)) = \delta$; one can then show that

$$\lim_{A \rightarrow \infty} \frac{\log \mathbb{E}[\exp(-2x\overline{T}_A)]}{\log \mathbb{E}[\exp(-2xT_A)]} = 1,$$

which completes the proof [1]. $\square$

4.5. Probability distribution and GFF thick points

The inverse Laplace transform $P_A(t)$ of (55), with respect to $2x$, is the probability density of $T_A = -\log(\varepsilon_A/\varepsilon_0)$ such that $P_A(t)dt := \mathbb{P}(T_A \in [t, t+dt])$. Its explicit form is [1]:

$$P_A(t) = \frac{A}{\sqrt{2\pi t^3}} \exp\left[-\frac{1}{2t}(A - a_\gamma t)^2\right]. \quad (59)$$

This probability density is associated with the variations of the Euclidean area

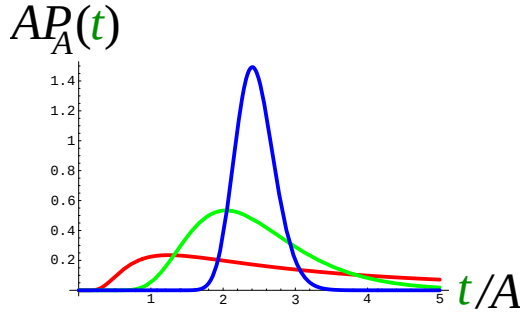


Fig. 13. Probability distribution $P_A(t)$ in reduced units, for $A = 2$ (left), 20 (middle), 200 (right).

$\pi\varepsilon^2 = \pi\varepsilon_0^2 e^{-2t}$ for a given quantum area $\delta := e^{-\gamma A}$.

From (59) one deduces that for A large (i.e., δ and ε small), the variable $\frac{A}{T_A} = \frac{\log \delta}{\gamma \log(\varepsilon_A/\varepsilon_0)}$ is concentrated in (55),

$$\mathbb{E}[\exp(-2xT_A)1_{T_A < \infty}] = \int_0^\infty dt P_A(t) \exp(-2xt), \quad (60)$$

near the saddle point value $A/t := a_\gamma + \gamma\Delta_\gamma$. Reverse engineering to Brownian motion $\mathcal{B}_t$ via (47), one finds that $-\mathcal{B}_t/t = A/t - a_\gamma$ is concentrated near $\gamma\Delta_\gamma$, and finally, using the identity in law (37) that the GFF circle average $h_\varepsilon(z)/\log \varepsilon^{-1} \stackrel{(\text{in law})}{=} \mathcal{B}_t/t + \gamma$ has a typical value $\gamma - \gamma\Delta_\gamma$. One thus concludes [1, 2] that a point z that is typical with respect to the quantum measure is an α -thick point of h (see Ref. 130):

$$\alpha := \lim_{\varepsilon \rightarrow 0} \log h_\varepsilon(z) / \log \varepsilon^{-1},$$

with a typical value

$$\alpha = \gamma - \gamma \Delta_\gamma \quad (61)$$

for a fractal of quantum scaling dimension Δ_γ .

5. Boundary KPZ

5.1. Boundary quantum measure.

Most of the results about random measures on D have straightforward analogs about random measures on the boundary ∂D . The proofs are essentially identical, and we will only sketch the arguments here [1, 2].

Suppose that D is a domain with a piecewise linear boundary ∂D and that h is an instance of the GFF on D , now with *free* boundary conditions, and normalized to have mean zero. This means that $h = \sum_n \alpha_n f_n$ where the α_n are i.i.d. zero mean unit variance normal random variables and the f_n are an orthonormal basis, with respect to the inner product

$$(f_1, f_2)_\nabla = (2\pi)^{-1} \int_D \nabla f_1(z) \cdot \nabla f_2(z) dz,$$

of the Hilbert space closure $H(D)$ of the space of C^∞ bounded real-valued (but *not* necessarily compactly supported) functions on D with mean zero.

We can also make sense of $h_\varepsilon(z)$, for a point $z \in \partial D$ on the boundary of D , to be the mean value of $h(z)$ on the *semicircle* $\partial B_\varepsilon(z) \cap D$ of radius ε centered at z and contained in the domain D . In this case, if $\varepsilon \leq \varepsilon_0$, where ε_0 is small enough so that $B_{\varepsilon_0}(z)$ contains no corners of D and exactly one semi-disc of $B_{\varepsilon_0}(z)$ lies in D , we can consider the geometry of D to be locally equivalent to that of the upper-half plane $\mathbb{H}$.

We then have, for fixed $z \in \partial D$, that the Gaussian variable $h_\varepsilon(z) - h_{\varepsilon_0}(z)$ is a *standard Brownian motion* $\mathcal{B}_{2t}$ in time $2t := -2\log(\varepsilon/\varepsilon_0)$ for $\varepsilon \leq \varepsilon_0$, i.e., the identity in law

$$z \in \partial D, \quad h_\varepsilon(z) - h_{\varepsilon_0}(z) \stackrel{(\text{in law})}{=} \mathcal{B}_{2t} \quad (62)$$

$$-\log(\varepsilon/\varepsilon_0) = t \quad (63)$$

$$0 = \mathcal{B}_0. \quad (64)$$

Thus at a point on the interior of one of the boundary lines of D , the variance of $h_\varepsilon(z)$ scales like $-2\log\varepsilon$ instead of $-\log\varepsilon$. We then define the *regularized boundary quantum measure*

$$d\mu_{\gamma,\varepsilon}^B := \varepsilon^{\gamma^2/4} e^{\gamma h_\varepsilon(z)/2} dz, \quad (65)$$

where in this case dz is Lebesgue measure on the boundary ∂D of D , with the conformal factor $e^{\gamma h_\varepsilon(z)/2}$ needed for integrating a quantum length instead of an area ; as before, the power of ε chosen makes the expectation $\varepsilon^{\gamma^2/4} \mathbb{E}_h e^{\gamma h_\varepsilon(z)/2}$

regular in ε (compare to (30)). In other words, the factor preceding dz in definition (65) is an exponential martingale in time $t = -\log \varepsilon$.

We define the boundary quantum measure $d\mu_\gamma^B$ to be the (weak) limit as $\varepsilon \rightarrow 0$ of the measures $d\mu_{\gamma,\varepsilon}^B$. The existence of this limit when $0 \leq \gamma < 2$ is insured by a complete analog of Proposition 1.1 for the boundary case [1].

5.2. Boundary scaling and KPZ

For $z \in \partial D$ we write $\hat{B}_\varepsilon(z) := B_\varepsilon(z) \cap \partial D$ and we define $\hat{B}^\delta(z)$ to be the (largest) set $\hat{B}_\varepsilon(z)$ whose μ_γ^B measure is δ .

Likewise define

$$\hat{B}_\varepsilon(X) = \{z \in \partial D : \hat{B}_\varepsilon(z) \cap X \neq \emptyset\}$$

and

$$\hat{B}^\delta(X) = \{z \in \partial D : \hat{B}^\delta(z) \cap X \neq \emptyset\}.$$

We say that a (deterministic or random) fractal subset X of the boundary of D has **Euclidean expectation dimension** $1 - \tilde{x}$ and **Euclidean scaling exponent** $\tilde{x}$ in the boundary sense if the expected measure of $\hat{B}_\varepsilon(X)$ decays like $\varepsilon^{\tilde{x}}$, i.e.,

$$\lim_{\varepsilon \rightarrow 0} \frac{\log \mathbb{E} \mu_0(\hat{B}_\varepsilon(X))}{\log \varepsilon} = \tilde{x}.$$

In other words, if z is chosen uniformly in ∂D , and independently of X , then the probability of boundary intersection of the Euclidean segment $\hat{B}_\varepsilon(z)$ and of X scales as

$$\mathbb{P}\{\hat{B}_\varepsilon(z) \cap X \neq \emptyset\} \asymp \varepsilon^{\tilde{x}}, \quad (66)$$

in the sense that

$$\lim_{\varepsilon \rightarrow 0} \frac{\log \mathbb{P}\{\hat{B}_\varepsilon(z) \cap X \neq \emptyset\}}{\log \varepsilon} = \tilde{x}.$$

We say that X has **boundary quantum scaling exponent** $\tilde{\Delta}$ if when X and μ_γ^B (as defined above) are chosen independently we have

$$\lim_{\delta \rightarrow 0} \frac{\log \mathbb{E} \mu_\gamma^B(\hat{B}^\delta(X))}{\log \delta} = \tilde{\Delta}.$$

For z chosen uniformly in ∂D , and independently of X and μ_γ^B , the joint probability of intersection of the boundary quantum interval $\hat{B}^\delta(z)$ and of X then scales as

$$\mathbb{P}\{\hat{B}^\delta(z) \cap X \neq \emptyset\} \asymp \delta^{\tilde{\Delta}}, \quad (67)$$

in the sense that

$$\lim_{\delta \rightarrow 0} \frac{\log \mathbb{P}\{\hat{B}^\delta(z) \cap X \neq \emptyset\}}{\log \delta} = \tilde{\Delta}.$$

The validity of KPZ relation (4) for the pair $(\tilde{x}, \tilde{\Delta})$, as anticipated in [21], is insured by the following theorem [1]

Theorem 5.1. *Fix $\gamma \in [0, 2)$ and $\partial\hat{D}$ a closed subinterval of one of the boundary line segments of ∂D . If $X \cap \partial\hat{D}$ has Euclidean scaling exponent $\tilde{x} \geq 0$ then it has quantum scaling exponent $\tilde{\Delta} = \tilde{\Delta}_\gamma$, where $\tilde{\Delta}_\gamma$ is the non-negative solution to*

$$\tilde{x} = \frac{\gamma^2}{4} \tilde{\Delta}_\gamma^2 + \left(1 - \frac{\gamma^2}{4}\right) \tilde{\Delta}_\gamma. \quad (68)$$

The proofs in the boundary case proceed exactly the same as in the interior point case, up to factors of 2 in various places, which recombine and give the same KPZ relation as in the bulk case [1].

6. Liouville quantum duality

6.1. $\gamma' = 4/\gamma$ Liouville duality

For $\gamma > 2$, the Liouville measure (30) corresponds to the so-called “other” gravitational dressing of the Liouville potential [18, 19, 47, 48]. The corresponding random surface is meant to be the scaling limit of random simply connected surfaces with large amounts of area cut off by small bottlenecks [35, 40–44]. [See also [39].] This surface turns out to be a tree-like foam of Liouville quantum bubbles of *dual* parameter

$$\gamma' := 4/\gamma, \quad \gamma' < 2 < \gamma, \quad (69)$$

(“baby-universes”) connected to each other at “pinch points” and rooted at a “principal bubble” parameterized by D . A precise mathematical description requires additional machinery and will appear elsewhere. Here we relate γ to γ' formally, following Ref. [2].

The definition of quantum balls in (38) and (41) makes sense when $\gamma > 2$. Noting that $Q_{\gamma'} = Q_\gamma =: Q$, we have

$$\mu_{\gamma' \odot}(B_\varepsilon(z)) = \pi \varepsilon^{\gamma' Q} e^{\gamma' h_\varepsilon(z)} = \mu_{\gamma \odot}(B_\varepsilon(z))^{\gamma'/\gamma} = \mu_{\gamma \odot}^{4/\gamma^2} \quad (70)$$

(up to an irrelevant power of π) — i.e., a γ -quantum ball of size δ has γ' -quantum size

$$\delta' := \delta^{4/\gamma^2}. \quad (71)$$

(Intuitively, the ball contains about a δ' fraction of the total γ' -quantum area but only a $\delta < \delta'$ fraction of the γ -quantum area because the latter also includes points on non-principal bubbles.) The number of γ' -quantum size- δ' balls (i.e., γ -quantum size- δ balls) needed to cover the principal bubble D thus scales as

$$N_{\gamma'}(\delta') \asymp \delta'^{-1} = \delta^{-4/\gamma^2}. \quad (72)$$

If the fractal random subset $X \subset D$ has scaling exponent x , then (43), established for $\gamma' < 2$, essentially says [see (11-12)] that the expected number $N_{\gamma'}(\delta', X)$

of γ' -quantum size- δ' balls (i.e., number $N_\gamma(\delta, X)$ of γ -quantum size- δ balls) required to cover X , when normalized by $N_{\gamma'}(\delta')$, scales, as in (11), as

$$\frac{N_{\gamma'}(\delta', X)}{N_{\gamma'}(\delta')} \asymp \delta'^{\Delta_{\gamma'}}. \quad (73)$$

From (71), (72) follows the scaling identity

$$N_{\gamma'}(\delta', X) \asymp \delta'^{\Delta_{\gamma'}-1} = \delta^{\Delta_\gamma-1} \asymp N_\gamma(\delta, X), \quad (74)$$

with the scaling exponent Δ_γ defined by the *duality relation*

$$\Delta_\gamma - 1 := \frac{4}{\gamma^2}(\Delta_{\gamma'} - 1), \quad \gamma' = 4/\gamma. \quad (75)$$

Eq. (74) is thus the generalization of the quantum scaling (12) to the case $\gamma > 2$.

Rewriting the KPZ relation (4) as

$$x - 1 = \frac{\gamma^2}{4}(\Delta - 1)^2 + \left(1 + \frac{\gamma^2}{4}\right)(\Delta - 1),$$

we note that the triple $(x, \gamma, \Delta_\gamma)$, thanks to (75), satisfies (4) since $(x, \gamma', \Delta'_{\gamma'})$ does. The quantum scaling exponent Δ_γ , as defined by the duality relation (75), then *coincides* with the quantum scaling exponent (56), (58) *generalized to $\gamma > 2$* :

$$\begin{aligned} \Delta_\gamma &= \gamma^{-1}[(a_\gamma^2 + 4x)^{1/2} - a_\gamma], \\ \gamma &> 2, \quad a_\gamma = 2/\gamma - \gamma/2 < 0. \end{aligned} \quad (76)$$

The duality relation (75) can indeed be written as

$$\gamma(\Delta_\gamma - 1) = \gamma'(\Delta_{\gamma'} - 1), \quad (77)$$

which is easily checked for (76). Dual quantum exponents appear in [19, 21, 47, 48]. In [21], the duality relation (75) was introduced in the natural context of boundary exponents associated with the *dense phase* of the $O(N)$ model, or with the Stochastic Loewner Evolution SLE_κ , for $\kappa \geq 4$. (See also the Lecture Notes [62].)

6.2. Brownian approach to duality [2]

When $\gamma > 2$, the ε -regularized measures $M_\varepsilon(z)dz$ (30) converge to zero. If we choose the pair (z, h) from the weighted measure $M_\varepsilon(z)dhdz$ as in (34) and consider the Brownian description (47), we find that $a_\gamma < 0$ for $\gamma > 2$, i.e., the drift term runs in a direction opposite to $A > 0$, so that $T_A = \infty$ for large A . The weighted measure is thus *singular*, i.e., there is a quantum area of at least δ *localized* at z for small enough δ .

The Brownian martingale result (55) still holds for $\gamma > 2$

$$\begin{aligned} \mathbb{E}[\exp(-2xT_A)1_{T_A < \infty}] &= \delta^{\Delta_\gamma} \\ [(a_\gamma^2 + 4x)^{1/2} - a_\gamma]/\gamma &= \Delta_\gamma. \end{aligned} \quad (78)$$

For $x = 0$, it gives the probability, at a given z , for T_A to be finite:

$$\mathbb{P}(T_A < \infty) = \mathbb{E}[1_{T_A < \infty}] = \delta^{\Delta_\gamma(0)} = \delta^{1-4/\gamma^2} = \delta/\delta' < 1, \quad (79)$$

where $\Delta_\gamma(0) = (|a_\gamma| - a_\gamma)/\gamma = 1 - 4/\gamma^2$.

We may define a δ -regularized measure $M^\delta(z)dhdz$ as $M_{\varepsilon_A}(z)dhdz$ restricted to the event $T_A < \infty$. This means that there is *strictly less* than quantum area δ localized at point z . Replacing γ with $\gamma' = 4/\gamma$ and δ with δ' has the same effect as multiplying $M^\delta(z)$ by $\delta'/\delta = \delta^{4/\gamma^2-1}$, so

$$\delta^{4/\gamma^2-1} M^\delta(z)dz = \frac{\delta'}{\delta} M^\delta(z)dz =: d\mu_{\gamma', \varepsilon_A} \quad (80)$$

does converge to $d\mu_{\gamma'}$ with $\gamma' < 2$.

From (78) and (79), we deduce the *conditional* expectation

$$\frac{\mathbb{E}[\exp(-2xT_A)1_{T_A < \infty}]}{\mathbb{E}[1_{T_A < \infty}]} = \delta^{\Delta_\gamma} \times \frac{\delta'}{\delta} = \delta'^{\Delta_{\gamma'}}, \quad (81)$$

scaling now in terms of the γ' -quantum area δ' with the *dual* quantum scaling exponent $\Delta_{\gamma'}$, $\gamma' < 2$, in complete agreement with the definition of measure (80) above.

The Liouville quantum gravity is expected to describe, for $\gamma > 2$, the continuum limit of critical discrete matrix models where a non-standard string susceptibility exponent $\gamma_{\text{str}} > 0$ appears; this happens precisely at special values of the couplings in those models, where a fine tuning of the fugacity controlling the occurrence of bottlenecks leads to a critical proliferation of “baby-universes”. (See [18, 19, 21, 35, 40–44, 47, 48].)

Finally, for the SLE_κ process $\gamma = \sqrt{\kappa}$ [1], so that the present Liouville $\gamma\gamma' = 4$ duality and the $\text{SLE}_{\kappa\kappa' = 16}$ duality [57, 60] coincide.

Conclusion and Perspectives

This contribution has explained the KPZ relation for continuum Liouville quantum gravity. Outstanding open problems include the relations of discrete models and of SLE to Liouville quantum gravity, as described in [1, 2]. The combinatorial approach to discrete maps, for instance, is well-suited to the study of geodesics [72, 86, 89, 91, 92], and allows a detailed mathematical construction of continuum scaling limits [80, 88, 93]. A conjecture due to Peter Jones states that SLE can be defined via conformal welding and the boundary measure induced by the Gaussian free field [131]. We hope these questions can be addressed by the methods described here [1, 2].

Acknowledgements

This Proceedings contribution describes joint work with SCOTT SHEFFIELD, of the MIT Department of Mathematics, and it is a great pleasure to thank him for his collaboration over the last two years. It is also a pleasure to thank Michael Aizenman, Omer Angel, Jacques Franchi, Peter Jones, Igor Klebanov, Ivan Kostov, Greg

Lawler, Jean-François Le Gall, Andrei Okounkov, Tom Spencer, Paul Wiegmann, Michel Zinsmeister and the late Oded Schramm for useful discussions. I also thank Emmanuel Guitter for his generous help with the figures, and Nam-Gyu Kang for his generous help in creating the GFF and Liouville quantum gravity movies shown during the ICMP presentation. Last, but not least, it is a pleasure to thank Pavel Exner for his outstanding efficiency in organizing this Congress. This research has made use of NASA's Astrophysics Data System.

References

- [1] B. Duplantier and S. Sheffield, *arXiv:0808.1560* (2008).
- [2] B. Duplantier and S. Sheffield, *Phys. Rev. Lett.* **102**, p. 150603 (2009).
- [3] A. M. Polyakov, *Phys. Lett. B* **103**, 207 (1981).
- [4] A. M. Polyakov, *Phys. Lett. B* **103**, 211 (1981).
- [5] A. M. Polyakov, *Mod. Phys. Lett. A* **2**, p. 893 (1987).
- [6] A. M. Polyakov, *Gauge fields and strings* (Harwood Academic Publishers, Chur, 1987).
- [7] V. G. Knizhnik, A. M. Polyakov and A. B. Zamolodchikov, *Mod. Phys. Lett. A* **3**, 819 (1988).
- [8] F. David, *Mod. Phys. Lett. A* **3**, 1651 (1988).
- [9] J. Distler and H. Kawai, *Nucl. Phys. B* **321**, 509 (1989).
- [10] A. Polyakov, Two-dimensional quantum gravity. Superconductivity at high T_C , in *Fields, Strings and Critical Phenomena (Les Houches, Session XLIX, 1988)*, (North-Holland, Amsterdam, 1989) pp. 305–368.
- [11] N. Seiberg, *Progr. Theoret. Phys. Suppl.* **102**, 319 (1990).
- [12] P. Ginsparg and G. Moore, Lectures on 2d gravity and 2d string theory (TASI 1992), in *Recent direction in particle theory, Proceedings of the 1992 TASI*, eds. J. Harvey and J. Polchinski (World Scientific, Singapore, 1993)
- [13] F. David, Random matrices and two-dimensional gravity, in *Fundamental problems in statistical mechanics, VIII (Altenberg, 1993)*, (North-Holland, Amsterdam, 1994) pp. 105–126.
- [14] F. David, Simplicial quantum gravity and random lattices, in *Gravitation et quantifications (Les Houches, Session LVII, 1992)*, eds. B. Julia and J. Zinn-Justin (Elsevier B.V., Amsterdam, 1995) pp. 679–749.
- [15] J. Ambjørn, J. Jurkiewicz and Y. Watabiki, *Nucl. Phys. B* **454**, 313 (1995).
- [16] J. Ambjørn and Y. Watabiki, *Nucl. Phys. B* **445**, 129 (1995).
- [17] P. Di Francesco, P. Ginsparg and J. Zinn-Justin, *Phys. Rep.* **254**, 1 (1995).
- [18] I. R. Klebanov, *Phys. Rev. D* **51**, 1836 (1995).
- [19] I. R. Klebanov and A. Hashimoto, *Nucl. Phys. B Proc. Suppl.* **45**, 135 (1996).
- [20] J. Ambjørn, B. Durhuus and T. Jonsson, Quantum geometry, a statistical field theory approach, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 1997) pp. xiv+363.
- [21] B. Duplantier, Conformal fractal geometry & boundary quantum gravity, in *Fractal geometry and applications: a jubilee of Benoît Mandelbrot, Part 2*, , Proc. Sympos. Pure Math. Vol. 72 (Amer. Math. Soc., Providence, RI, 2004) pp. 365–482.
- [22] B. Duplantier, Conformal random geometry, in *Mathematical Statistical Physics (Les Houches Summer School, Session LXXXIII, 2005)*, eds. A. Bovier, F. Dunlop, F. den Hollander, A. van Enter and J. Dalibard (Elsevier B.V., Amsterdam, 2006) pp. 101–217.

- [23] E. Brézin, C. Itzykson, G. Parisi and J. B. Zuber, *Commun. Math. Phys.* **59**, 35 (1978).
- [24] J. Ambjørn, B. Durhuus and J. Fröhlich, *Nucl. Phys. B* **257**, 433 (1985).
- [25] V. A. Kazakov, I. K. Kostov and A. A. Migdal, *Phys. Lett. B* **157**, 295 (1985).
- [26] F. David, *Phys. Lett. B* **159**, 303 (1985).
- [27] D. V. Boulatov, V. A. Kazakov, I. K. Kostov and A. A. Migdal, *Nucl. Phys. B* **275**, 641 (1986).
- [28] D. V. Boulatov, V. A. Kazakov, I. K. Kostov and A. A. Migdal, *Phys. Lett. B* **174**, 87 (1986).
- [29] V. A. Kazakov, *Phys. Lett. A* **119**, 140 (1986).
- [30] B. Duplantier and I. Kostov, *Phys. Rev. Lett.* **61**, 1433 (1988).
- [31] B. Duplantier and I. K. Kostov, *Nucl. Phys. B* **340**, 491 (1990).
- [32] M. Gaudin and I. Kostov, *Phys. Lett. B* **220**, 200 (1989).
- [33] I. K. Kostov, *Mod. Phys. Lett. A* **4**, 217 (1989).
- [34] I. K. Kostov, *Nucl. Phys. B* **326**, 583 (1989).
- [35] S. R. Das, A. Dhar, A. M. Sengupta and S. R. Wadia, *Mod. Phys. Lett. A* **5**, 1041 (1990).
- [36] G. Moore, N. Seiberg and M. Staudacher, *Nucl. Phys. B* **362**, 665 (1991).
- [37] V. A. Kazakov and I. K. Kostov, *Nucl. Phys. B* **386**, 520 (1992).
- [38] B. Eynard and J. Zinn-Justin, *Nucl. Phys. B* **386**, 558 (1992).
- [39] S. Jain and S. D. Mathur, *Phys. Lett. B* **286**, 239 (1992).
- [40] G. P. Korchemsky, *Phys. Lett. B* **296**, 323 (1992).
- [41] G. P. Korchemsky, *Mod. Phys. Lett. A* **7**, 3081 (1992).
- [42] L. Alvarez-Gaumé, J. L. F. Barbón and Č. Crnković, *Nucl. Phys. B* **394**, 383 (1993).
- [43] B. Durhuus, *Nucl. Phys. B* **426**, 203 (1994).
- [44] J. Ambjørn, B. Durhuus and T. Jonsson, *Mod. Phys. Lett. A* **9**, 1221 (1994).
- [45] J.-M. Daul, *arXiv:hep-th/9502014* (1995).
- [46] B. Eynard and C. Kristjansen, *Nucl. Phys. B* **455**, 577 (1995).
- [47] I. R. Klebanov and A. Hashimoto, *Nucl. Phys. B* **434**, 264 (1995).
- [48] J. L. F. Barbón, K. Demeterfi, I. R. Klebanov and C. Schmidhuber, *Nucl. Phys. B* **440**, 189 (1995).
- [49] J. Ambjørn, K. N. Anagnostopoulos, U. Magnea and G. Thorleifsson, *Phys. Lett. B* **388**, 713 (1996).
- [50] B. Duplantier, *Phys. Rev. Lett.* **81**, 5489 (1998).
- [51] B. Duplantier, *Phys. Rev. Lett.* **82**, 3940 (1999).
- [52] B. Duplantier, *Phys. A* **263**, 452 (1999), STATPHYS 20 (Paris, 1998).
- [53] B. Duplantier, *Phys. Rev. Lett.* **82**, 880 (1999).
- [54] B. Eynard and G. Bonnet, *Phys. Lett. B* **463**, 273 (1999).
- [55] V. A. Kazakov and P. Zinn-Justin, *Nucl. Phys. B* **546**, 647 (1999).
- [56] I. K. Kostov, *Nucl. Phys. B* **575**, 513 (2000).
- [57] B. Duplantier, *Phys. Rev. Lett.* **84**, 1363 (2000).
- [58] P. Di Francesco, O. Golinelli and E. Guitter, *Nucl. Phys. B* **570**, 699 (2000).
- [59] B. Duplantier and I. A. Binder, *Phys. Rev. Lett.* **89**, p. 264101 (2002).
- [60] B. Duplantier, *J. Stat. Phys.* **110**, 691 (2003), Special issue in honor of Michael E. Fisher's 70th birthday (Piscataway, NJ, 2001).
- [61] I. K. Kostov, *J. Stat. Mech.* **08**, p. P08023 (2007).
- [62] I. K. Kostov, Boundary $O(n)$ models and 2D quantum gravity, in *Exact methods in low-dimensional statistical physics and quantum theory (Les Houches Summer School, Session LXXXIX, 2008)*, eds. S. Ouvry, J. Jacobsen, V. Pasquier, D. Serban and L. Cugliandolo (Oxford University Press, Great Clarendon Street, 2009).

- [63] R. Cori and B. Vauquelin, *Canad. J. Math.* **33**, 1023 (1981).
- [64] G. Schaeffer, *PhD Thesis, Univ. Bordeaux I, Talence* (1998).
- [65] C. Banderier, P. Flajolet, G. Schaeffer and M. Soria, *Random Struct. & Algorithms* **19**, 194 (2001).
- [66] P. Flajolet, B. Salvy and G. Schaeffer, *Electron. J. Comb.* **11**, #R34,1 (2004).
- [67] J. Bouttier, P. Di Francesco and E. Guitter, *Nucl. Phys. B* **645**, 477 (2002).
- [68] M. Bousquet-Mélou and G. Schaeffer, The degree distribution in bipartite planar maps: applications to the Ising model, in *Proceedings of FPSAC 03, (Formal power series and algebraic combinatorics)*, Vadstena, Sweden, June 2003, K. Eriksson and S. Linusson eds. (*arXiv:math/0211070*), 2003.
- [69] O. Angel and O. Schramm, *Comm. Math. Phys.* **241**, 191 (2003).
- [70] O. Angel, *Geom. & Funct. Ann.* **13**, 935 (2003).
- [71] J. Bouttier, P. Di Francesco and E. Guitter, *Nucl. Phys. B* **655**, 313 (2003).
- [72] J. Bouttier, P. Di Francesco and E. Guitter, *Nucl. Phys. B* **663**, 535 (2003).
- [73] P. Chassaing and G. Schaeffer, *Probab. Th. Rel. Fields* **128**, 161 (2004).
- [74] P. Chassaing and B. Durhuus, *Ann. Probab.* **34**, 879 (2006).
- [75] M. Krikun, *arXiv:math/0512304* (2005).
- [76] L. Ménard, *arXiv:0812.0965 [math.PR]* (2008).
- [77] P. Di Francesco and E. Guitter, *Phys. Rep.* **415**, 1 (2005).
- [78] J. Bouttier, P. Di Francesco and E. Guitter, *J. Phys. A* **40**, 7411 (2007).
- [79] G. M. Miermont, *arXiv:0712.3688* (2007).
- [80] J.-F. Le Gall, *Invent. Math.* **169**, 621 (2007).
- [81] J.-F. Marckert and G. Miermont, *Ann. Probab.* **35**, 1642 (2007).
- [82] O. Bernardi, *arXiv:math/0601684* (2006).
- [83] O. Bernardi, *Ann. Comb.* **12**, 139 (2008).
- [84] O. Bernardi, *Ann. Comb.* **12**, 17 (2008).
- [85] O. Bernardi, *Electron. J. Comb.* **15**, #R109, (2008).
- [86] J. Bouttier and E. Guitter, *J. Phys. A* **41**, p. 145001 (2008).
- [87] G. Miermont and M. Weill, *Electron. J. Probab.* **13**, no. 4, 79 (2008).
- [88] G. M. Miermont, *Electron. Commun. Probab.* **13**, 248 (2008).
- [89] J. Bouttier and E. Guitter, *J. Stat. Mech.* **7**, p. P07020 (2008).
- [90] J.-F. Le Gall and F. Paulin, *Geom. & Funct. Ann.* **18**, 893 (2008).
- [91] J.-F. Le Gall, *Acta Math. (to appear)* (2008), *arXiv:math.PR/0804.3012*.
- [92] J. Bouttier and E. Guitter, *J. Stat. Mech.* **3**, p. P03001 (2009).
- [93] J.-F. Le Gall and G. Miermont, *arXiv:math.PR/0907.3262* (2009).
- [94] O. Bernardi and M. Bousquet-Mélou, *arXiv:0909.1695* (2009).
- [95] G. Chapuy, M. Marcus and G. Schaeffer, *arXiv:0712.3649* (2007).
- [96] G. Chapuy, *arXiv:0805.0352* (2008).
- [97] G. Chapuy, *arXiv:0804.0546* (2008).
- [98] B. Eynard and N. Orantin, *arXiv:math-ph/0702.045* (2007).
- [99] B. Eynard and N. Orantin, *J. High Energy Phys.* **6**, p. 37 (2008).
- [100] B. Eynard, *J. High Energy Phys.* **3**, p. 3 (2009).
- [101] A. M. Polyakov, *arXiv:0812.0183* (2008).
- [102] L. A. Takhtajan, *Mod. Phys. Lett. A* **8**, 3529 (1993).
- [103] I. K. Kostov, *Nucl. Phys. B* **658**, 397 (2003).
- [104] I. K. Kostov, B. Ponsot and D. Serban, *Nucl. Phys. B* **683**, 309 (2004).
- [105] M. Goulian and M. Li, *Phys. Rev. Lett.* **66**, 2051 (1991).
- [106] H. Dorn and H.-J. Otto, *Nucl. Phys. B* **429**, 375 (1994).
- [107] J. Teschner, *Phys. Lett. B* **363**, 65 (1995).
- [108] A. B. Zamolodchikov and A. B. Zamolodchikov, *Nucl. Phys. B* **477**, 577 (1996).

- [109] V. Fateev, A. B. Zamolodchikov and A. B. Zamolodchikov, *arXiv:hep-th/0001012* (2000).
- [110] J. Teschner, *Class. Quant. Grav.* **18**, R153 (2001).
- [111] K. Hosomichi, *J. High Energy Phys.* **11**, p. 44 (2001).
- [112] B. Ponsot and J. Teschner, *Nucl. Phys. B* **622**, 309 (2002).
- [113] A. B. Zamolodchikov, *Int. J. Mod. Phys. A* **19**, 510 (2004).
- [114] I. K. Kostov, *Nucl. Phys. B* **689**, 3 (2004).
- [115] Y. Nakayama, *Int. J. Mod. Phys. A* **19**, 2771 (2004).
- [116] G. F. Lawler, O. Schramm and W. Werner, *Acta Math.* **187**, 237 (2001).
- [117] G. F. Lawler, O. Schramm and W. Werner, *Acta Math.* **187**, 275 (2001).
- [118] G. F. Lawler, O. Schramm and W. Werner, *Ann. Inst. H. Poincaré Probab. Statist.* **38**, 109 (2002).
- [119] I. Benjamini and O. Schramm, *Commun. Math. Phys.* **289**, 46 (2009).
- [120] R. Rhodes and V. Vargas, *arXiv:0807.1036* (2008).
- [121] F. David and M. Bauer, *J. Stat. Mech.* **3**, p. P03004 (2009).
- [122] K. Anagnostopoulos, P. Bialas and G. Thorleifsson, *J. Stat. Phys.* **94**, 321 (1999).
- [123] W. Janke and D. A. Johnston, *Phys. Lett. B* **460**, 271 (1999).
- [124] S. Sheffield, *Probab. Th. Rel. Fields* **139**, 521 (2007).
- [125] M. Bauer, *PhD Thesis, Université Paris 7* (1990).
- [126] B. Simon, *The $P(\Phi)_2$ Euclidean (quantum) field theory* (Princeton University Press, Princeton, 1974).
- [127] R. Høegh-Krohn, *Commun. Math. Phys.* **21**, 244 (1971).
- [128] S. Albeverio, G. Gallavotti and R. Høegh-Krohn, *Commun. Math. Phys.* **70**, 187 (1979).
- [129] S. Albeverio, R. Høegh-Krohn, S. Paycha and S. Scarlatti, *Acta Appl. Math.* **26**, 103 (1992).
- [130] X. Hu, J. Miller and Y. Peres, *arXiv:0902.3842* (2009).
- [131] K. Astala, P. Jones, A. Kupiainen and E. Saksman, *arXiv:0909.1003* (2009).

UNIVERSALITY OF WIGNER RANDOM MATRICES

LÁSZLÓ ERDŐS

*Institute of Mathematics, University of Munich
Munich, D-80333, Germany
E-mail: lerdos@math.lmu.de*

We consider $N \times N$ symmetric or hermitian random matrices with independent, identically distributed entries where the probability distribution for each matrix element is given by a measure ν with a subexponential decay. We prove that the local eigenvalue statistics in the bulk of the spectrum for these matrices coincide with those of the Gaussian Orthogonal Ensemble (GOE) and the Gaussian Unitary Ensemble (GUE), respectively, in the limit $N \rightarrow \infty$. Our approach is based on the study of the Dyson Brownian motion via a related new dynamics, the local relaxation flow. We also show that the Wigner semicircle law holds locally on the smallest possible scales and we prove that eigenvectors are fully delocalized and eigenvalues repel each other on arbitrarily small scales.

Keywords: Wigner random matrix, Dyson Brownian Motion.

1. Introduction

A central question concerning random matrices is the universality conjecture which states that local statistics of eigenvalues of large $N \times N$ square matrices H are determined by the symmetry type of the ensembles but are otherwise independent of the details of the distributions.

There are two types of universalities: the edge universality and the bulk universality concerning the interior of the spectrum. The edge universality is commonly approached via the fairly robust moment method [33, 34]; very recently an alternative proof was given [36]. The bulk universality is a subtler problem. In the hermitian case, it states that the local k -point correlation functions of the eigenvalues, after appropriate rescaling, are given by the determinant of the *sine kernel*

$$\det \left(K(x_\ell - x_j) \right)_{\ell, j=1}^k, \quad K(x) = \frac{\sin \pi x}{\pi x}, \quad (1)$$

independently of the distribution of the entries. Similar statement holds for the symmetric matrices but the explicit formulae are somewhat more complicated.

For ensembles that remain invariant under the transformations $H \rightarrow U^* H U$ for any unitary matrix U , the joint probability density function of all the N eigenvalues can be explicitly computed. These ensembles are typically given by the probability

density

$$P(H)dH \sim \exp(-N\text{Tr } V(H))dH$$

where V is a real function with sufficient growth at infinity and dH is the flat measure. The eigenvalues are strongly correlated and they are distributed according to a Gibbs measure with a long range logarithmic interaction potential. The joint probability density of the eigenvalues of H can be computed explicitly:

$$f(\lambda_1, \lambda_2, \dots, \lambda_N) = \text{const.} \prod_{i < j} (\lambda_i - \lambda_j)^\beta \prod_{j=1}^N e^{-N \sum_{j=1}^N V(\lambda_j)}, \quad (2)$$

where $\beta = 1$ for symmetric and $\beta = 2$ for hermitian ensembles. The local statistics can be obtained via a detailed analysis of orthogonal polynomials on the real line with respect to the weight function $\exp(-V(x))$. Quadratic V corresponds to the Gaussian ensembles. This approach was originally applied [26] for ensembles that lead to classical orthogonal polynomials (e.g. GUE leads to Hermite polynomials). Later a general method using orthogonal polynomials has been developed to tackle a very general class of unitary ensembles (see, e.g. [5, 9, 10, 26, 29] and references therein).

Many natural matrix ensembles are typically not unitarily invariant; the most prominent example is the Wigner matrices. These are symmetric or hermitian matrices whose entries above the diagonal are independent, identically distributed random variables. The only unitarily invariant Wigner ensembles are the Gaussian ensembles. For general Wigner matrices, no explicit formula is available for the joint eigenvalue distribution. Thus the basic algebraic connection between eigenvalue ensembles and orthogonal polynomials is missing and completely new methods needed to be developed.

The bulk universality for *hermitian* Wigner ensembles has been established jointly with J. Ramirez, S. Péché, B. Schlein and H.T. Yau, and independently by Tao-Vu [20, 21, 35]. These works rely on the Wigner matrices with Gaussian divisible distribution, i.e. ensembles of the form

$$\hat{H} + sV, \quad (3)$$

where $\hat{H}$ is a Wigner matrix, V is an independent standard GUE matrix and s is a positive constant. Johansson [24] (see also [4]) proved bulk universality for the eigenvalues of such matrices using an *explicit* formula by Brézin-Hikami [6, 24] on the correlation functions. Unfortunately, the similar formula for symmetric matrices is not very explicit and the technique of [20, 24] cannot be extended to prove universality for symmetric Wigner matrices.

A key observation of Dyson is that if the parameter s in the matrix $\hat{H} + sV$ is varied and s^2 is interpreted as time, then the evolution of the eigenvalues is given by a coupled system of stochastic differential equations, commonly called the Dyson Brownian motion (DBM) [12]. If we replace the Brownian motions by the Ornstein-Uhlenbeck processes to keep the variance constant, then the resulting dynamics on

the eigenvalues, which we still call DBM, has the GUE eigenvalue distribution as the invariant measure. Thus the result of Johansson can be interpreted as stating that the local statistics of GUE is reached via DBM for time of order one. In fact, by analyzing the dynamics of DBM with ideas from the hydrodynamical limit, we have extended Johansson's result to $s^2 \gg N^{-3/4}$ [19]. The key observation of [19] is that the local statistics of eigenvalues depend exclusively on the approach to local equilibrium which in general is faster than reaching global equilibrium. Unfortunately, the identification of local equilibria still uses explicit representations of correlation functions by orthogonal polynomials (following e.g. [29]), and the extension to other ensembles is not a simple task.

Therefore, the universality for symmetric random matrices remained open and the only partial result is given by Tao-Vu (Theorem 23 in [35]) for Wigner matrices with the first four moments of the matrix elements matching those of GOE.

In [18], together with B. Schlein and H.T. Yau, we have introduced a general approach based on a new stochastic flow, the local relaxation flow, which locally behaves like DBM, but has a faster decay to equilibrium. This approach completely circumvents explicit formulae. It is thus applicable to prove the universality for a very broad class of matrices that includes hermitian, symmetric and symplectic Wigner matrices and in principle it works also for Wishart matrices and general β -ensembles. The heart of the proof is a convex analysis and the model specific information involve only estimates on the accuracy of the local density of states. For simplicity of the presentation, we will focus on the hermitian and symmetric cases, the necessary modifications for the other cases are technical.

We present results only about the convergence of the local correlation functions; this implies, among others, that the distribution of the gap (difference between neighboring eigenvalues) is universal as well (Wigner surmise). In particular, short gaps are suppressed, i.e. the eigenvalues tend to repel each other. This feature is characteristic to the strongly correlated point process of eigenvalues of random matrices in contrast to the Poisson process of independent points.

Universality of local eigenvalue statistics is believed to hold for a much broader class of matrix ensembles than we have introduced. Wigner has originally invented random matrices to mimic the eigenvalues of the then unknown Hamiltonian of heavy nuclei; lacking any information, he assumed that the matrix elements are i.i.d. random variables subject to the hermitian condition. Conceivably, the matrix elements need not be fully independent or identically distributed for universality. There is little known about matrices with correlated entries, apart from the unitary invariant ensembles that represent a very specific correlation. In case of a certain class of Wigner matrices with weakly correlated entries, the semicircle law and its Gaussian fluctuation have been proven [31, 32].

Much more studied are various classes of random matrices with independent but not identically distributed entries. The most prominent example is the Anderson model [2], i.e. a Schrödinger operator on a regular square lattice with a

random potential. Restricted to a finite box, it can be represented by a matrix whose diagonal elements are i.i.d. random variables; the deterministic off-diagonal elements are given by the Laplacian. In space dimensions three or higher and for weak randomness, the Anderson model is conjectured to exhibit metal-insulator transition. Near the spectral edges, the eigenfunctions are localized [1, 22] and the local eigenvalue statistics is Poissonian [28]; in particular there is no level repulsion. It is conjectured, but not yet proven, that in the middle of the spectrum the eigenfunctions are extended (some results on the quantum diffusion and delocalization of eigenfunctions are available in a certain scaling limit [7, 14]). Furthermore, in the delocalization regime the local eigenvalue statistics are expected to be given by GUE or GOE statistics, depending whether the time reversal symmetry is broken by magnetic field or not. Based upon this conjecture, local eigenvalue statistics is used to compute the phase diagram numerically. It is very remarkable that the random Schrödinger operator, represented by a very sparse random matrix, exhibits the same universality class as the full Wigner matrix, at least in a certain energy range.

An intermediate class of ensembles between these two extremes is the family of random band matrices. These are hermitian or symmetric random matrices H with independent but not identically distributed entries. The variance of H_{ij} depends only on $|i - j|$ and it becomes negligible if $|i - j|$ exceeds a given parameter W , the band-width; for example, $\mathbb{E}|H_{ij}|^2 \sim \exp(-|i - j|/W)$. It is conjectured that for narrow bands, $W \ll \sqrt{N}$, the local eigenvalue statistics is Poisson, while for broad bands, $W \gg \sqrt{N}$ it is given by GUE or GOE, depending on the symmetry class. (Localization properties of H for $W \ll N^{1/8}$ has been recently shown [30] but not local statistics.) To mimic the three dimensional Anderson model, the rows and columns of H may be labelled by a finite domain of the three dimensional lattice, $i, j \in \Lambda \subset \mathbb{Z}^3$. The only rigorous result for this three dimensional band matrix concerns the density of states by establishing that Wigner semicircle law holds as $W \rightarrow \infty$ [11].

Finally, we mention that universality of local eigenvalue statistics is often investigated by supersymmetric techniques in the physics literature. These methods are extremely powerful to extract the results by saddle point computations, but the analysis justifying the saddle point approximation usually lacks mathematical rigor. It is a challenge to the mathematical physics community to put the supersymmetric method on a solid mathematical basis; so far only the density of states has been investigated rigorously by using this technique [11].

2. Local semicircle law, delocalization and level repulsion

Each approach that proves bulk universality for general Wigner matrices requires first to analyze the local density of eigenvalues. The Wigner semicircle law [38] (and its analogue for Wishart matrices, the Marchenko-Pastur law [25]) has traditionally been among the first results established on random matrices. Typically, however,

the empirical density is shown to converge weakly on macroscopic scales, i.e. on intervals that contain $O(N)$ eigenvalues. Based upon our results [15–17], here we show that the semicircle law holds on much smaller scales as well.

To fix the notation, we assume that in the symmetric case the matrix elements of H are given by

$$h_{\ell k} = N^{-1/2} x_{\ell k}, \quad (4)$$

where $x_{\ell k}$ for $\ell < k$ are independent, identically distributed random variables with the distribution ν that has zero expectation and variance 1. The diagonal elements $x_{\ell \ell}$ are also i.i.d. with distribution $\tilde{\nu}$ that has zero expectation and variance two. In the hermitian case we assume that

$$h_{\ell k} = N^{-1/2} (x_{\ell k} + i y_{\ell k}) \quad (5)$$

where $x_{\ell k}$ and $y_{\ell k}$ are real i.i.d. random variables with zero expectation and variance $\frac{1}{2}$. The diagonal elements are also centered and have variance one. The eigenvalues of H will be denoted by $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$. The Gaussian ensembles (GUE and GOE) are special Wigner ensembles with Gaussian single-site distribution.

We will often need to assume that the distributions ν and $\tilde{\nu}$ have Gaussian decay, i.e. there exists $\delta_0 > 0$ such that

$$\int_{\mathbb{R}} \exp [\delta_0 x^2] d\nu(x) < \infty, \quad \int_{\mathbb{R}} \exp [\delta_0 x^2] d\tilde{\nu}(x) < \infty. \quad (6)$$

In several statements we can relax this condition to assuming only subexponential decay, i.e. that there exists $\delta_0 > 0$ and $\gamma > 0$ such that

$$\int_{\mathbb{R}} \exp [\delta_0 |x|^\gamma] d\nu(x) < \infty, \quad \int_{\mathbb{R}} \exp [\delta_0 |x|^\gamma] d\tilde{\nu}(x) < \infty. \quad (7)$$

The matrix elements have thus variance of order $1/N$. This normalization guarantees that the spectrum remains bounded as $N \rightarrow \infty$, in fact the spectrum converges to $[-2, 2]$ almost surely. Therefore the typical spacing between neighboring eigenvalues is of order $1/N$.

For any $I \subset \mathbb{R}$ let $\mathcal{N}_I$ denote the number of eigenvalues in I . Wigner's theorem [38] states that for any fixed interval I

$$\frac{\mathcal{N}_I}{N} \rightarrow \int_I \varrho_{sc}(x) dx$$

almost surely as $N \rightarrow \infty$, where

$$\varrho_{sc}(x) := \frac{1}{2\pi} \sqrt{(4 - x^2)_+}$$

is the density of the semicircle law. This result can be interpreted as a law of large numbers for the empirical eigenvalue density on macroscopic scales, i.e. for intervals that contain $O(N)$ eigenvalues. The following result shows that the semicircle law holds on intervals I of length $|I| = \eta \geq K/N$ for sufficiently large K .

Theorem 2.1. [17, Theorem 3.1] Suppose that (6) holds. Let $\kappa > 0$ and fix an energy $E \in [-2 + \kappa, 2 - \kappa]$. Consider the interval $I = [E - \frac{\eta}{2}, E + \frac{\eta}{2}]$ of length η about E . Then there exist positive constants C, c , depending only on κ , and a universal constant c_1 such that for any $\delta \leq c_1 \kappa$ there is $K = K_\delta$ such that

$$\mathbb{P}\left\{\left|\frac{\mathcal{N}_I}{N\eta} - \varrho_{sc}(E)\right| \geq \delta\right\} \leq Ce^{-c\delta^2\sqrt{N\eta}} \quad (8)$$

holds for all η satisfying $K/N \leq \eta \leq 1/K$.

In particular, this result shows that $\mathcal{N}_I/N\eta$ converges to $\varrho_{sc}(E)$ in probability as long as $\eta = \eta(N)$ is such that $\eta(N) \rightarrow 0$ and $N\eta(N) \rightarrow \infty$. The Gaussian decay condition (6) can be relaxed to (7) if $\eta \geq N^{-1+\varepsilon}$ with any $\varepsilon > 0$ at the expense of a weaker bound on the right hand side of (8), see Section 5 of [20]. The estimate also deteriorates if the energy is close to the edge, see Proposition 4.1 of [19] for a more precise statement. Based upon our proofs, similar estimates were given in [35, Theorem 56] for energies in the bulk and somewhat stronger bounds in [36, Theorem 1.7] for the edge.

Sketch of the proof. For any $z = E + i\eta$, $\eta > 0$, let

$$m(z) = m_N(z) = \frac{1}{N} \text{Tr} \frac{1}{H - z} = \frac{1}{N} \sum_{\alpha=1}^N \frac{1}{\lambda_\alpha - z} \quad (9)$$

be the Stieltjes transform of the empirical density of states and let

$$m_{sc}(z) = \int \frac{\varrho_{sc}(x)}{x - z} dx$$

be the Stieltjes transform of the semicircle law. Clearly $\varrho_\eta(E) = \frac{1}{\pi} \text{Im} \, m(z)$ gives the normalized density of states of H around E regularized on a scale η . Therefore it is sufficient to establish the convergence of $m(z)$ to $m_{sc}(z)$ for small $\eta = \text{Im} \, z$.

The first step of the proof is to provide an upper bound on $\mathcal{N}_I$. Let $B^{(k)}$ denote the $(N-1) \times (N-1)$ minor of H after removing the k -th row and k -th column. Let $\lambda_\alpha^{(k)}$, $\alpha = 1, 2, \dots, N-1$ denote the eigenvalues of $B^{(k)}$ and $\mathbf{u}_\alpha^{(k)}$ denote its eigenvectors. Computing the (k, k) diagonal element of the resolvent $(H - z)^{-1}$ we easily obtain the following expression for $m(z)$

$$m(z) = \frac{1}{N} \sum_{k=1}^N \frac{1}{H - z}(k, k) = \frac{1}{N} \sum_{k=1}^N \left[h_{kk} - z - \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{\xi_\alpha^{(k)}}{\lambda_\alpha^{(k)} - z} \right]^{-1}, \quad (10)$$

where

$$\xi_\alpha^{(k)} = N |\mathbf{a}^{(k)} \cdot \mathbf{u}_\alpha^{(k)}|^2, \quad (11)$$

and $\mathbf{a}^{(k)}$ is the k -th column of H without the diagonal element h_{kk} . Taking the imaginary part, and using $\mathcal{N}_I \leq C \text{Im} \, m(z)$, we have

$$\mathcal{N}_I \leq CN\eta^2 \sum_{k=1}^N \left| \sum_{\alpha: \lambda_\alpha^{(k)} \in I} \xi_\alpha^{(k)} \right|^{-1}. \quad (12)$$

It is an elementary fact that the eigenvalues of H and $B^{(k)}$, for each fixed k , are interlaced, i.e. the number of $\lambda_\alpha^{(k)}$ in I is at least $\mathcal{N}_I - 1$. For each fixed k the random variables $\{\xi_\alpha^{(k)} : \alpha = 1, 2, \dots, N-1\}$ are almost independent and have expectation value one, thus the probability of the event

$$\Omega_k := \left\{ \sum_{\alpha : \lambda_\alpha^{(k)} \in I} \xi_\alpha^{(k)} \leq \delta(\mathcal{N}_I - 1) \right\}$$

is negligible for small δ [17, Lemma 4.7]. On the complement of all Ω_k we thus have from (12) that

$$\mathcal{N}_I \leq \frac{CN^2\eta^2}{\delta(\mathcal{N}_I - 1)},$$

from which it follows that $\mathcal{N}_I \leq CN\eta$ with very high probability.

The second step of the proof is to establish that $m(z)$ and $m_{sc}(z)$ are close. Let $m^{(k)}(z)$ denote the Stieltjes transform of the empirical distribution of the eigenvalues $\lambda_\alpha^{(k)}$ of $B^{(k)}$. Then it follows from (10) that

$$m(z) = \frac{1}{N} \sum_{k=1}^N \frac{1}{h_{kk} - z - \left(1 - \frac{1}{N}\right)m^{(k)}(z) - X_k} \quad (13)$$

holds, where

$$X_k = \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{\xi_\alpha^{(k)} - 1}{\lambda_\alpha^{(k)} - z}.$$

Fixing the matrix $B^{(k)}$, we view X_k as a random variable of the independent $\mathbf{a}^{(k)}$ vector alone. Using again that the nominators $\xi_\alpha^{(k)} - 1$ are almost independent and have zero expectation, we obtain that X_k is bounded by $(N\eta)^{-1}$ with high probability [17, Lemma 6.1]. The interlacing property guarantees that $m(z)$ and $m^{(k)}(z)$ are close. Since h_{kk} is also small, we obtain from (13) that

$$m(z) = -\frac{1}{N} \sum_{k=1}^N \frac{1}{m(z) + z + \varepsilon_k}. \quad (14)$$

where ε_k are small with very high probability. Note that the Stieltjes transform of the semicircle law is the solution of the equation

$$m_{sc}(z) = -\frac{1}{m_{sc}(z) + z} \quad (15)$$

that is stable away from the spectral edges, $z = \pm 2$. Comparing the solution of (14) and (15) we obtain that $|m - m_{sc}|$ is small. Strictly speaking, this argument applies only for $\eta \geq (\log N)^4/N$ since the smallness of each ε_k is guaranteed only apart from a set of probability $e^{-c\sqrt{N\eta}}$ [17, Lemma 4.2] and there are N possible values of k . On very short scale, our proof uses an additional expansion of the denominators in (14) up to second order and we use that the expectation of X_k , the main contribution to ε_k , vanishes [17, Section 6]. $\square$

The second result concerns the delocalization of eigenvectors. The motivation comes from the Anderson model. In the infinite volume, the extended states regime is usually characterized by the absolute continuity of the spectrum; such characterization is meaningless for finite matrices. However, the lack of concentration of the eigenfunctions for the finite volume approximations of the Anderson Hamiltonian is already a signature of the extended states regime.

If $\mathbf{v}$ is an ℓ^2 -normalized eigenvector of H , then the size of the ℓ^p -norm of $\mathbf{v}$, for $p > 2$, gives information about delocalization. Complete delocalization occurs when $\|\mathbf{v}\|_p \lesssim N^{-1/2+1/p}$ (note that $\|\mathbf{v}\|_p \geq CN^{-1/2+1/p}\|\mathbf{v}\|_2$). The following result shows that eigenvectors are fully delocalized with a very high probability.

Theorem 2.2. *[17, Corollary 3.2] Under the conditions of Theorem 2.1, for any $|E| < 2$, fixed K and $2 < p < \infty$ we have*

$$\mathbb{P}\left\{\exists \mathbf{v} : H\mathbf{v} = \lambda\mathbf{v}, |\lambda - E| \leq \frac{K}{N}, \|\mathbf{v}\|_2 = 1, \|\mathbf{v}\|_p \geq MN^{-\frac{1}{2} + \frac{1}{p}}\right\} \leq Ce^{-c\sqrt{M}}$$

for M and N large enough.

The proof is an easy consequence of Theorem 2.1 and will be omitted here.

The local semicircle law asserts that the empirical density on scales $\eta \gg O(1/N)$ is close to the semicircle density. On even smaller scales $\eta \leq O(1/N)$, the empirical density fluctuates, but its average, $\mathbb{E} \varrho_\eta(E)$, remains bounded uniformly in η . This is a type of Wegner estimate that plays a central role in the localization theory of random Schrödinger operators. In particular, it says that the probability of finding at least one eigenvalue in an interval I of size $\eta = \varepsilon/N$ is bounded by $C\varepsilon$ uniformly in N and $\varepsilon \leq 1$, i.e. no eigenvalue can stick to any value. Furthermore, if the eigenvalues were independent (Poisson process), then the probability of finding $n = 1, 2, 3, \dots$ eigenvalues in I were proportional with ε^n . For random matrices in the bulk of the spectrum this probability is much smaller. This phenomenon is known as level repulsion and the precise statement is the following:

Theorem 2.3. *[17, Theorem 3.4 and 3.5] Suppose (6) holds and the measure ν is absolutely continuous with a strictly positive and smooth density. Let $|E| < 2$ and $I = [E - \eta/2, E + \eta/2]$ with $\eta = \varepsilon/N$. Then for any fixed n ,*

$$\mathbb{P}(\mathcal{N}_I \geq n) \leq \begin{cases} C_n \varepsilon^{n^2} & [\text{hermitian case}] \\ C_n \varepsilon^{n(n+1)/2} & [\text{symmetric case}] \end{cases} \quad (16)$$

uniformly in $\varepsilon \leq 1$ and for all sufficiently large N .

The exponents are optimal as one can easily see from the Vandermonde determinant in the joint probability density (2) for unitary ensembles. The sine kernel behavior (1) implies level repulsion (and even a lower bound on $\mathbb{P}(\mathcal{N}_I \geq n)$), but usually not on arbitrarily small scales since sine kernel is typically proven only as a weak limit (see (22) later).

Sketch of the proof. The starting point is formula (10) together with

$$\mathcal{N}_I \leq CN\eta \operatorname{Im} m(E + i\eta).$$

This implies

$$\mathcal{N}_I \leq C\eta \sum_{k=1}^N \frac{1}{(a_k^2 + b_k^2)^{1/2}} \quad (17)$$

with

$$a_k := \eta + \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{\eta \xi_{\alpha}^{(k)}}{(\lambda_{\alpha}^{(k)} - E)^2 + \eta^2}, \quad b_k := h_{kk} - E - \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{(\lambda_{\alpha}^{(k)} - E) \xi_{\alpha}^{(k)}}{(\lambda_{\alpha}^{(k)} - E)^2 + \eta^2},$$

where a_k and b_k are the imaginary and real part, respectively, of the reciprocal of the summands in (10) and $\xi_{\alpha}^{(k)}$ was defined in (11). The proof of Theorem 2.1 relied only on the imaginary part, i.e. b_k in (17) was neglected. In the proof of Theorem 2.3, however, we make an essential use of b_k as well. Since typically $1/N \lesssim |\lambda_{\alpha}^{(k)} - E|$, we note that a_k^2 is much smaller than b_k^2 if $\eta \ll 1/N$ and this is the relevant regime for the Wegner estimate and for the level repulsion.

Assuming a certain smoothness condition on the distribution $d\nu$, the distribution of the variables $\xi_{\alpha}^{(k)}$ will also be smooth even if we fix an index k and we condition on the minor $B^{(k)}$, i.e. if we fix the eigenvalues $\lambda_{\alpha}^{(k)}$ and the eigenvectors $\mathbf{u}_{\alpha}^{(k)}$. Although the random variables $\xi_{\alpha}^{(k)} = N |\mathbf{a}^{(k)} \cdot \mathbf{u}_{\alpha}^{(k)}|^2$ are not independent for different α 's, they are sufficiently decorrelated so that the distribution of b_k inherits some smoothness from $\mathbf{a}^{(k)}$. Sufficient smoothness on the distribution of b_k makes the expectation value $(a_k^2 + b_k^2)^{-p/2}$ finite for any $p > 0$. This will give a bound on the p -th moment on $\mathcal{N}_I$ which will imply (16).

We present this idea for hermitian matrices and for the simplest case $k = 1$. From (17) we have

$$\mathbb{P}(\mathcal{N}_I \geq 1) \leq \mathbb{E} \mathcal{N}_I^2 \leq C(N\eta)^2 \mathbb{E} \frac{1}{a_1^2 + b_1^2}.$$

Dropping the superscript $k = 1$ and introducing the notation

$$d_{\alpha} = \frac{N(\lambda_{\alpha} - E)}{N^2(\lambda_{\alpha} - E)^2 + \varepsilon^2}, \quad c_{\alpha} = \frac{\varepsilon}{N^2(\lambda_{\alpha} - E)^2 + \varepsilon^2},$$

we have

$$\mathbb{P}(\mathcal{N}_I \geq 1) \leq C\varepsilon^2 \mathbb{E} \left[\left(\sum_{\alpha=1}^{N-1} c_{\alpha} \xi_{\alpha} \right)^2 + \left(h - E - \sum_{\alpha=1}^{N-1} d_{\alpha} \xi_{\alpha} \right)^2 \right]^{-1}. \quad (18)$$

From the local semicircle law we know that with very high probability, there are several eigenvalues λ_{α} within a distance of $O(1/N)$ of E . Choosing four such eigenvalues, we can guarantee that for some index γ

$$c_{\gamma}, c_{\gamma+1} \geq C\varepsilon, \quad d_{\gamma+2}, d_{\gamma+3} \geq C \quad (19)$$

for some positive constant C . If ξ_α 's were indeed independent and distributed according to the square of a complex random variable z_α with a smooth and decaying density $d\mu(z)$ on the complex plane, then the expectation in (18) would be bounded by

$$\sup_E \int \frac{1}{(c_\gamma |z_\gamma|^2 + c_{\gamma+1} |z_{\gamma+1}|^2)^2 + (E - d_{\gamma+2} |z_{\gamma+2}|^2 - d_{\gamma+3} |z_{\gamma+3}|^2)^2} \prod_{j=0}^3 d\mu(z_{\gamma+j}). \quad (20)$$

Simple calculation shows that this integral is bounded by $C\varepsilon^{-1}$ assuming the lower bounds (19). Combining this bound with (18), we obtain (16) for $n = 1$. The proof for the general n goes by induction. The difference between the hermitian and the symmetric cases manifests itself in the fact that ξ_α 's are squares of complex or real variables, respectively. This gives different estimates for integrals of the type (20), resulting in different exponents in (16). $\square$

3. Sine kernel universality

Let $f(\lambda_1, \lambda_2, \dots, \lambda_N)$ denote the symmetric joint density function of the eigenvalues of the $N \times N$ Wigner matrix H . For any $k \geq 1$ we define the k -point correlation functions (marginals) by

$$p_N^{(k)}(\lambda_1, \dots, \lambda_k) = \int_{\mathbb{R}^{N-k}} f(\lambda_1, \lambda_2, \dots, \lambda_N) d\lambda_{k+1} \dots d\lambda_N.$$

We will use the notation $p_{N,GUE}^{(k)}$ and $p_{N,GOE}^{(k)}$ for the correlation functions of the GUE and GOE ensembles.

We consider the rescaled correlation functions about a fixed energy E under a scaling that guarantees that the local density is one. The sine-kernel universality for the GUE ensemble states that the rescaled correlation functions converge weakly to the determinant of the sine-kernel, $K(x) = \frac{\sin \pi x}{\pi x}$, i.e.

$$\frac{1}{[\varrho_{sc}(E)]^k} p_{N,GUE}^{(k)} \left(E + \frac{x_1}{N\varrho_{sc}(E)}, \dots, E + \frac{x_k}{N\varrho_{sc}(E)} \right) \rightarrow \det \left(K(x_\ell - x_j) \right)_{\ell,j=1}^k \quad (21)$$

as $N \rightarrow \infty$ for any fixed energy $|E| < 2$ in the bulk of the spectrum [13, 27]. Similar result holds for the GOE case; the sine kernel being replaced with a similar but somewhat more complicated universal function, see [26]. Our main result is that universality (21) holds for general hermitian or symmetric Wigner matrices after averaging in the energy E :

Theorem 3.1. [18] *Let H be an $N \times N$ symmetric or hermitian Wigner matrix with normalization defined at the beginning of Section 2. Suppose that the distribution ν of the matrix elements has subexponential decay (7). Let $k \geq 1$ and $O : \mathbb{R}^k \rightarrow \mathbb{R}$ be a continuous, compactly supported function. Then for any $|E| < 2$,*

we have

$$\lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{2\delta} \int_{E-\delta}^{E+\delta} dv \int_{\mathbb{R}^k} d\alpha_1 \dots d\alpha_k O(\alpha_1, \dots, \alpha_k) \quad (22)$$

$$\times \frac{1}{[\varrho_{sc}(v)]^k} \left(p_N^{(k)} - p_{N,\#}^{(k)} \right) \left(v + \frac{\alpha_1}{N\varrho_{sc}(v)}, \dots, v + \frac{\alpha_k}{N\varrho_{sc}(v)} \right) = 0,$$

where $\#$ stands for GOE or GUE for the symmetric or hermitian cases, respectively.

For the *hermitian case*, the first result on universality beyond the GUE was due to Johansson [24] (based upon [6]) under the condition that ν has a Gaussian component with a positive variance independent of N . His method was extended in [4] to Wishart matrices. The variance of the necessary Gaussian component was reduced to $N^{-3/4+\varepsilon}$ in [19] under the additional technical assumptions that the measure ν is smooth and it satisfies the logarithmic Sobolev inequality. The local statistics was identified via orthogonal polynomials. The Gaussian component assumption was first removed completely in [20] under the condition that the density of the probability measure ν is positive and it possesses a certain number of derivatives. Shortly after [20] appeared on the arXiv, the same result using a different method has been posted [35] without any regularity condition on ν provided that the third moment vanishes and ν is supported on at least three points. Combining the two methods, all conditions on ν apart from the subexponential decay (7) were removed in a short joint paper [21].

The methods of [20] and [35] both rely on the explicit formula of Brézin and Hikami [6], exploited also in [24], for the correlation functions of the Wigner matrix with Gaussian convolution. This formula reduces the problem to a saddle point analysis. The saddle points are identified by solving an equation involving the Stieltjes transform $m_N(z)$ (9) with $\eta = \text{Im } z$ corresponding to the variance of the Gaussian component: precise information on $m_N(z)$ for a smaller η implies that a smaller Gaussian component is sufficient.

In our work [20] we used the convergence of $m_N(z)$ to $m_{sc}(z)$ for very small $\eta = N^{-1+\varepsilon}$ established along the proof of Theorem 2.1. To remove this tiny Gaussian component, we have compared the local eigenvalue statistics of a given Wigner matrix H with that of $\widehat{H}_s + sV$ for which the saddle point analysis applies. Here $s^2 = \eta = N^{-1+\varepsilon}$ and the new Wigner matrix $\widehat{H}_s$ was chosen such that the law of $\widehat{H}_s + sV$ be very close to H . Since Gaussian convolution corresponds to running a heat flow on the matrix elements, $\widehat{H}_s$ could, in principle, be obtained by running the *reverse heat flow* on the elements of H . Although the reverse heat flow is undefined for most initial conditions, one can construct an approximation to the reverse heat flow that is well defined and yields $\widehat{H}_s$ with a required precision assuming sufficient smoothness on ν . Technically, we use Ornstein-Uhlenbeck process instead of the heat flow to keep the variance constant. We also mention that the result of [20] is valid for any fixed energy E , i.e. dv averaging in (22) is not necessary.

Tao and Vu [35] have directly compared local statistics of the Wigner matrix H and that of the matrix with order one Gaussian component for which Johansson

has already proved universality. Their main technical result [35, Theorem 15] states that the local eigenvalue statistics of two Wigner matrices coincide as long as the first four moments of their single site distributions match. It is then an elementary lemma from probability theory (cf. [35, Corollary 23] based upon [8]) to match to order four a given random variable with another random variable with a Gaussian component.

The proof of Theorem 3.1 for the *symmetric* case requires a new idea since the formula of Brézin and Hikami is not available. While the four moment theorem of [35] also applies to this case, there is no reference ensemble available. In the next sections we describe our new approach that proves universality for both hermitian and symmetric matrices without relying on any explicit formulae.

4. Dyson Brownian motion

The joint distribution of the eigenvalues $\mathbf{x} = (x_1, x_2, \dots, x_N)$ of the Gaussian ensembles is given by the following measure

$$\mu = \mu_N(d\mathbf{x}) = \frac{e^{-\mathcal{H}(\mathbf{x})}}{Z_\beta} d\mathbf{x}, \quad \mathcal{H}(\mathbf{x}) = N \left[\beta \sum_{i=1}^N \frac{x_i^2}{4} - \frac{\beta}{N} \sum_{i < j} \log |x_j - x_i| \right] \quad (23)$$

where $\beta = 1$ for GOE and $\beta = 2$ for GUE. For definiteness, we consider the $\beta = 1$ GOE case and we assume that the eigenvalues are ordered, i.e. μ is restricted to $\Sigma_N = \{\mathbf{x} \in \mathbb{R}^N : x_1 < x_2 < \dots < x_N\}$.

Suppose the matrix elements evolve according to the Ornstein-Uhlenbeck process on $\mathbb{R}$, i.e. the density of their distribution $\nu_t = u_t(x)dx$ satisfies

$$\partial_t u_t = \mathcal{L} u_t, \quad \mathcal{L} = \frac{1}{2} \frac{\partial^2}{\partial x^2} - \frac{x}{2} \frac{\partial}{\partial x}. \quad (24)$$

The Ornstein-Uhlenbeck process (24) induces a stochastic process, the Dyson Brownian motion, on the eigenvalues with a generator given by

$$L = \sum_{i=1}^N \frac{1}{2N} \partial_i^2 + \sum_{i=1}^N \left(-\frac{\beta}{4} x_i + \frac{\beta}{2N} \sum_{j \neq i} \frac{1}{x_i - x_j} \right) \partial_i \quad (25)$$

acting on $L^2(\mu)$. The measure μ is invariant and reversible with respect to the dynamics generated by L . Let

$$D(f) = - \int f L f d\mu = \sum_{j=1}^N \frac{1}{2N} \int (\partial_j f)^2 d\mu \quad (26)$$

be the corresponding Dirichlet form. Denote the distribution of the eigenvalues at time t by $f_t(\mathbf{x})\mu(d\mathbf{x})$. Then f_t satisfies

$$\partial_t f_t = L f_t \quad (27)$$

with initial condition f_0 given by the eigenvalue density of the Wigner ensemble. Dyson Brownian motion is the corresponding system of stochastic differential equations for the eigenvalues $\mathbf{x}(t)$ that is given by (see, e.g. Section 12.1 of [23])

$$dx_i = \frac{dB_i}{\sqrt{N}} + \left[-\frac{\beta}{4}x_i + \frac{\beta}{2N} \sum_{j \neq i} \frac{1}{x_i - x_j} \right] dt, \quad 1 \leq i \leq N, \quad (28)$$

where $\{B_i : 1 \leq i \leq N\}$ is a collection of independent Brownian motions. Note that the equations (27) and (28) are defined for any $\beta \geq 1$, independently of the original matrix models. Our main technical result (Theorem 5.1) holds for general $\beta \geq 1$.

5. Local Relaxation Flow

The Hamiltonian of the invariant measure μ of the Dyson Brownian motion is convex, with Hessian bounded from below

$$\text{Hess } \mathcal{H} \geq \frac{\beta N}{2}$$

on the set Σ_N . By the Bakry-Emery criterion, this guarantees that μ satisfies the logarithmic Sobolev inequality and the relaxation time to equilibrium is of order one (note the additional $1/N$ factor in the Dirichlet form (26) that rescales time).

We now introduce the local relaxation measure, which has the local statistics of GOE (or GUE) but generates a faster decaying dynamics. Let γ_j be the semicircle location of the j -th eigenvalue, i.e.

$$\gamma_j = n_{sc}^{-1}(j/N), \quad n_{sc}(E) := \int_{-\infty}^E \varrho_{sc}(x) dx.$$

We fix a regularization parameter $\eta \ll 1$ and we replace the interaction potential between x_j and far away particles by a regularized mean field potential

$$W_j(x) = -\frac{\beta}{N} \sum_{k: |k-j| \geq N\eta} \log(|x - \gamma_k| + \eta) \quad (29)$$

Strictly speaking, $W_j(x)$ is defined by this formula only in an interval of size $N\eta$ about γ_j and we use a quadratic extension beyond, but we leave this technicality aside.

The local relaxation measure $\omega_N = \omega$ is a Gibbs measure defined by the Hamiltonian

$$\tilde{\mathcal{H}} = N \sum_{j=1}^N \left\{ \beta \frac{x_j^2}{4} + W_j(x_j) \right\} + \beta \sum_{i < j} \log |x_i - x_j| - \frac{\beta}{2} \sum_i \sum_{j: |j-i| > N\eta} \log(|x_i - x_j| + \eta).$$

We often write $\omega = \psi \mu$ where ψ is the Radon-Nikodym derivative. The local relaxation flow is defined to be the reversible dynamics w.r.t. ω characterized by the

generator $\tilde{L}$ defined by

$$\int f \tilde{L} g d\omega = -\frac{1}{2N} \sum_j \int \partial_j f \partial_j g d\omega. \quad (30)$$

Explicitly, $\tilde{L}$ is given by

$$L = \tilde{L} + \sum_j b_j \partial_j, \quad b_j = \frac{1}{N} \sum_{k: |k-j| > N\eta} \frac{\text{sgn}(x_j - x_k)}{|x_j - x_k| + \eta} + W'_j(x_j), \quad (31)$$

Simple calculation shows that the mean field potential is uniformly convex with

$$\inf_j \inf_{x \in \mathbb{R}} W''_j(x) \geq c\eta^{-1/3}. \quad (32)$$

This will guarantee that the relaxation time to equilibrium ω for the $\tilde{L}$ dynamics is of order $\eta^{-1/3}$.

We recall the definition of the relative entropy of with respect to any probability measure $d\lambda$

$$S_\lambda(f) = \int f \log f d\lambda, \quad S_\lambda(f|\psi) = \int f \log(f/\psi) d\lambda$$

Our main technical result is the following theorem that states that the relaxation time τ for specific local observables is much shorter than order one.

Theorem 5.1 (Universality of Dyson Brownian Motion for Short Time).

Suppose that $S_\mu(f_0|\psi) \leq CN^m$ for some m fixed. Let $\tau = \eta^{1/3}N^\varepsilon$ with some $\varepsilon > 0$ and assume that $\eta \geq N^{-3/55+\varepsilon}$. Assume that there is a positive number Λ such that

$$\sup_{0 \leq t \leq \tau} N \sum_j \int b_j^2 f_t d\mu \leq C\eta^{-2}\Lambda. \quad (33)$$

Let G be a bounded smooth function with compact support. Then for any fixed $n \geq 1$ and $J \subset [1, \dots, N]$ we have

$$\left| \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) f_\tau d\mu - \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) d\mu \right| \leq \sqrt{\frac{C\Lambda}{N^{1-\varepsilon}\eta^{5/3}}}.$$

We emphasize that Theorem 5.1 applies to *all* $\beta \geq 1$ ensembles and the only assumption concerning the distribution f_t is in (33). In case of the original Wigner ensembles $\beta = 1, 2$, the critical constant Λ can be estimated under an additional assumption.

Lemma 5.1. Let f_0 be the joint density of the eigenvalues of a Wigner matrix. Suppose that the measure $d\nu$ of its single site distribution satisfies the logarithmic Sobolev inequality. Then the constant Λ in (33) can be estimated as

$$\Lambda \leq C_\sigma \eta^{-2} N^{4/5+\sigma} \quad (34)$$

for any $\sigma > 0$.

For the proof of this lemma, we can estimate b_j as

$$|b_j| \leq \frac{1}{N} \sum_{k: |k-j| > N\eta} \left| \frac{\text{sgn}(x_j - x_k)}{|x_j - x_k| + \eta} - \frac{\text{sgn}(x_j - \gamma_k)}{|x_j - \gamma_k| + \eta} \right| \leq C\eta^{-2} \frac{1}{N} \sum_{k=1}^N |x_k - \gamma_k| \quad (35)$$

as long as x_k is sufficiently near γ_k so that $\text{sgn}(x_j - x_k) = \text{sgn}(x_j - \gamma_k)$ holds for $|j - k| > N\eta$. The average difference between x_k and $\mathbb{E}x_k$ can be estimated using the logarithmic Sobolev inequality for ν . The average of $|\mathbb{E}x_k - \gamma_k|$ is estimated in Proposition 4.2 of [17] that was a consequence of the local semicircle law. Combining these results with information on the lowest and largest eigenvalues [37], we can show that $\frac{1}{N} \sum_k |x_k - \gamma_k| \leq N^{-3/5+\varepsilon}$ and this yields (34). $\square$

Combining Lemma 5.1 with Theorem 5.1 and choosing η appropriately, we see that the local eigenvalue statistics of f_τ with $\tau \geq N^{-1/55+\varepsilon}$ coincides with that of the global equilibrium measure, i.e. with GOE or GUE. For hermitian matrices, the same statement was already proven in [19] even for $\tau \geq N^{-1+\varepsilon}$ by using Brézin-Hikami formula, but the current approach is purely analytical and it applies to symmetric matrices as well. Using the reverse heat flow argument, we can show that the local statistics of f_0 is also given by GOE or GUE assuming that the initial distribution ν is sufficiently smooth. The smoothness condition and the additional requirement that ν satisfies the logarithmic Sobolev inequality can be removed by applying the four moment theorem of [35].

6. Proof of Theorem 5.1

We first list the key new ideas of behind the proof of Theorem 5.1, then we formulate the corresponding results.

- I. The key concept is the introduction of the local relaxation flow (30) which has the following two properties: (1) The invariant measure for this flow, the local relaxation measure ω has the same local eigenvalue statistics as the GOE or GUE. (2) The relaxation time of the local relaxation flow is much shorter than that of the DBM, which is of order one.
- II. Suppose we have a density q w.r.t. ω that evolves with the local relaxation flow. Then, by differentiating the Dirichlet form w.r.t. ω we will prove that the difference between the local statistics of $q\omega$ and ω can be estimated in terms of the Dirichlet form of q w.r.t. ω . Hence if the Dirichlet form is small, the local statistics of $q\omega$ is independent of q .
- III. It remains to show that the Dirichlet form of $q = f_t\mu$ w.r.t. ω is small for t sufficiently large (but still much less than order one). To do that, we study the evolution of the entropy of $f_t\mu$ relative to ω . This provides estimates on the entropy and Dirichlet form which serve as inputs for the Step II to conclude the universality.

The first ingredient to prove Theorem 5.1 is the analysis of the local relaxation flow which satisfies the logarithmic Sobolev inequality and the following dissipation estimate.

Theorem 6.1. *Suppose (32) holds. Consider the equation*

$$\partial_t q_t = \tilde{L}q_t \quad (36)$$

with reversible measure ω . Then we have the following estimates

$$\partial_t D_\omega(\sqrt{q_t}) \leq -C\eta^{-1/3} D_\omega(\sqrt{q_t}) - \frac{1}{2N^2} \int \sum_{|i-j| \leq N\eta} \frac{1}{(x_i - x_j)^2} (\partial_i \sqrt{q_t} - \partial_j \sqrt{q_t})^2 d\omega, \quad (37)$$

$$\frac{1}{2N^2} \int_0^\infty ds \int \sum_{|i-j| \leq N\eta} \frac{1}{(x_i - x_j)^2} (\partial_i \sqrt{q_s} - \partial_j \sqrt{q_s})^2 d\omega \leq D_\omega(\sqrt{q_0}) \quad (38)$$

and the logarithmic Sobolev inequality

$$S_\omega(q) \leq C\eta^{1/3} D_\omega(\sqrt{q}) \quad (39)$$

with a universal constant C . Thus the time to equilibrium is of order $\eta^{1/3}$:

$$S_\omega(q_t) \leq e^{-Ct\eta^{-1/3}} S_\omega(q_0). \quad (40)$$

The proof follows the standard argument in [3] (used in this context in [19]). The key input is the following lower bound on the Hessian of $\tilde{\mathcal{H}}$

$$\frac{1}{2N^2} \left\langle \mathbf{v}, (\nabla^2 \tilde{\mathcal{H}}) \mathbf{v} \right\rangle \geq C\eta^{-1/3} \frac{1}{N} \|\mathbf{v}\|^2 + \frac{1}{2N^2} \sum_{|i-j| \leq N\eta} \frac{1}{(x_i - x_j)^2} (v_i - v_j)^2. \quad (41)$$

The first term is due to convexity of the mean field potential (32). The second term comes from the additional convexity of the local interaction and it corresponds to “local Dirichlet form dissipation”. The estimate (38) on this additional term plays a key role in the next theorem.

Theorem 6.2. *Suppose that the density q_0 satisfies $S_\omega(q_0) \leq CN^m$ with some $m > 0$ fixed. Let G be a bounded smooth function with compact support and let $J \subset \{1, 2, \dots, N\}$. Set $\tau = \eta^{1/3} N^\varepsilon$. Then for any $n \geq 1$ fixed we have*

$$\left| \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) d\omega - \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) q_0 d\omega \right| \leq C \sqrt{\frac{D_\omega(\sqrt{q_0})\tau}{N}} + Ce^{-cN^\varepsilon}. \quad (42)$$

Sketch of the proof. Let q_t satisfy

$$\partial_t q_t = \tilde{L}q_t$$

with an initial condition q_0 . Thanks to the exponential decay of the entropy on time scale $\tau \gg \eta^{1/3}$, see (40), difference between the local statistics w.r.t. $q_\tau \omega$ and

$q_\infty \omega = \omega$ is subexponentially small in N . To compare q_0 with q_τ , by differentiation, we have

$$\begin{aligned} \int \frac{1}{N} \sum_i G(N(x_i - x_{i+n})) q_\tau d\omega - \int \frac{1}{N} \sum_i G(N(x_i - x_{i+n})) q_0 d\omega \\ = \int_0^\tau ds \int \frac{1}{N} \sum_i G'(N(x_i - x_{i+n})) [\partial_i q_s - \partial_{i+n} q_s] d\omega. \end{aligned}$$

From the Schwarz inequality and $\partial q = 2\sqrt{q}\partial\sqrt{q}$ the last term is bounded by

$$\begin{aligned} 2 \left[\int_0^\tau ds \int \frac{1}{N^2} \sum_i \frac{1}{(x_i - x_{i+n})^2} [\partial_i \sqrt{q_s} - \partial_{i+n} \sqrt{q_s}]^2 d\omega \right]^{1/2} \\ \times \left[\int_0^\tau ds \int \sum_i G'(N(x_i - x_{i+n}))^2 (x_i - x_{i+n})^2 q_s d\omega \right]^{1/2} \leq C \sqrt{\frac{D_\omega(\sqrt{q_0})\tau}{N}}, \quad (43) \end{aligned}$$

where we have used (38) and $G'(N(x_i - x_{i+n}))^2 (x_i - x_{i+n})^2 \leq C/N^2$. $\square$

Notice if we use only the entropy dissipation and Dirichlet form, the main term on the right hand side of (42) will become $\sqrt{S\tau}$. Hence by exploiting the local Dirichlet form dissipation coming from the second term on the r.h.s. of (37), we gain the crucial factor $N^{-1/2}$ in the estimate.

The final ingredient to prove Theorem 5.1 is the following entropy and Dirichlet form estimates.

Theorem 6.3. *Suppose the assumptions of Theorem 5.1 hold. Let $\tau = \eta^{1/3} N^\varepsilon$ and let $g_t = f_t/\psi$ so that $S_\mu(f_t|\psi) = S_\omega(g_t)$. Then the entropy and the Dirichlet form satisfy the estimates:*

$$S_\omega(g_{\tau/2}) \leq C\eta^{-5/3}\Lambda, \quad D_\omega(\sqrt{g_\tau}) \leq C\eta^{-2}\Lambda. \quad (44)$$

Sketch of the proof. Recall that $\partial_t f_t = Lf_t$. The standard estimate on the entropy of f_t with respect to the invariant measure is obtained by differentiating it twice and using the logarithmic Sobolev inequality. The entropy and the Dirichlet form in (44) are, however, computed with respect to the measure ω . This yields the additional second term in the following identity [39] that holds for any probability density ψ_t :

$$\partial_t S_\mu(f_t|\psi_t) = -\frac{2}{N} \sum_j \int (\partial_j \sqrt{g_t})^2 \psi_t d\mu + \int g_t (L - \partial_t) \psi_t d\mu,$$

where $g_t = f_t/\psi_t$. In our application we set $\psi_t = \psi = \omega/\mu$, hence we have

$$\partial_t S_\omega(g_t) = -\frac{2}{N} \sum_j \int (\partial_j \sqrt{g_t})^2 d\omega + \int \tilde{L} g_t d\omega + \sum_j \int b_j \partial_j g_t d\omega.$$

Since ω is invariant, the middle term on the right hand side vanishes, and from the Schwarz inequality

$$\partial_t S_\omega(g_t) \leq -D_\omega(\sqrt{g_t}) + CN \sum_j \int b_j^2 g_t d\omega. \quad (45)$$

Together with (39) and (33), we have

$$\partial_t S_\omega(g_t) \leq -C\eta^{-1/3} S_\omega(g_t) + C\eta^{-2} \Lambda. \quad (46)$$

which, after integrating it from $t = 0$ to $\tau/2$, proves the first inequality in (44). The second inequality can be obtained from integrating (45) from $t = \tau/2$ to $t = \tau$ and using the monotonicity of the Dirichlet form in time. $\square$

Finally, we sketch the proof of Theorem 5.1. With the choice of $\tau = \eta^{1/3} N^\varepsilon$ and $q_0 = f_\tau/\psi$, Theorems 6.1, 6.2 and 6.3 directly imply

$$\begin{aligned} \left| \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) f_\tau d\mu - \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) d\omega \right| \\ \leq \sqrt{\frac{C\Lambda}{N^{1-\varepsilon}\eta^{5/3}}} + Ce^{-cN^\varepsilon}, \end{aligned} \quad (47)$$

i.e. the local statistics of $f_\tau \mu$ and ω are close. Clearly, equation (47) also holds for the special choice $f_0 = 1$ (for which $f_\tau = 1$), i.e. local statistics of μ and ω can also be compared. This completes the proof of Theorem 5.1. $\square$

Acknowledgements

This work is partially supported by the SFB-TR 12 Grant of the German Research Council.

References

- [1] Aizenman, M., and Molchanov, S.: Localization at large disorder and at extreme energies: an elementary derivation, *Commun. Math. Phys.* **157**, 245–278 (1993)
- [2] Anderson, P.: Absences of diffusion in certain random lattices, *Phys. Rev.* **109**, 1492–1505 (1958)
- [3] Bakry, D., Émery, M.: Diffusions hypercontractives. in: Séminaire de probabilités, XIX, 1983/84, **1123** Lecture Notes in Mathematics, Springer, Berlin, 1985, 177–206.
- [4] Ben Arous, G., Pécché, S.: Universality of local eigenvalue statistics for some sample covariance matrices. *Comm. Pure Appl. Math.* **LVIII**. (2005), 1–42.
- [5] Bleher, P., Its, A.: Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model. *Ann. of Math.* **150** (1999): 185–266.
- [6] Brézin, E., Hikami, S.: Correlations of nearby levels induced by a random potential. *Nucl. Phys. B* **479** (1996), 697–706, and Spectral form factor in a random matrix theory. *Phys. Rev. E* **55** (1997), 4067–4083.

- [7] Chen, T.: Localization lengths and Boltzmann limit for the Anderson model at small disorders in dimension 3. *J. Stat. Phys.* **120**, no.1-2, 279–337 (2005).
- [8] Curto, R., Fialkow, L.: Recursiveness, positivity and truncated moment problems. *Houston J. Math.* **17**, no. 4., 603–635 (1991).
- [9] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.* **52** (1999):1335–1425.
- [10] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: Strong asymptotics of orthogonal polynomials with respect to exponential weights. *Comm. Pure Appl. Math.* **52** (1999): 1491–1552.
- [11] Disertori, M., Pinson, H., Spencer, T.: Density of states for random band matrices. *Commun. Math. Phys.* **232**, 83–124 (2002)
- [12] Dyson, F.J.: A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.* **3**, 1191–1198 (1962).
- [13] Dyson, F.J.: Correlations between eigenvalues of a random matrix. *Commun. Math. Phys.* **19**, 235–250 (1970).
- [14] L. Erdős, M. Salmhofer, H.-T. Yau, Quantum diffusion for the Anderson model in scaling limit. *Ann. Inst. H. Poincaré* **8** no. 4, 621–685 (2007)
- [15] Erdős, L., Schlein, B., Yau, H.-T.: Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37**, No. 3, 815–852 (2008)
- [16] Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for Wigner random matrices. *Commun. Math. Phys.* **287**, 641–655 (2009)
- [17] Erdős, L., Schlein, B., Yau, H.-T.: Wegner estimate and level repulsion for Wigner random matrices. To appear in *Int. Math. Res. Notices* (2008). Preprint arxiv.org/abs/0811.2591
- [18] Erdős, L., Schlein, B., Yau, H.-T.: Universality of random matrices and local relaxation flow. Submitted to *Inv.Math.* arxiv.org/abs/0907.5605
- [19] Erdős, L., Ramirez, J., Schlein, B., Yau, H.-T.: *Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation*. Submitted to *Electr. J. Probab.* arxiv.org/abs/0905.2089
- [20] L. Erdős, J. Ramirez, S. Péché, B. Schlein and H.-T. Yau, Bulk universality for Wigner matrices. Submitted to *Comm. Pure Appl. Math.* Preprint arXiv.org:0905.4176.
- [21] L. Erdős, J. Ramirez, B. Schlein, T. Tao, V. Vu and H.-T. Yau, Bulk universality for Wigner hermitian matrices with subexponential decay. To appear in *Math. Res. Letters*. Preprint arXiv:0906.4400
- [22] J. Fröhlich and T. Spencer, *Absence of diffusion in the Anderson tight binding model for large disorder or low energy*, *Commun. Math. Phys.* **88**, 151–184 (1983)
- [23] Guionnet, A.: Large random matrices: Lectures on Macroscopic Asymptotics. École d’Été de Probabilités de Saint-Flour XXXVI-2006. Springer.
- [24] Johansson, K.: Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.* **215** (2001), no.3. 683–705.
- [25] Marchenko, V.A., Pastur, L.: The distribution of eigenvalues in a certain set of random matrices. *Mat. Sb.* **72**, 507–536 (1967).
- [26] Mehta, M.L.: Random Matrices. Academic Press, New York, 1991.
- [27] Mehta, M.L., Gaudin, M.: On the density of eigenvalues of a random matrix. *Nuclear Phys.* **18**, 420–427 (1960).
- [28] Minami, N.: Local fluctuation of the spectrum of a multidimensional Anderson tight binding model. *Commun. Math. Phys.* **177**, 709–725 (1996).
- [29] Pastur, L., Shcherbina M.: Bulk universality and related properties of Hermitian

matrix models. *J. Stat. Phys.* **130** (2008), no.2., 205-250.

- [30] Schenker, J.: Eigenvector localization for random band matrices with power law band width. *Commun. Math. Phys.* **290**, 1065-1097 (2009)
- [31] Schenker, J. and Schulz-Baldes, H.: Semicircle law and freeness for random matrices with symmetries or correlations. *Math. Res. Letters* **12**, 531-542 (2005)
- [32] Schenker, J. and Schulz-Baldes, H.: Gaussian fluctuations for random matrices with correlated entries. *Int. Math. Res. Not. IMRN* 2007, **15**, Art. ID rnm047.
- [33] Sinai, Y. and Soshnikov, A.: A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge. *Functional Anal. and Appl.* **32** (1998), no. 2, 114-131.
- [34] Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** (1999), no.3. 697-733.
- [35] Tao, T. and Vu, V.: Random matrices: Universality of the local eigenvalue statistics. Preprint arXiv:0906.0510.
- [36] Tao, T. and Vu, V.: Random matrices: Universality of local eigenvalue statistics up to the edge. Preprint arXiv:0908.1982
- [37] Vu, V.: Spectral norm of random matrices. *Combinatorica*, **27** (6) (2007), 721-736.
- [38] Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.* **62** (1955), 548-564.
- [39] Yau, H. T.: Relative entropy and the hydrodynamics of Ginzburg-Landau models, *Lett. Math. Phys.* **22** (1991) 63-80.

USES OF FREE PROBABILITY IN RANDOM MATRIX THEORY

ALICE GUIONNET

UMPA, CNRS UMR 5669, ENS Lyon, 46 allée d'Italie, 69007 Lyon, France

This article gives a short introduction to free probability theory and emphasizes its role as a natural framework to study random matrices with size going to infinity. We motivate the uses of free probability by a few applications, taken from joint works with Jones, Khrishnapur, Maurel-Segala, Shlyakhtenko and Zeitouni. The first concerns the study of non-normal matrices; we show that the empirical measure of the eigenvalues of non-normal random matrices whose law is invariant by conjugation converges and that the limit can be described by the so-called R -diagonal operators, which are random variables taken from free probability. This in particular generalizes the single ring theorem of Feinberg and Zee [1]. The second application deals with the topological expansion of Brézin, Itzykson, Parisi and Zuber [2]. We show that such expansion can be turned into an asymptotic expansion and that the limit can be constructed by 'free' Langevin dynamics. In particular, the convergence holds also in non-perturbative but convex situations.

Keywords: Random matrices, map enumeration.

1. Introduction

Large random matrices were first studied in theoretical physics by Wigner, in connection with quantum mechanics, as a model for the energy level of large nuclei. In an independent development in the early 1970s, Hugh Montgomery showed that, assuming the Riemann Hypothesis, and modulo some technicalities, the pair correlation function for the spacings between the (normalized) zeros of the Riemann zeta function on the critical line behaves like the pair correlation function for the (normalized) eigenvalues of a random matrix. At about the same time, G. 't Hooft and Brézin, Itzykson, Parisi and Zuber, showed that Gaussian random matrix integrals are generating functions for the enumeration of graphs embedded into surfaces which are sorted by their genus (the so-called topological expansion). Since that time an extraordinary variety of mathematical, physical and engineering systems have been related with Random Matrix Theory; it has emerged as an interdisciplinary scientific activity par excellence.

In this review, we shall concentrate on the global asymptotics of the spectrum of large random matrices, for instance reflected by the asymptotics of the (normalized) trace of powers of these matrices. We will argue that such asymptotics can be described in great generality in the framework of free probability and therefore that this theory is extremely useful to study large random matrices. We will illustrate this point by two problems from theoretical physics. The first concerns the study

of the spectrum of non-normal large random matrices which appears in several areas of physics [3–10]. It was shown by Feinberg and Zee [1] that the spectral distribution of certain ensembles of random non-normal converges to a radially symmetric deterministic measure whose support is a single ring. This result came as a big surprise because the support of the singular values of these ensembles can be as disconnected as wished. This question is related with the law of non-commutative variables called 'R-diagonal operators' in free probability. Relying on a study of such operators by Haagerup and Larsen [11], we can generalize and prove rigorously Feinberg and Zee's result. Another range of applications of free probability concerns the first order of the topological expansion of Brézin, Itzykson, Parisi and Zuber which relates the asymptotics of matrix integrals with the enumeration of planar graphs. Indeed, again the limit can be described in terms of free variables, which in turns allows us to prove that the topological expansion can be turned into an asymptotic expansion, that it holds in non-perturbative but convex situations and finally, under such convexity hypothesis, that it has some properties, such as the connectivity of the support of the limiting matrices.

Free probability emerged in the eighties when Voiculescu realized that certain questions appearing in operator algebra theory could be phrased in probabilistic terms. Of course, such a probability theory should be concerned with non-commutative random variables, as is non-commutative probability theory. However, free probability theory differs from the latter by the notion of freeness. Similarly to the classical notion of independence, freeness is defined by certain relations between traces of words. These two components are the basis for a probability theory for noncommutative variables where many concepts taken from probability theory such as the notions of laws, convolution, convergence in law, independence, central limit theorem, Brownian motion, entropy, and more can be naturally defined. For instance, the law of one self-adjoint variable is simply given by the traces of its powers (which generalizes the definition through moments of compactly supported probability measures on the real line), and the joint law of several self-adjoint non-commutative variables is defined by the collection of traces of words in these variables. The joint law of free variables is defined by the law of each of them and a condition on the joint moments which defines them uniquely from the moments of the marginals. Convergence in law just means that the trace of any word in the noncommutative variables converges towards the right limit.

About ten years later, Voiculescu showed that free probability is the right framework to consider the asymptotics of random matrices as their size go to infinity. More specifically, he proved that an m -tuple of random matrices whose eigenvectors are genuinely independent, namely with a covariance matrix following independent unitary Haar distributed random matrices, converges to an m -tuple of free random variables. From that point, many concepts from standard probability theory could be brought to free probability, once suitably transposed through random matrices. Hence, in some manner, many concepts in free probability are inspired from random matrices. But free probability also developed on its own, or in relation with operator

algebra theory, and provide now the natural framework to study the asymptotics of random matrices. We shall in the next section describe more precisely the basics of free probability.

2. Free probability theory

2.1. Basics of free probability

Free probability is a non-commutative probability theory with a notion of freeness very analogous to independence in classical probability theory. Variables in such a theory include finite size random matrices but also 'matrices with size going to infinity'. This notion represents the weak limit of matrices sequences. Namely m $N \times N$ Hermitian matrices $(M_1^N, \dots, M_m^N)$ converge to $(M_1, \dots, M_m)$ iff for all polynomial P

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr} (P(M_1^N, \dots, M_m^N)) =: \tau(P(M_1, \dots, M_m)).$$

The limit τ , when it exists, is just a linear form on the set of polynomials. It is analogous to a classical expectation. Indeed, because it is obtained as the limit of traces of matrices, it satisfies, for all polynomials P, Q in m non-commutative variables,

$$\tau(PQ) = \tau(QP) \quad \tau(PP^*) \geq 0 \quad \tau(I) = 1.$$

The second and third properties generalize the properties of positivity and mass of the classical expectation. τ is called a tracial state. As in classical probability, one likes to think about expectation of random variables and of laws of random variables. In fact, τ can be seen as the law of non-commutative variables in the sense that, if the matrices $(M_i^N, N \geq 0, 1 \leq i \leq m)$ are uniformly bounded for the operator norm by some constant R , one can construct a Hilbert space H with scalar product $\langle \cdot, \cdot \rangle$, a vector $\zeta \in H$ and bounded linear operators $(M_1, \dots, M_m)$ on H so that for all polynomial P ,

$$\tau(P) = \langle P(M_1, \dots, M_m)\zeta, \zeta \rangle. \quad (1)$$

This construction of H and $(M_1, \dots, M_m)$ is called the Gelfand-Naimark-Seigal construction. Note that in the case where $m = 1$, Riesz's theorem asserts that τ is a classical probability measure on $[-R, R]$. One way to construct M_1 as a bounded operator on a Hilbert space is to take $H = L^2(\tau)$, once quotiented by the left ideal $\{h \geq 0, \tau(h) = 0\}$, and to put M_1 to be the left multiplication by x , so that for all $h \in L^2(\tau)$, $M_1 h(x) = xh(x)$.

In the sequel, τ will be a linear form on the set of polynomials and we will assume that we have constructed a Hilbert space H so that (1) holds and our random variables all live in $B(H)$.

Definition 2.1. $\mathbf{X} = (X_1, \dots, X_m)$ and $\mathbf{Y} = (Y_1, \dots, Y_n) \in (B(H), \tau)$ are free iff for all polynomials $P_1, \dots, P_\ell$ and $Q_1, \dots, Q_\ell$ so that $\tau(P_i(\mathbf{X})) = 0$ and $\tau(Q_i(\mathbf{Y})) =$

0 for all i , we have

$$\tau(P_1(\mathbf{X})Q_1(\mathbf{Y}) \cdots P_\ell(\mathbf{X})Q_\ell(\mathbf{Y})) = 0.$$

Note that the joint law $(\tau(P(\mathbf{X}, \mathbf{Y})), P)$ is uniquely determined by the collection of the marginals $\tau(P(\mathbf{X}))$ and $\tau(Q(\mathbf{Y}))$, where Q, P runs over the set of polynomials in $\mathbf{X}$ or $\mathbf{Y}$.

The notion of freeness we just defined is related with the classical notion of freeness in groups. Indeed, let G be a group with generators $g_1, \dots, g_m$, a neutral element e and define

$$\tau(g) = 1_{g=e}, \quad \text{for } g \in G.$$

Then we claim that $g_1, \dots, g_m$ are free under τ iff they are free in the group G . Indeed, $g_1, \dots, g_m$ are free in G iff it is not possible that a non trivial word in $g_1, \dots, g_m$ is trivial. But a non trivial word is just a word which can be reduced into a product $P_1(g_{i_1}) \cdots P_k(g_{i_k})$ of words in the g_i , $i_{k+1} \neq i_k$. Because each of the word P_i in g_i is not trivial, $\tau(P_i) = 0$ and then we see that the condition that $P_1(g_{i_1}) \cdots P_k(g_{i_k})$ is not the neutral element exactly means that $\tau(P_1(g_{i_1}) \cdots P_k(g_{i_k})) = 0$ which is the freeness condition.

Next we show that freeness also appears when one considers random matrices with 'independent' basis of eigenvectors, asymptotically when the size of the matrices goes to infinity. Let $(U_1^N, \dots, U_m^N)$ be $N \times N$ independent matrices following the Haar measure on the unitary group and $A_1^N, \dots, A_m^N$ deterministic Hermitian matrices with spectral measures $\hat{\mu}_k^N = \frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j^N(A_k^N)}$ which converge to μ_i for $1 \leq i \leq m$ in moments, that is

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}(A_i^N)^k = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N (\lambda_j(A_i^N))^k = \int x^k d\mu_i(x).$$

Then, Voiculescu [12]

Theorem 2.1 (Voiculescu 91'). *Let $X_i^N = U_i^N A_i^N (U_i^N)^*$. For any polynomial P in m non-commutative variables,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}(P(X_1^N, \dots, X_m^N)) = \tau(P(X_1, \dots, X_m)) \quad a.s.$$

with $(X_1, \dots, X_m)$ free and with marginal distribution $(\mu_i)_{1 \leq i \leq m}$.

When (random) matrices converge and their limit are free, we say that they are asymptotically free. In fact, Theorem 2.1 applies to many standard matrix models, such as for instance the classical Gaussian ensembles. Indeed, the ensemble of Hermitian $N \times N$ matrices with independent Gaussian entries with covariance N^{-1} taken from the GUE is by definition invariant under unitary conjugation; hence a matrix taken from the GUE can be written as $X^N = U^N A^N (U^N)^*$ where A^N is the diagonal matrix with entries given by the eigenvalues of X^N and U^N an independent matrix following the Haar measure on the unitary group. By Wigner's theorem [13]

Theorem 2.2 (Wigner 58'). *For all integer number d*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}((X^N)^d) = \begin{cases} 0 & \text{if } d \text{ is odd,} \\ C_{\frac{d}{2}} & \text{otherwise,} \end{cases} \quad \text{a.s.}$$

where C_p denote the Catalan numbers. Moreover, the right hand side is also equal to $\int x^d d\sigma(x)$ with σ the semicircle law given by

$$d\sigma(x) = (2\pi)^{-1} 1_{|x| \leq 2} \sqrt{4 - x^2} dx.$$

In particular, if one considers m independent matrices taken from the GUE, they will be asymptotically free. More precisely

Corollary 2.1. *m independent matrices taken from the GUE $(X_1^N, \dots, X_m^N)$ converge to m free semicircle variables.*

This corollary in fact generalizes to independent matrices with independent centered entries with finite moments and covariance N^{-1} , see [14, Theorem 5.4.2]. From Theorem 2.2 and its corollary, it is no surprise that freeness can be used to describe many asymptotics of random matrix problems. Reciprocally, it is clear that probability theory can be brought to the non-commutative setting by applying probability concepts to random matrices and then taking the large N limit. We next show how a few classical notions of probability theory generalize in free probability.

2.2. Free Brownian motion

Following Corollary 2.1, we are going to define the free Brownian motion as the limit of the Hermitian Brownian motion. $\{H_t^N, t \geq 0\}$, the Hermitian Brownian motion, is a $N \times N$ Hermitian matrix valued process whose entries are i.i.d. complex Brownian motions.

$$H_t^N(k\ell) = \frac{B_t(k\ell) + i\tilde{B}_t(k\ell)}{\sqrt{2N}} \quad k < \ell, \quad H_t^N(kk) = \frac{B_t(kk)}{\sqrt{N}}.$$

It is not hard to check that for all $N \times N$ Hermitian matrix A such that $N^{-1} \text{Tr}(A^2) = 1$, $\{\text{Tr}(AH_t^N), t \geq 0\}$ follows a real Brownian motion.

From Corollary 2.1, and the scaling property of Brownian motion, we deduce

Corollary 2.2. *For all $t_1, \dots, t_p \in \mathbb{R}^+$, the following limit exists*

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[\frac{1}{N} \text{Tr}(H_{t_1}^N \cdots H_{t_p}^N) \right] =: \tau(S_{t_1} \cdots S_{t_p})$$

$(S_t, t \geq 0)$ is a continuous process with values in $B(H)$ with free increments distributed according to the appropriately rescaled semicircle law.

This result is a direct consequence of Corollary 2.1 since the increments of the Hermitian Brownian motion are independent matrices distributed according to the GUE. One of the great application of the classical Brownian motion is based on its

relation with the heat equation which governs its density, which allowed to get a new insight on the latter. Such a connection can be performed thanks to Itô's calculus which shows that the small variation of a function evaluated at the Brownian motion after a small time increment is given by the sum of the derivative of this function (multiplied by the increment of the Brownian motion) and the second derivative of this function (multiplied by the time increment). The last term reflects the lack of smoothness of the Brownian motion and provides the heat equation motion. A similar calculus holds with the free Brownian motion, but this time the additional term is given by a second order non-commutative differential, which differs from the Laplacian. For instance, if one evaluates the trace $v_k(t) := \tau((S_t)^k)$ of monomials in the free Brownian, it satisfies the Smoluchowski's type equation

$$\partial_t v_k(t) = \frac{k}{2} \sum_{p=0}^{k-2} v_p(t) v_{k-2-p}(t)$$

The differential calculus of the free Brownian motion can also be generalized to consider differential equation driven by the free Brownian motion.

Let $K : B(H)^m \rightarrow B(H)^m$ be a bounded Lipschitz function on $B(H)^m$

$$\max_{1 \leq i \leq m} \|K^i(X) - K^i(Y)\|_{B(H)} \leq \|K\|_L \max_{1 \leq i \leq m} \|X_i - Y_i\|_{B(H)} \wedge 1.$$

Assume also that $K_i(X_1, \dots, X_m)$ is self-adjoint for any m -tuple $(X_1, \dots, X_m)$ of self-adjoint elements of $B(H)$. Then, for any $X_0 \in B(H)^m$, there exists a unique solution $X_\cdot$ to

$$X_t^i = X_0^i + S_t^i + \int_0^t K^i(X_s) ds$$

with $(S^1, \dots, S^m)$ m free Brownian motions. Note again that X_t can be seen to be the limit of the $N \times N$ Hermitian matrices so that X_0 is the limit of the $N \times N$ Hermitian matrices X_0^N and $X_\cdot^N$ is the unique strong solution to the Langevin dynamics

$$X_t^{N,i} = X_0^{N,i} + H_t^{N,i} + \int_0^t K^i(X_s^N) ds$$

with i.i.d Hermitian Brownian motions $\{H^{N,i}, 1 \leq i \leq m\}$.

In the case where $m = 1$, K is just a bounded Lipschitz function on $\mathbb{R}$ and if we let τ_t be the spectral distribution of X_t (recall that τ_t is the probability measure on $\mathbb{R}$ so that $\tau_t(x^k) = \tau(X_t^k)$), Itô's calculus now gives, for any bounded twice continuously differentiable function f ,

$$\partial_t \int f(x) d\tau_t(x) = \frac{1}{2} \int \int \frac{f'(x) - f'(y)}{x - y} d\tau_t(x) d\tau_t(y) + \int f'(y) K(y) d\tau_t(y).$$

2.3. Free convolution

Functional analysis can also be developped in free probability theory. For instance one can wonder what is the distribution of $A + B$, A and B being free variables with a prescribed distribution. This is, by Theorem 2.1, the limit of the spectral measure of $A_N + U_N B_N U_N^*$ when N goes to infinity. A similar question can be asked about the distribution of the product AB , A and B being free variables. Because AB is not self-adjoint, the moments of AB do not give the spectral measure of AB but in fact those of $A^{\frac{1}{2}} B A^{\frac{1}{2}}$ when $A \geq 0$. Indeed, $\tau((A^{\frac{1}{2}} B A^{\frac{1}{2}})^n)$ equals $\tau((AB)^n)$ since τ is tracial and the spectral measure of $A^{\frac{1}{2}} B A^{\frac{1}{2}}$ is given by its moments since it is self-adjoint. Hence, the law of AB describes the asymptotics of the spectral measure of $A_N^{\frac{1}{2}} U_N B_N U_N^* A_N^{\frac{1}{2}}$ as N goes to infinity and the spectral measure of A_N (resp. B_N) converges to μ_A (resp. μ_B) and U_N independent unitary.

The law of $A+B$ (resp. AB) is denoted $\mu_A \boxplus \mu_B$ (resp. $\mu_A \boxtimes \mu_B$). These probability measures are described by the R -transform and the S -transform respectively, which play the same role as a log-Fourier transform for the standard convolution. For example, we put

$$S_C(z) := \frac{1+z}{z} m_C^{-1}(z)$$

if $m_C(z) = \sum_{n \geq 1} \mu_C(x^n) z^n$. S_C is well defined at list when $\mu_C(x) \neq 0$ since then m_C^{-1} exists at list in a neighborhood of the origin by the implicit function theorem. Moreover, the knowledge of S_C on a set with accumulation points defines uniquely m_C and therefore the law μ_C . It is then known [14, Lemma 5.3.30] that, at list for small z 's,

$$S_{A^{\frac{1}{2}} B A^{\frac{1}{2}}}(z) = S_A(z) S_B(z).$$

Free probability theory offers many other interesting developments [14–17].

3. Single ring theorem

This section deals with one application of free probability results to the analysis of large non-normal random matrices. We consider the ensemble of random, non-normal $N \times N$ matrices with law

$$dP_N(X_N) = \frac{1}{Z_N} e^{-N \text{Tr}(V(X_N X_N^*))} dX_N$$

where $dX_N = \prod_{1 \leq i, j \leq N} d\Re(X_N(ij)) d\Im(X_N(ij))$ is the Lebesgue measure on the set of $N \times N$ matrices with complex entries. V is a polynomial going to infinity at infinity so that Z_N is finite for each N . We consider the eigenvalues $\{\lambda_i^N\}_{1 \leq i \leq N}$ of X_N and their empirical measure $L_N^V = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i^N}$. The following theorem was proved, albeit not entirely rigorously, by Feinberg and Zee [1].

Theorem 3.1 (Feinberg-Zee 97'). *Assume V is a polynomial. Then: L_N^V converges to a deterministic, rotationally invariant μ_V whose support consists of a*

single ring: there exists constants $0 \leq a < b < \infty$ so that

$$\text{supp}(\mu_V) = \{re^{i\theta} : a \leq r \leq b, \theta \in [0, 2\pi[\}.$$

If $V(x) = gx^2 + mx$, phase transition occurs when the support changes from a disc to an annulus.

This result is rather surprising since one could imagine a priori that if the potential presents sufficiently deep attracting wells, the eigenvalues of X_N should concentrate in these wells so that the support of the limiting spectral measure should be disconnected. This is true when one considers the singular values of X_N , that is the eigenvalues of $\sqrt{X_N X_N^*}$. Let $\sigma_1^N \geq \sigma_2^N \geq \dots \geq \sigma_N^N \geq 0$ denote the singular values of X_N . Their joint distribution is

$$\frac{1}{Z_N} \prod_{i < j} [(\sigma_i^N - \sigma_j^N)(\sigma_i^N + \sigma_j^N)]^2 e^{-N \sum_{i=1}^N V((\sigma_i^N)^2)} \prod_{i=1}^N \sigma_i^N d\sigma_i^N$$

One easily deduces (see e.g. [14, Section 2.6]) the convergence of the empirical measure of the singular values to the probability measure σ_V which minimizes

$$\mu \rightarrow \int V(x^2) d\mu(x) - \int \log |x^2 - y^2| d\mu(x) d\mu(y)$$

on the set of probability measures on the half line $\mathbb{R}^+$. If V has k sufficiently deep wells, it is easy to show that the support of this minimizer will consist of at list k intervals!

The surprising connectivity of the support of non-normal operators whose singular values have a very disconnected support was already observed in free probability by Haagerup and Larsen [11]. Indeed, note that under the Feinberg-Zee model, $X_N = U_N D_N V_N$ with D_N a real diagonal matrix with converging spectral distribution ($D_N = \text{diag}(\sigma^N)$) and independent U_N, V_N , unitary matrices following Haar distribution. It is natural to wonder what can be said about the convergence of the empirical measure $L_{X_N} = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i^N}$ of the eigenvalues of $X_N = U_N D_N V_N$ when D_N is a real diagonal matrix with converging spectral distribution and independent Haar distributed matrices U_N, V_N , independent of D_N .

A similar question was considered in the framework of free probability. Indeed, Nica and Speicher [17] defined the so-called R -diagonal operators, which are operators which can be decomposed as $R = UD$ with D Hermitian and U unitary, U and D being free. Note that the distribution of the eigenvalues of UDV and UD is the same if U, V are two free unitary variables. The distribution of such operators was studied by Haagerup and Larsen [11].

Theorem 3.2 (Haagerup-Larsen 00'). *Take $X_D := UDV$ or UD , U, V, D free, D self-adjoint with law μ_D , U, V unitaries. The Brown measure (continuous analogue of the spectral measure) of X_D is rotation invariant and radially supported*

on an annulus

$$\mu_{X_D}(B(0, f(t))) = t$$

where $f(t) = 1/\sqrt{S_{D^2}(t-1)}$.

$$\text{supp}(\mu_{X_D}) = \{re^{i\theta}, r \in [(\mu_D(x^{-2}))^{-\frac{1}{2}}, (\mu_D(x^2))^{\frac{1}{2}}], \theta \in [0, 2\pi[\}$$

The Brown measure of a non-normal operator X is given as follows. For $z \in \mathbb{C} \setminus \mathbb{R}$, let ν^z be the spectral measure of $(z - X)(z - X)^*$. Since $(z - X)(z - X)^*$ is self-adjoint, ν^z can easily be computed from its moments for instance. Then, the Brown measure of X is the compactly supported probability measure μ_X on $\mathbb{C}$ so that for any smooth compactly supported function ψ

$$\int \psi(z) d\mu_X(z) = \frac{1}{4\pi} \int_{\mathbb{C}} \Delta\psi(z) \int \log |x| d\nu^z(x) dz.$$

One way to see that the Brown measure is related to our problem is the Green formula which reads, if $(\lambda_i^N)_{1 \leq i \leq N}$ are the eigenvalues of X_N ,

$$\begin{aligned} \sum_{i=1}^N \psi(\lambda_i^N) &= \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z) \log \left| \prod_{i=1}^N (z - \lambda_i^N) \right| dz \\ &= \frac{1}{4\pi} \int_{\mathbb{C}} \Delta\psi(z) \left(\sum_{i=1}^N \log |z - \lambda_i^N|^2 \right) dz \\ &= \frac{N}{4\pi} \int_{\mathbb{C}} \Delta\psi(z) \int \log |x| dL_{(z - X_N)(z - X_N)^*}(x) dz \end{aligned} \quad (2)$$

where $L_{(z - X_N)(z - X_N)^*}(x)$ is the spectral measure of the self-adjoint operator $(z - X_N)(z - X_N)^*$. Taking $X_N = U_N D_N V_N$, we know by Theorem 2.1 that the spectral measures of the Hermitian matrices $(z - X_N)(z - X_N)^*$ converge to the law of $(z - UDV)(z - UDV)^*$. Hence, we expect $N^{-1} \sum_{i=1}^N \delta_{\lambda_i^N}$ to converge to the Brown measure of UDV , U, D free unitary variables, free with D . This is the content of the following theorem [18]

Theorem 3.3 (G., Krishnapur, Zeitouni '09). *Take $X_N = U_N D_N V_N$. Assume some technical conditions on D_N and that $L_{D_N} = \frac{1}{N} \sum \delta_{D_N(ii)}$ converges to a probability measure μ_D on $\mathbb{R}^+$ so that for some $\kappa, \kappa' > 0$, all $z \in \mathbb{C}^+$ so that $\Im z \geq N^{-\kappa'}$,*

$$\left| \frac{1}{N} \sum_{i=1}^N \frac{1}{z - D_N(ii)} - \int \frac{1}{z - x} d\mu_D(x) \right| \leq \frac{1}{N^\kappa |\Im z|}.$$

Then $L_{X_N} = \frac{1}{N} \sum \delta_{\lambda_i^N}$ converges weakly to μ_{X_D} in probability, i.e for any bounded continuous function f on $\mathbb{C}$, $N^{-1} \sum_{i=1}^N f(\lambda_i^N)$ converges to $\int f(z) d\mu_{X_D}(z)$ in probability.

As a corollary of Theorem 3.3, we prove the the Feinberg-Zee “single ring theorem”.

Corollary 3.1. *Let V denote a polynomial with positive leading coefficient. Let the N -by- N complex matrix X_N be distributed according to the law*

$$dP_N(X) = \frac{1}{Z_N} \exp(-N \operatorname{Tr} V(XX^*)) dX,$$

where Z_N is a normalization constant and dX the Lebesgue measure on N -by- N complex matrices. Then, the spectral measure L_{X_N} of X_N with law P_N converges weakly in probability to μ_{X_D} as in Theorem 3.3 with $\Theta = \sigma_V$.

The proof of Theorem 3.3 is not straightforward since Theorem 2.1 only guarantees the weak convergence of $L_{(z-X_N)(z-X_N)^*}(x)$ whereas $\log|x|$ is not bounded. Hence, a complete proof requires to control the small eigenvalues of $(z - X_N)(z - X_N)^*$. Such a problem was also encountered in the case of a $N \times N$ matrix X_N with i.i.d entries. When the entries are Gaussian, Ginibre [19] obtained the joint distribution of the eigenvalues of X_N , from which it was proved by Mehta [20] that the spectral measure of X_N converges in probability to the so-called circular law, that is the uniform measure on the disk with radius one. Extending this result to non Gaussian entries happened to be a difficult task. The idea to use the Green formula (2) is due to Girko [21], who however did not address the difficulties related with the unboundedness of the logarithm. The circular law was proved under some conditions by Bai [22] and finally, in full generality, by Gotze and Tikhomirov [23] and Tao and Vu [24], by dealing with this question.

4. Enumeration of maps

The uses of matrix integrals as generating functions for the enumeration of graphs in physics and the so-called topological expansion are diverse. Let us give a few examples: the enumeration of triangulations following Brézin, Itzykson, Parisi and Zuber, the enumeration of meanders (Di Francesco et al), the study of loop configurations and the $O(n)$ model (Eynard, Kostov ...), the application to knots theory (Zuber, Zinn Justin ...), the relation with algebraic geometry and topological string theory (the famous Dijkgraaf-Vafa conjecture states that Gromov-Witten invariants generating functions should be matrix integrals), Harer and Zagier (1986) in their article on the Euler characteristic of the moduli space of curves, and the famous work of Kontsevich. It became a cornerstone in free probability when Voiculescu (1984) found out that the combinatorics of moments of several independent matrices are, when their size goes to infinity, the same as the combinatorics of free variables. Since then, random matrices and their combinatorics appeared as a central tool in free probability (see e.g. the recent book of Nica and Speicher [17]). In this section, we shall precise some elements of the relation between these two fields. Let us first recall the key result of Brézin, Itzykson, Parisi and Zuber [2].

Theorem 4.1 (Brézin, Itzykson, Parisi and Zuber 78'). *Let $V(X_1, \dots, X_m) = \sum \beta_i q_i(X_1, \dots, X_m) + \frac{1}{2} \sum X_i^2$ with words $(q_i)_{1 \leq i \leq n}$ and*

$$d\mu_V^N(X_1, \dots, X_m) = (Z_V^N)^{-1} e^{-N \text{Tr}(V(X_1, \dots, X_m))} dX_1 \dots dX_m.$$

For any monomial P , we have as a formal expansion in β_i and N

$$\int \left[\frac{1}{N} \text{Tr}(P(X_1, \dots, X_m)) \right] d\mu_V^N = \sum_{g \geq 0} \frac{1}{N^{2g}} \tau_{\beta, g}(P) \quad (3)$$

with, for some integer numbers $M_g((1, P), (k_i, q_i))$,

$$\tau_{\beta, g}(P) = \sum_{k_1, \dots, k_n \geq 0} \prod_{i=1}^p \frac{(-\beta_i)^{k_i}}{k_i!} M_g((1, P), (k_i, q_i))$$

The integer number $M_g((1, P), (k_i, q_i))$ is the number of certain graphs that we now describe.

First, let us define what is a polygon of type q for a word q in m letters $(X_i, 1 \leq i \leq m)$. It is a polygon drawn on the sphere with one marked side and colored sides constructed as follows. We associate to each index $i, 1 \leq i \leq m$, a color (called 'color i ') and to each letter $X_i, 1 \leq i \leq m$ a side with color i . A polygon of type $q = X_{i_1} \dots X_{i_k}$ is a polygon embedded in the sphere by drawing the sides corresponding to the letters of q successively; following the orientation of the sphere, we first draw a side with color i_1 , then of color i_2 until the end where the loose end of the side of color i_k is glued with the loose end of the side with color i_1 . $M_g((k_i, q_i)_{1 \leq i \leq p})$ is the number of coverings (also called maps) of a surface with genus g by k_i polygons of type $q_i, 1 \leq i \leq p$ where only sides of the same color can be glued together. The counting is done for labelled sides. Such enumeration question is highly non trivial, in particular when polygons are colored. They are related at criticality with statistical models on $\mathbb{Z}^2$ by the Knizhnik-Polyakov-Zamolodchikov relation [25].

In their seminal article [2], Brézin, Itzykson, Parisi and Zuber used their result to count some planar maps, that is coverings of the sphere, by estimating the related matrix integrals. This assumes that the large N limit can be taken in (3). This point was justified recently in a series of papers [26–31].

Theorem 4.2. *Let $V = \sum \beta_i q_i(X_1, \dots, X_m) + \frac{1}{2} \sum X_i^2$ with words $(q_i)_{1 \leq i \leq n}$ and*

$$d\mu_V^N(X_1, \dots, X_m) = (Z_V^N)^{-1} e^{-N \text{Tr}(V(X_1, \dots, X_m))} dX_1 \dots dX_m.$$

For any $\ell > 0$, if the β_i 's are small enough and V strictly convex, then for any word P ,

$$\int \left[\frac{1}{N} \text{Tr}(P(X_1, \dots, X_m)) \right] d\mu_V^N = \sum_{g=0}^{\ell} \frac{1}{N^{2g}} \tau_{\beta, g}(P) + o\left(\frac{1}{N^{2\ell}}\right) \quad (4)$$

with, for interesting integer numbers $M_g((1, P), (k_i, q_i))$,

$$\tau_{\beta, g}(P) = \sum_{k_1, \dots, k_n \geq 0} \prod_{i=1}^p \frac{(-\beta_i)^{k_i}}{k_i!} M_g((1, P), (k_i, q_i))$$

Here, V is strictly convex iff there exists $c > 0$ so that for any $p \times p$ Hermitian matrices $X_1, \dots, X_m$, any $p \in \mathbb{N}$, $\text{Tr} V(X_1, \dots, X_m)$ is real and

$$\text{Hess}(\text{Tr} V(X_1, \dots, X_m)) \geq cI.$$

(4) implies in particular that the free energy $\frac{1}{N^2} \log Z_V^N$ converges as N goes to infinity under the hypotheses of the theorem. We prove below, by using free probability ideas, that such a convergence extends to the case where V is strictly convex or even “locally strictly convex” provided we add a cutoff. If V is just some polynomial going to infinity at infinity so that the free energy is well defined, it is still an open problem to establish this convergence. The idea to extend the convergence to this convex situation is to use dynamics and type of Monte-Carlo approximation argument. More precisely, in strictly convex situations, it is well known that Langevin dynamics converge as time goes to infinity to their unique equilibrium measure, whatever is their initial distribution. This is true at the level of large matrices but also at the level of free probability limit. The idea is thus to approximate Gibbs measures by their dynamics, use that they converge to the free dynamics as dimension goes to infinity, and then that the free dynamics are close to their equilibrium state when time is large. In other words, we want to show that the diagram below is approximately commutative

$$\begin{array}{ccc} X_t^N = H_t^N - \frac{1}{2} \int_0^t DV(X_s^N) ds & \xrightarrow{t \rightarrow \infty} & \mu_V^N \\ \downarrow & & \downarrow \\ N \rightarrow \infty & & N \rightarrow \infty \\ \downarrow & & \downarrow \\ X_t = S_t - \frac{1}{2} \int_0^t DV(X_s) ds & \xrightarrow{t \rightarrow \infty} & \tau_{\beta, 0} \end{array}$$

Of course, this approach can only be valid because the above approximations hold uniformly, with constants depending only on the convexity of the potential.

We say that V is locally strictly convex iff $\exists c > 0$

$$\text{Hess}(\text{Tr} V(X_1, \dots, X_m)) \geq cI$$

on the set of self-adjoint operators $X_1, \dots, X_m$ (in a C^* algebra) so that $\|X_i\|_\infty \leq L(c)$ for some $L(c)$. We let

$$d\mu_V^{N, L}(X_1, \dots, X_m) = (Z_V^{N, L})^{-1} \mathbf{1}_{\|X_i\|_\infty \leq L} e^{-N \text{Tr}(V(X_1, \dots, X_m))} dX_1 \dots dX_m.$$

We then can prove [32]

Theorem 4.3. *If $V = \frac{1}{2} \sum X_i^2 + \sum \beta_i q_i$ is “locally strictly convex”, for L large enough (but finite if V is not globally strictly convex),*

$$\lim_{N \rightarrow \infty} \int \frac{1}{N} \text{Tr}(P(X_1, \dots, X_m)) d\mu_V^{N,L} = \tau_{\beta,0}(P)$$

with $\tau_{\beta,0}(P)$ the analytic extension of the generating function for the enumeration of coverings of the sphere.

The key step to prove this theorem is to show the convergence of the free dynamics to the tracial state $\tau_{\beta,0}$, which in turns also gives some non trivial properties of the latter.

Theorem 4.4. *Let $V = \frac{1}{2} \sum X_i^2 + \sum \beta_i q_i$ be “locally strictly convex”, and X_t be the solution of*

$$X_t^i = S_t^i - \frac{1}{2} \int_0^t D_i V(X_s) ds.$$

Then

- $\tau(P(X_t))$ converges, as $t \rightarrow \infty$, to $\tau_{\beta,0}(P(X))$.
- For small β_i ’s,

$$\tau_{\beta,0}(P) = \sum_{k_1, \dots, k_n \geq 0} \prod_{i=1}^p \frac{(-\beta_i)^{k_i}}{k_i!} M_0((1, P), (k_i, q_i))$$

and the above series converges absolutely. $\tau_{\beta,0}(P)$ extends as an analytic function on the domain of local strict convexity of the potential.

- X_i has a connected spectrum under $\tau_{\beta,0}$, and in fact more precisely any polynomial $P(X_1, \dots, X_m)$ of $X_1, \dots, X_m$ under $\tau_{\beta,0}$ is an operator with connected spectrum.

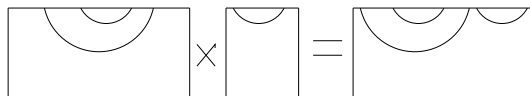
The last point is an amazing application of operator algebra theory. Indeed, we show that $\tau_{\beta,0}$ is the law of the solution of the free diffusion as time goes to infinity and in fact that the convergence holds for the operator norm. This in turn guarantees that this limit belongs to the C^* -algebra generated by the free increments of the free Brownian motion. The result follows since it is well known [33] that such a C^* -algebra does not contain a projection.

The fact that $\tau_{\beta,0}$ can be constructed as a limit of matrix models or free dynamics can also be used

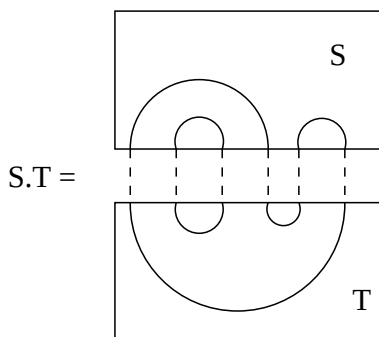
- to compute $\tau_{\beta,0}(P)$ (see the enumeration of triangulations [2]),
- to show that some generating functions of combinatorial numbers are tracial states.

We next consider the second application. Even though this point can also be proved by combinatorial arguments, constructing matrix models for these enumeration questions often appear to be a short cut to prove that generating functions of

interesting numbers are indeed tracial state. The goal of this application is to construct II_1 factors, and more precisely towers of factors with prescribed index. Recall that factors are von Neumann algebras (that is weakly closed algebras of bounded operators on a Hilbert space equipped with an involution and a neutral element) with a trivial center. They are said to be of type II_1 if they are equipped with a tracial state. A tower of factors is a sequence of factors $(M_n)_{n \geq 0}$ which are embedded in each other ($M_n \subset M_{n+1}$). The index $[M_n; M_{n+1}]$ measures somehow the 'size' of M_{n+1} with respect to M_n as follows. It can be seen that M_{n+1} is generated by M_n and a projection e_{n+1} and then for all $x \in M_n$, $\text{tr}(xe_{n+1}) = \lambda \text{tr}(x)$ with $\lambda = 1/[M_n; M_{n+1}]$. It was shown by Jones [34] that the index can only take the values $\{4 \cos^2 \pi/n, n = 3, 4, \dots\} \cup [4, \infty)$. With Jones and Shlyakhtenko we have constructed a tower of subfactors for any possible values of the index, based on planar algebra structure (and in fact mainly Temperley-Lieb algebras). Temperley-Lieb elements are boxes containing non-intersecting strings. We can endow this set with the multiplication given by simply drawing the drawings next to each other



We also endow this algebra by the involution which is given by taking the symmetric picture of the element. We denote $S.R$ the drawing obtained, for two Temperley-Lieb diagrams S, R with the same number of boundary points, by drawing these two diagrams in front of each other and gluing the boundary points pairwise by straight lines.



We then obtain a collection of non intersecting loops (two in the above picture). We then consider the trace given by

$$\tau(S) = \sum_{R \in \text{TL}} \delta^{\# \text{ loops in } S.R}$$

where the exponent in δ is the number of loops in the drawing defined by $S.R$ and we sum over all Temperley-Lieb diagrams R with the same number of boundary points than S . The next result proves [35] that if we take the weak completion of

the resulting algebra we obtain a factor. Moreover, we can construct a tower by considering the k th multiplication to be given by gluing by embedded arches the k nearest neighbouring boundary points of the two elements and defining similarly a trace by summing over Temperley-Lieb elements except for the k boundary points at the two extremes which are glued by embedded arches.

Theorem 4.5 (G-Jones-Shlyakhtenko 07'). *Take*

$$\delta \in I := \{2 \cos(\frac{\pi}{n})\}_{n \geq 4} \cup]2, \infty[$$

- τ is a tracial state.

- The corresponding von Neumann algebra is a factor. A tower of factors with index δ^2 can be built.

The fact that τ is a tracial state was first proved by using matrix models [35] but then a combinatorial proof was given [36]. The matrix models approach follows the idea of the planar algebra of a graph [37]. In fact, the idea to get the construction for integer values of δ is to use the embedding from Temperley-Lieb diagrams into the set of polynomials in δ variables as follows. Suppose that we are given a box B with $2k$ boundary points. Assume also that there are k non-crossing curves inside B which connect pairs of boundary points together, hence yielding a Temperley-Lieb element. Let π be the associated non-crossing pairing of $\{1, \dots, 2k\}$ and denote $p \approx \ell$ if (p, ℓ) is a block of π . We associate to B the non commutative polynomial

$$P_B(X_1, \dots, X_n) := \sum_{\substack{1 \leq i_1, \dots, i_{2k} \leq n \\ i_\ell = i_p \text{ if } \ell \approx p}} X_{i_1} \cdots X_{i_{2k}}.$$

Taking the $(X_i, 1 \leq i \leq n)$ to be independent GUE matrices and letting the size going to infinity, we know that the expectation of the renormalized trace of polynomials in $(X_1, \dots, X_n)$ converge to the number of non crossing pairings of the letters that can be build above this polynomial so that only all pairing contain only the same letter. By symmetry, it is not hard to see that when summing over all these graphs, each loop will come δ times, hence yielding the trace τ . For more general δ 's, one has to sum over the vertices of a graph whose adjacency matrix has eigenvalue δ [35, 37].

This approach can be generalized as follows. Let $S_1, \dots, S_n$ be Temperley-Lieb elements. Let $\beta_1, \dots, \beta_n$ be small real numbers and for any Temperley-Lieb element S , define

$$\text{Tr}_\beta(S) = \sum_{n_i \geq 0} \sum_{1 \leq i \leq n} \prod \frac{\beta_i^{n_i}}{n_i!} \delta^{\#\text{loops}}$$

where we sum over all connected planar diagrams build over n_i diagrams S_i and one diagram S by matching the boundary points of these diagrams and we count the number of loops of the full picture. Then, we can prove

Theorem 4.6 (G-Jones-Shlyakhtenko 09').

Take $\delta \in I := \{2 \cos(\frac{\pi}{n})\}_{n \geq 4} \cup]2, \infty[$. Tr_β is a tracial state, as a limit of matrix (or free probability) models.

The construction is made by considering, instead of independent Gaussian random matrices, random matrices interacting via a potential chosen appropriately.

5. Conclusion

In this review, we tried to advertise free probability theory to the physicists community. Indeed, it is particularly convenient to describe the asymptotics of random matrices with genuinely independent random eigenvectors which are given by free operators. Hence, many of such limits have already been studied in free probability, cf. the so-called R -diagonal operators which describe the limit of non-normal matrices. Moreover, because free probability has developed many powerful tools from classical probability, it can give new ideas to study random matrices, cf. Monte-Carlo type of ideas to generate a Gibbs measure by Langevin dynamics, which in turn allows to study several matrix models in non-perturbative situations. The relation between free probability theory and operator algebra theory, equipped with the classical notion of freeness, is also very important to analyze the asymptotics of random matrices, cf. the connectivity of the support of the spectral measure of random matrices interacting via a convex potential. Vice versa, the relation between random matrices and combinatorics developed by the so-called topological expansion allows to get more insight in operator algebra, cf. the construction of the tower of factors. Of course, there are much more applications and developments around these themes and we refer the readers to review articles and books [14–17]. To conclude, we would like however to point out that the range of applications of free probability mainly concerns random matrices whose eigenvectors are approximately uniformly distributed on the sphere (which correspond to Haar distributed eigenvectors). This is well known to be the case for instance for Wigner matrices with independent entries with high enough moments, which have “delocalized” eigenvectors. When the entries have no second moment, the asymptotic distribution is different [38–40], the eigenvectors more localized [41] and it is not clear how to interpret the limit in the free probability context.

References

- [1] J. Feinberg and A. Zee, *Nuclear Phys. B* **501**, 643 (1997).
- [2] G. P. E. Brézin, C. Itzykson and J. B. Zuber, *Comm. Math. Phys.* **59**, 35 (1978).
- [3] H.-J. Sommers, A. Crisanti, H. Sompolinsky and Y. Stein, *Phys. Rev. Lett.* **60**, 1895 (1988).
- [4] Y. V. Fyodorov, B. A. Khoruzhenko and H.-J. Sommers, *Phys. Rev. Lett.* **79**, 557 (1997).
- [5] Y. V. Fyodorov, B. A. Khoruzhenko and H.-J. Sommers, *Phys. Lett. A* **226**, 46 (1997).
- [6] M. A. Halasz, A. D. Jackson and J. J. M. Verbaarschot, *Phys. Lett. B* **395**, 293 (1997).
- [7] M. A. Halasz, A. D. Jackson and J. J. M. Verbaarschot, *Phys. Rev. D* (3) **56**, 5140 (1997).
- [8] K. Efetov, *Supersymmetry in disorder and chaos* (Cambridge University Press, Cambridge, 1997).
- [9] R. A. Janik, M. A. Nowak, G. Papp and I. Zahed, *Nuclear Phys. B* **501**, 603 (1997).

- [10] S. Grossmann and M. Robnik, *J. Phys. A* **40**, 409 (2007).
- [11] U. Haagerup and F. Larsen, *J. Funct. Anal.* **176**, 331 (2000).
- [12] D. Voiculescu, *Invent. Math.* **104**, 201 (1991).
- [13] E. P. Wigner, *Annals Math.* **67**, 325 (1958).
- [14] A. Anderson, G. W. and Guionnet and O. Zeitouni, *An introduction to random matrices* (Cambridge University Press, Cambridge, 2009).
- [15] D. Voiculescu, *Bull. London Math. Soc.* **34**, 257 (2002).
- [16] D. V. Voiculescu, K. J. Dykema and A. Nica, *Free random variables*, CRM Monograph Series, Vol. 1 (American Mathematical Society, Providence, RI, 1992). A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.
- [17] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, London Mathematical Society Lecture Note Series, Vol. 335 (Cambridge University Press, Cambridge, 2006).
- [18] A. Guionnet, M. Krishnapur and O. Zeitouni, The single ring theorem, arXiv:0909.2214, (2009).
- [19] J. Ginibre, *J. Mathematical Phys.* **6**, 440 (1965).
- [20] M.L. Mehta, *Random Matrices*, Academic Press, NY (1967)
- [21] V. L. Girko, *Teor. Veroyatnost. i Primenen.* **29**, 669 (1984).
- [22] Z. D. Bai, *Ann. Probab.* **25**, 494 (1997).
- [23] F. Götze and A. Tikhomirov, arXiv:0709.3995v3 [math.PR] (2007).
- [24] T. Tao and V. Vu, *Commun. Contemp. Math.* **10**, 261 (2008).
- [25] V. G. Knizhnik, A. M. Polyakov and A. B. Zamolodchikov, *Modern Phys. Lett. A* **3**, 819 (1988).
- [26] J. Ambjørn, L. Chekhov, C. F. Kristjansen and Y. Makeenko, *Nuclear Phys. B* **404**, 127 (1993).
- [27] N. M. Ercolani and K. D. T.-R. McLaughlin, *Int. Math. Res. Not.* **14**, 755 (2003).
- [28] P. L. Albeverio S. and S. M., *Comm. Math. Phys.* **224**, 271 (2001), Dedicated to Joel L. Lebowitz.
- [29] A. Guionnet and E. Maurel Segala, *Alea* **1**, 241 (2006).
- [30] A. Guionnet and E. Maurel Segala, *Ann. Probab.* **35**, 2160 (2007).
- [31] E. Maurel Segala, High order asymptotics of matrix models and enumeration of maps, arXiv:math/0608192v1 [math.PR], (2006).
- [32] A. Guionnet and D. Shlyakhtenko, *Geom. Funct. Anal.* **18**, 1875 (2009).
- [33] M. Pimsner and D. Voiculescu, *J. Operator Theory* **8**, 131 (1982).
- [34] V. F. R. Jones, *Invent. Math.* **72**, 1 (1983).
- [35] A. Guionnet, V. Jones and D. Shlyakhtenko, Free probability, planar algebras and subfactors, arXiv:0712.2904, (2007).
- [36] V. Jones, D. Shlyakhtenko and K. Walker, An orthogonal approach to the subfactor of a planar algebra., arXiv:0807.4146, (2008).
- [37] V. F. R. Jones, The planar algebra of a bipartite graph, in *Knots in Hellas '98 (Delphi)*, , Ser. Knots Everything Vol. 24 (World Sci. Publ., River Edge, NJ, 2000) pp. 94–117.
- [38] J. Bouchaud and P. Cizeau, *Phys. Rev. E* **50**, 1810 (1994).
- [39] G. Ben Arous and A. Guionnet, *Comm. Math. Phys.* **278**, 715 (2008).
- [40] S. Belinschi, A. Dembo and A. Guionnet, *Comm. Math. Phys.* **289**, 1023 (2009).
- [41] A. Soshnikov, Poisson statistics for the largest eigenvalues in random matrix ensembles, in *Mathematical physics of quantum mechanics*, , Lecture Notes in Phys. Vol. 690 (Springer, Berlin, 2006) pp. 351–364.

THE PHYSICS OF DECISION MAKING: STOCHASTIC DIFFERENTIAL EQUATIONS AS MODELS FOR NEURAL DYNAMICS AND EVIDENCE ACCUMULATION IN CORTICAL CIRCUITS

PHILIP HOLMES*, PHILIP ECKHOFF, and K.F. WONG-LIN

*Program in Applied Mathematics, Neuroscience Institute
and Department of Mechanical and Aerospace Engineering,
Princeton University, Princeton NJ 08544, USA*

**E-mail: pholmes@math.princeton.edu
www.princeton.edu/mae/people/faculty/holmes/*

RAFAL BOGACZ

*Department of Computer Science, University of Bristol,
Bristol BS8 1UB, U.K.*

MIRIAM ZACKSENHOUSE

*Faculty of Mechanical Engineering, Technion –
Israel Institute of Technology, Haifa 32000, Israel.*

JONATHAN D. COHEN

*Department of Psychology and Neuroscience Institute,
Princeton University, Princeton NJ 08544, USA*

We describe how drift-diffusion (DD) processes – systems familiar in physics – can be used to model evidence accumulation and decision-making in two-alternative, forced choice tasks. We sketch the derivation of these stochastic differential equations from biophysically-detailed models of spiking neurons. DD processes are also continuum limits of the sequential probability ratio test and are therefore optimal in the sense that they deliver decisions of specified accuracy in the shortest possible time. This leaves open the critical balance of accuracy and speed. Using the DD model, we derive a speed-accuracy tradeoff that optimizes reward rate for a simple perceptual decision task, compare human performance with this benchmark, and discuss possible reasons for prevalent sub-optimality, focussing on the question of uncertain estimates of key parameters. We present an alternative theory of robust decisions that allows for uncertainty, and show that its predictions provide better fits to experimental data than a more prevalent account that emphasises a commitment to accuracy. The article illustrates how mathematical models can illuminate the neural basis of cognitive processes.

Keywords: Accumulator, decision making, dimension reduction, drift-diffusion process, speed-accuracy tradeoff, spiking neuron model

1. Introduction

In this article we review work done over the past ten years by many students and colleagues, in which we attempt to model decision-making and cognitive control as part of a program to better understand the neural substrates and mechanisms that constitute human and primate brains. We approach this from multiple directions and viewpoints, using both cellular level models of individual neurons and networks, informed by electrophysiological data from experimental neuroscience, and “high level” models of cortical areas implicated in specific cognitive tasks, motivated and parameterized by behavioral, imaging, and EEG experiments by psychologists. We are interested in the general question: “How do neural spikes generate behavior?”

Before such a vague question can be addressed, it must be made more precise. This is typically done by studying a particular sensory, motor, or cognitive function in a specific animal *in vivo*, or by probing cellular or network dynamics in a slice of cortical tissue *in vitro*. Mathematical models are playing increasingly important roles in planning, executing, and analyzing the behavioral and electrophysiological data recorded in such work. In this paper we focus on a simple perceptual decision task in which subjects are trained (or asked, if human) to correctly identify a stimulus, drawn at random from a set of two possibilities, in each of a sequence of trials. Under the free-response paradigm, subjects respond “in their own time”, but are often encouraged to do so quickly. Such two-alternative forced-choice (2AFC) tasks are common in the psychological literature, and random walk and drift-diffusion (DD) processes have long been used to model subjects’ reaction time distributions and error rates [1–5]. 2AFC tasks with visual motion stimuli are also frequently used in primate studies [6–11], and *in vivo* recordings show that neural spike rates in certain oculo-motor regions, including the lateral intraparietal area (LIP) and frontal eye fields (FEF), evolve like sample paths of a DD process, rising to a “decision threshold” prior to response initiation.

2. Decision-making models and an optimal speed-accuracy tradeoff

Our autonomous and sensori-motor nervous systems are remarkably complex. The human brain contains $\mathcal{O}(10^{11})$ neurons (nerve cells), interconnected by $\mathcal{O}(10^{14})$ synapses. Individual neurons exhibit a wide range of morphologies, with extensively branching dendritic and axonal processes that collect information from other neurons, in the form of action potentials (APs) or spikes, and relay it to yet others. Neurons are also functionally diverse: they can spike repetitively, autonomously emit quasi-regular bursts, or spike only upon the arrival of specific triggering signals. Their collective behaviors are driven by, and result in, events spanning molecular to macroscopic scales, that evolve in milliseconds to hours, weeks and years (e.g., in the case of learning).

This complexity has prompted descriptions of neural dynamics over a similar range of spatial and temporal scales, most of which employ differential equations. Three broad classes of models are in common use. The extremes are “high-level”

(low-dimensional) connectionist networks that describe averaged activity levels in specific brain areas, and “low-level” (high-dimensional) biophysically-based models of individual neurons. The former include DD processes that mimic evidence accumulation (§1), while the latter, based on the classic work of Hodgkin and Huxley [12], can represent multiple ion channels and retain cell geometries by using multiple compartments to describe post-synaptic inputs along dendrites and APs traveling down axons. An intermediate class of integrate-and-fire (spiking) models replaces the millisecond AP dynamics by stereotypic delta functions, retaining only the sub-spike-threshold voltage dynamics in each cell, but allowing multiple molecular species to interact in synaptic connections.

2.1. *Leaky competing accumulators and drift-diffusion processes*

Leaky competing accumulators [13] are neurally-inspired high-level networks that can be parameterized to match behavioral data such as reaction-time distributions and error rates. In the simplest such model, appropriate for 2AFC tasks, units with activity levels (x_1, x_2) represent two pools of neurons selectively responsive to the two stimuli (e.g., in area LIP). The units mutually inhibit one another via input-output or frequency current (f-I) functions that express neural activity (e.g., short-term firing rates) in terms of input currents. Additional inputs include constant currents representing mean stimulus levels, and i.i.d. Wiener processes modeling noise that pollute the stimuli and enter the local circuit from other brain regions. In the absence of inputs, the activities decay to a baseline level which can be set at zero without loss of generality. The stochastic differential equations (SDEs) may be written:

$$dx_1 = [-\gamma x_1 - \beta f(x_2) + \mu_1] dt + \sigma dW_1, \quad (1)$$

$$dx_2 = [-\gamma x_2 - \beta f(x_1) + \mu_2] dt + \sigma dW_2, \quad (2)$$

where the state variables x_j denote unit activities (spike rates), γ, β are the leak and inhibition rates and μ_j, σ are the means and standard deviation of the noisy stimuli. A decision is supposed to be made when the first of the unit activities $x_j(t)$ exceeds a fixed threshold $x_{j,th}$. See [13] and [14, 15] for background on such connectionist networks.

The function $f(\cdot)$ is typically sigmoidal:

$$f(x) = \frac{1}{1 + \exp[-g(x - b)]}, \quad (3)$$

and if the gain g and bias b are set appropriately, Eqns. (1-2) without noise ($\sigma = 0$) can have one or two stable equilibria, separated by a saddle point in the latter case. These correspond to two “choice attractors” in the noisy system, and if γ and β are sufficiently large, a one-dimensional, attracting, slow manifold $\mathcal{M}$ exists that contains both stable and unstable equilibria and the solutions connecting them [16–18]. Thus, once rapid transients have decayed following stimulus onset, the dynamics is essentially that of a nonlinear, scalar Ornstein-Uhlenbeck process [19, 20]. The

leading terms in this SDE are obtained by linearizing (1-2) and subtracting to form a single equation for the difference in activity $x = x_1 - x_2$:

$$dx = [(\mu_1 - \mu_2) + (\beta - \gamma)x] dt + \sigma dW. \quad (4)$$

In the perfectly balanced case $\beta = \gamma$, Eqn. (4) becomes a DD process [21].

As described at length in [21] and references therein, the DD process is a continuum limit of the sequential probability ratio test (SPRT) [22, 23]: an optimal decision maker in the sense that it delivers responses of guaranteed average accuracy in the shortest possible time. This leaves open the question of how one should determine a *speed-accuracy tradeoff* (SAT), since increased accuracy typically demands slower and more careful information processing, and hence results in longer reaction times. In the work described here, we focus on SATs that maximize the *reward rate*: the average reward gained in each trial, divided by the average trial duration (see Eqn. (25) below).

2.2. A spiking neuron model and its reduction to competing accumulators

In [24] the spiking network from [25] was adapted to model evidence accumulation in oculo-motor cortical areas and to study norepinehrine (NE) modulation on decision making. NE can change cellular excitability and synaptic efficacy, thus altering behavioral performance [26–28]. The network contains 2000 leaky integrate-and-fire neurons divided into four groups: two stimulus-selective populations each containing 240 excitatory pyramidal cells, a non-selective pool of 1120 pyramidal cells, and an inhibitory population of 400 interneurons. The state variables are the cellular trans-membrane voltages $V_j(t)$ (mV), the internal synaptic variables $S_{AMPA,j}(t)$, $S_{NMDA,j}(t)$, and $S_{GABA,j}(t)$, and noisy external inputs $S_{AMPA,ext,j}(t)$ to all cells. See Fig. 1. The computational studies of [24] investigated the effects of tonic and phasic NE release when glutamatergic (AMPA and NMDA) and GABA-ergic synapse strengths are both simultaneously and separately modulated, showing that co-modulation can provide good performance over a substantial range, while allowing tuning of speed and accuracy of responses.

The ODEs describing the system are

$$C_j \frac{dV_j}{dt} = -g_L(V_j - V_L) + I_{syn,j}(t), \quad (5)$$

$$\frac{dS_{type,j}}{dt} = -\frac{S_{type,j}}{T_{type}} + \sum_l \delta(t - t_j^l), \text{ where} \quad (6)$$

$$I_{syn,j}(t) = \sum_{type,k} -g_{type,k}(V_k - V_E)S_{type,k} \quad (7)$$

and $type = AMPA, GABA, \text{ or } AMPA, ext$, T_{type} is the time constant for that synapse type and k ranges over all presynaptic neurons. Two ODEs are needed to

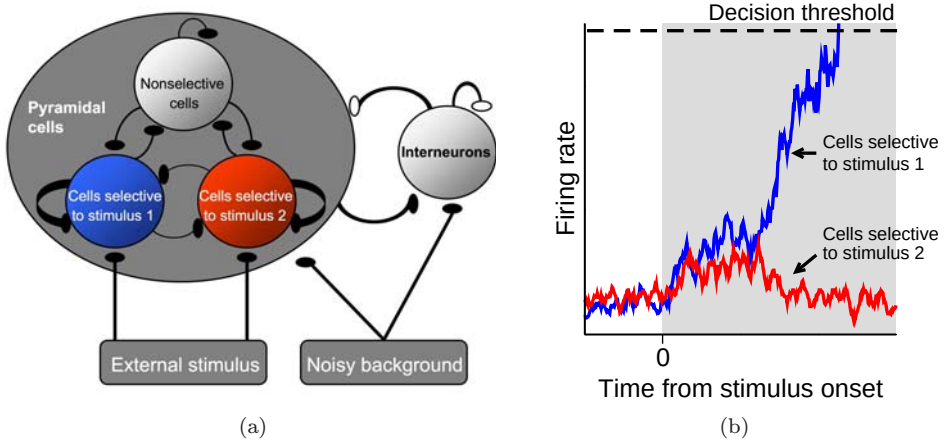


Fig. 1. (a) The integrate-and-fire network contains three populations of excitatory cells; each selective population responds preferentially to one stimulus, the third is nonselective to both stimuli. A pool of interneurons provides overall inhibition. Excitatory (NMDA- and AMPA-mediated) and inhibitory (GABA_A-mediated) synapses are denoted by filled and open ovals respectively. All cells receive noisy background excitation (AMPA-mediated); each cell connects to every other and selective populations have relatively stronger local recurrent excitation. (b) Stimuli excite both selective populations, but inhibition typically suppresses one population, producing winner-take-all dynamics. A decision is made when the first population crosses a fixed decision threshold. Figure adapted from [24].

describe the slower *NMDA* dynamics:

$$\frac{dS_{NMDA,j}(t)}{dt} = -\frac{S_{NMDA,j}(t)}{\tau_{NMDA,decay}} + \alpha x_j(t)[1 - S_{NMDA,j}(t)], \quad (8)$$

$$\frac{dx_j(t)}{dt} = -\frac{x_j(t)}{\tau_{NMDA,rise}} + \sum_l \delta(t - t_j^l). \quad (9)$$

When $V_j(t)$ crosses a threshold V_{thresh} at time t_j^l the cell emits a delta function $\delta(t - t_j^l)$, after which V_j is instantaneously reset and held at V_{reset} for an absolute refractory period τ_{ref} . The sums over delta functions in (6) and (9) represents spikes in presynaptic neurons, and the external currents are generated by Gaussian noise of mean μ and standard deviation σ :

$$dS_{AMPA,ext,j}(t) = -(S_{AMPA,ext,j} - \mu_j) \frac{dt}{\tau_{AMPA}} + \sigma_j \sqrt{\frac{dt}{\tau_{AMPA}}} \mathcal{N}(0, 1). \quad (10)$$

The stimulus is represented by the addition of terms $\mu(1 \pm E)$ to the mean inputs μ_1 and μ_2 to the selective cells, with appropriate adjustments to the variances σ_j . Here μ is the overall strength of the stimulus and E denotes its discriminability ($E = 1$ for perfectly clear stimuli with infinite signal-to-noise ratio; $E = 0$ for zero signal-to-noise ratio). Neuromodulation of synapses is introduced by scaling

the excitatory and inhibitory conductances in Eqn. (7) ($g_{AMPA,k}$, $g_{NMDA,k}$ and $g_{GABA,k}$ respectively) by multiplying them by factors γ_E and γ_I .

Using the all-to-all coupling structure and eliminating irrelevant $S_{type,j}$'s (excitatory neurons have no *GABA*-synapses, inhibitory neurons have no *AMPA*- or *NMDA*-synapses), Eqns. (5-10) still constitute a 9200-dimensional stochastic dynamical system that is analytically intractable and computationally intensive to simulate. We now sketch a sequence of low-dimensional reductions that preserve key physiological detail, permit bifurcation analyses, and relate the spiking network to the leaky accumulators described in §2.1. Full details, along with parameter values appropriate for modeling area LIP, appear in [29, 30].

We first reduce to a four-population model using a mean field approach from statistical physics [25], simplifying the self-consistency calculations of [31, 32] by employing a fixed average voltage $\bar{V} = (V_{reset} + V_{thresh})/2$ to estimate synaptic currents that enter each of the four cell populations as $J_{type,k} = -g_{type,k}(\bar{V} - V_{type})/1000$.^a These are then multiplied by the appropriate number N_j of presynaptic cells in each population and by an averaged synaptic variable $S_{type,j}$, and summed to create the incoming synaptic input currents to each population. Each term in the current takes the form

$$I_{type,k}(t) = N_j J_{type,k} \omega_{j,k} S_{type,j}(t), \quad (11)$$

where $\omega_{j,k}$ denotes the strength of the synaptic connection from population j to population k .

Individual neuron voltages are replaced by averaged firing rates determined by frequency-current (f-I) relationships, analogous to the input-output function $f(\cdot)$ of Eqn. (3). This yields an 11-dimensional system described by 4 firing rates $\nu_j(t)$, one inhibitory population-averaged synaptic variable $S_{GABA}(t)$, and two such variables $S_{AMPA,j}(t)$ and $S_{NMDA,j}(t)$ for each excitatory population (6 in all). The seven synaptic equations take the forms

$$\frac{dS_{NMDA,j}}{dt} = -\frac{S_{NMDA,j}}{T_{NMDA}} + 0.641(1 - S_{NMDA,j})\frac{\nu_j}{1000}, \quad (12)$$

$$\frac{dS_{AMPA,j}}{dt} = -\frac{S_{AMPA,j}}{T_{AMPA}} + \frac{\nu_j}{1000}, \quad (13)$$

$$\frac{dS_{GABA,I}}{dt} = -\frac{S_{GABA,I}}{T_{GABA}} + \frac{\nu_I}{1000}, \quad (14)$$

where $j = 1, 2, 3$, for the excitatory populations, and the four firing rates obey

$$\frac{d\nu_j}{dt} = \frac{-(\nu_j - \phi_j(I_{syn,j}))}{T_{AMPA}}, \quad (15)$$

with $j = 1, 2, 3$ and $j = I$ for the inhibitory population. Various f-I relationships $\phi_j(I_{syn,j})$ are considered in [29, 30] and ultimately piecewise-linear forms are found

^aHere and below, division by 10^3 is necessary to accomodate the conventional units of millivolts, nanoamps, and nanoseimens for conductances.

to be adequate, e.g., for the non-selective population:

$$\phi_3(I_{syn,3}) = 1 + c\theta(I_{syn,3} - I_{thresh,3}), \quad (16)$$

in which θ is a Heaviside step function that “turns on” at $I_{thresh,3}$ and the background firing rate in the absence of inputs is 1 Hz.

The final reduction to two populations is based on separation of time scales [16, 33]. The synaptic time constants for *AMPA* and *GABA* are fast ($T_{AMPA} = 2$ ms and $T_{GABA} = 5$ ms), while that for *NMDA* decay is relatively slow ($T_{NMDA} = 100$ ms). $S_{AMPA,j}(t)$ and $S_{GABA,j}(t)$ therefore converge rapidly to quasi-steady states that closely track the population firing rates in Eqns. (13-14), so that $S_{type,j} \approx \nu_j T_{type}/1000$. This eliminates three ODEs for the excitatory populations and one for the inhibitory population, producing a 7-dimensional system.

Firing rates also rapidly approach the values set by the f-I curves in Eqns. (15), since they too are dominated by T_{AMPA} , so we may set $\nu_j(t) \approx \phi_j(I_{syn,j}(t))$ for the nonselective and interneuron populations $j = 3$ and I and thereby remove the ODEs for ν_3 and ν_I . Now only three ODEs describing *NMDA* dynamics in the excitatory populations remain, along with the ODEs for ν_1 and ν_2 . Finally, when stimuli are on, the nonselective population $j = 3$ typically has a lower and less variable firing rate than the two selective populations, so that $S_{NMDA,3}$ can be replaced by its steady state value and its evolution equation may also be removed, leaving four ODEs:

$$\dot{S}_1 = -\frac{S_1}{T_{NMDA}} + 0.641(1 - S_1)\frac{\nu_1}{1000}, \quad (17)$$

$$\dot{S}_2 = -\frac{S_2}{T_{NMDA}} + 0.641(1 - S_2)\frac{\nu_2}{1000}, \quad (18)$$

$$\dot{\nu}_1 = -\frac{\nu_1 - \phi_{mod}(I_{syn,1})}{T_{pop2}}, \quad (19)$$

$$\dot{\nu}_2 = -\frac{\nu_2 - \phi_{mod}(I_{syn,2})}{T_{pop2}}, \quad (20)$$

where we write $S_{NMDA,j} = S_j$ for brevity.

To complete the reduction process, the input currents, including contributions from nonselective and inhibitory populations that no longer enter Eqns. (17-20), must be estimated in a self-consistent manner. This is complicated by the fact that $I_{syn,j}$ in (19-20) contains terms that depend on both S_j and $\phi_{mod}(I_{syn,j})$, so that the vectorfield is defined recursively. Ideally, we seek linear relationships of the form:

$$I_{syn,1} = \alpha_1 S_1 + \alpha_2 S_2 + \beta_1 \nu_1 + \beta_2 \nu_2 + I_{const,1}, \quad (21)$$

$$I_{syn,2} = \alpha_2 S_1 + \alpha_1 S_2 + \beta_2 \nu_1 + \beta_1 \nu_2 + I_{const,2}, \quad (22)$$

as in [25]. Here the use of piecewise-linear f-I curves is crucial, and the regions in parameter space in which the currents lie above ($I_{syn,3} > I_{thresh,3}$) and below threshold must be carefully delineated to determine if the terms ν_j in Eqns. (21-22) are active or not. The gain parameters γ_E, γ_I enter this system via the *AMPA*,

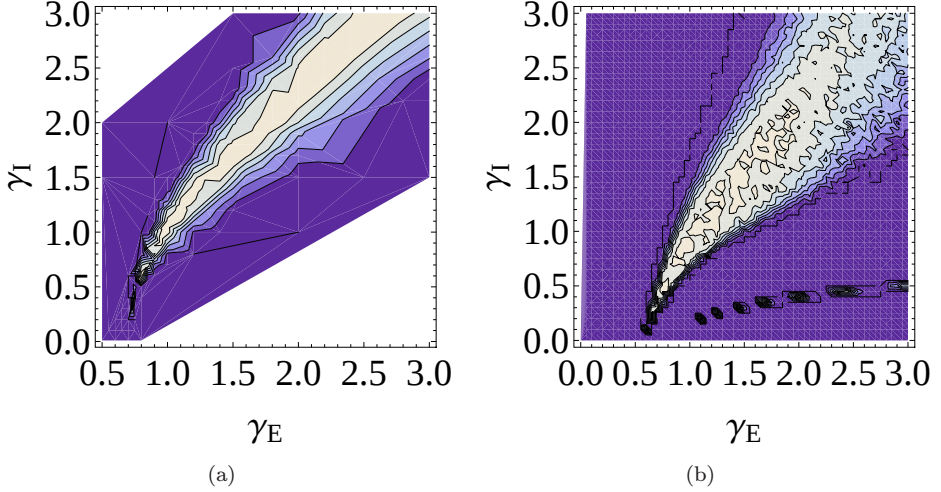


Fig. 2. Contour maps of reward rates over the glutamatergic (γ_E) and *GABA*-ergic (γ_I) gain modulation plane for the spiking network model (a), and for the 2-dimensional reduced system (b). At upper left inhibition is dominant and excitatory cells rarely exceed decision threshold, at lower right excitation dominates and excitatory cells rapidly cross threshold, leading to impulsive choices with high error rates. A ridge of high *RRs* (lighter shades) separates these low *RR* regions.

NMDA and *GABA* components $J_{type,k}$ of the currents $I_{const,j}$ and also the coefficients α_j and β_j in Eqns. (21-22).

In [30] we show that the dynamics of solutions of Eqns. (17-20) are dominated by their evolution on the (S_1, S_2) plane, and that this is largely explained by the structure of the nullclines [16] on which $\dot{S}_j = 0$ and $\nu_j = \phi_{mod}(I_{syn,j})$. The reduced state variables $S_j(t)$ therefore represent the neural activities of the two selective populations, providing explicit examples of the states $x_j(t)$ in the leaky competing accumulators of Eqns. (1-2).

That the resulting reduced four-variable ODE system captures key properties of the 9,200-dimensional spiking neuron simulation is verified by computing reward rates over the (γ_E, γ_I) -neuromodulation plane for both models, as illustrated in Fig. 2. Bifurcation diagrams, computed using AUTO [34], showing branches of fixed points of the noise-free system at the two locations $(\gamma_E, \gamma_I) = (1, 1)$ and $(2, 2)$, are shown in Fig. 3. These reveal that multiple stable states (attractors) can coexist both for $\mu = 0$ (stimulus off) and $\mu > 0$ (stimulus on). The state with S_j (and ν_j) both low represents lack of response (possibly awaiting stimulus appearance), while S_1 high and S_2 low and its partner $S_2 \gg S_1$ represent the two possible choices, and S_1, S_2 both high represents impulsive behavior in which both populations are so strongly activated that near-random choices occur. This latter behavior is prevalent when excitation exceeds inhibition in the lower right of the neuromodulation plane of Fig. 2.

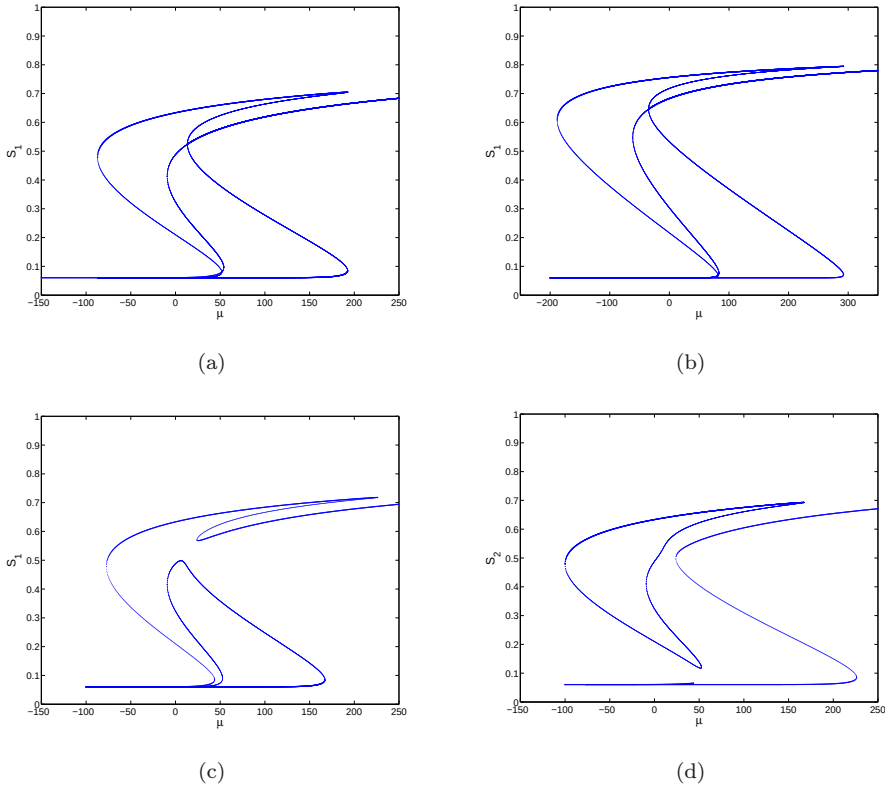


Fig. 3. Bifurcation diagrams for the noise-free two-population model at $E = 0$ and $(\gamma_E, \gamma_I) = (1, 1)$ (a) and $(2, 2)$ (b), showing S_1 vs. stimulus strength μ . For these zero-coherence cases the system is reflection-symmetric about the hyperplane $S_1 = S_2$, $\nu_1 = \nu_2$ and pitchfork bifurcations occur from the low-low and high-high branches. Panels (c) and (d) show S_1 and S_2 vs. μ for a non-symmetric, finite coherence case $E = 0.128$ at $(\gamma_E, \gamma_I) = (1, 1)$. Stable branches are shown bold.

2.3. An optimal speed-accuracy tradeoff

We now return to the DD process. With unbiased initial data, appropriate to experiments in which the stimuli appear with equal probability, this SDE may be written

$$dx = A dt + \sigma dW; \quad x(0) = 0, \quad (23)$$

where $A = \mu_1 - \mu_2$ denotes the drift rate (cf. Eqn. (4)). Responses are given when the solution first crosses one of the thresholds $x = \pm x_{th}$, and if $A > 0$ crossing of $+x_{th}$ is correct (response 1) and $-x_{th}$ is incorrect (response 2). For this simple SDE, the error rate and mean decision time, quantifying accuracy and speed are given by:

$$p(\text{err}) = \frac{1}{1 + \exp(2\eta\theta)} \quad \text{and} \quad \langle DT \rangle = \theta \left[\frac{\exp(2\eta\theta) - 1}{\exp(2\eta\theta) + 1} \right]; \quad (24)$$

see [19, 35] and [21, Appendix]. The three parameters A, σ and x_{th} reduce to two new parameters $\eta \equiv (A/\sigma)^2$ (the signal-to-noise ratio (SNR), having units of inverse time), and $\theta \equiv |x_{th}/A|$ (the threshold-to-drift ratio, being the first passage time for the noise-free process $x(t) = At$).

We follow [36] in proposing that, to optimize performance, subjects should maximize their *reward rates*: the average accuracy divided by average time between responses:

$$RR = \frac{1 - p(\text{err})}{\langle DT \rangle + T_0 + D_{RSI}}, \quad (25)$$

where T_0 is that part of the reaction time devoted to sensory and motor processing and D_{RSI} is the experimenter-imposed response-to-stimulus interval (RSI).^b This is appropriate for experiments in which D_{RSI} and the stimulus discriminability are held constant in each block of trials, and block durations are fixed. Since in this case T_0 and η also typically remain (approximately) constant for each subject, we may substitute Eqns. (24) into (25) and maximize RR for fixed η, T_0 and D_{RSI} , obtaining a unique value of θ for each pair (η, D_{tot}) , given by the expression

$$\exp(2\eta\theta) - 1 = 2\eta(D_{tot} - \theta), \text{ where } D_{tot} = T_0 + D_{RSI}. \quad (26)$$

Inverting the relationships (24) to obtain

$$\theta = \frac{\langle DT \rangle}{1 - 2p(\text{err})} \quad \text{and} \quad \eta = \frac{1 - 2p(\text{err})}{2\langle DT \rangle} \log \left[\frac{1 - p(\text{err})}{p(\text{err})} \right], \quad (27)$$

we can replace the model parameters θ, η in Eqn. (26) by the performance measures, $p(\text{err})$ and $\langle DT \rangle$ to yield a unique, parameter-free relationship that describes the speed-accuracy $(\langle DT \rangle, 1 - p(\text{err}))$ tradeoff that maximizes RR :

$$\frac{\langle DT \rangle}{D_{tot}} = \left[\frac{1}{p(\text{err}) \log \left[\frac{1 - p(\text{err})}{p(\text{err})} \right]} + \frac{1}{1 - 2p(\text{err})} \right]^{-1}. \quad (28)$$

This *optimal performance curve* (OPC), pictured in Fig. 4(a), was first presented in [37] and [21]. Each condition, specified by SNR η and total delay D_{tot} , determines an optimal threshold-to-drift ratio θ_{op} by Eq. (26). As task conditions vary, by manipulating η via stimulus discriminability or D_{tot} via RSI, θ_{op} changes and with it the SAT that maximizes RR . As shown in Fig. 4(a), RR s increase moving leftward along the OPC, but as noted above, each subject's SNR stabilizes after sufficient practice. Along with the RSI, the SNR uniquely determines location along the OPC and the maximum RR available under the given conditions. Optimality is therefore interpreted as follows: given η and D_{tot} , the subject's RR is maximized by selecting $\theta = \theta_{op}$; any other threshold, higher or lower, yields lower net rewards (see diamonds in Fig. 4(a)). In this sense, every performance point $(p(\text{err}), \langle DT \rangle / D_{tot})$ on the OPC is optimal.

^bAdditional penalty delays imposed after errors can also be included, as in [21].

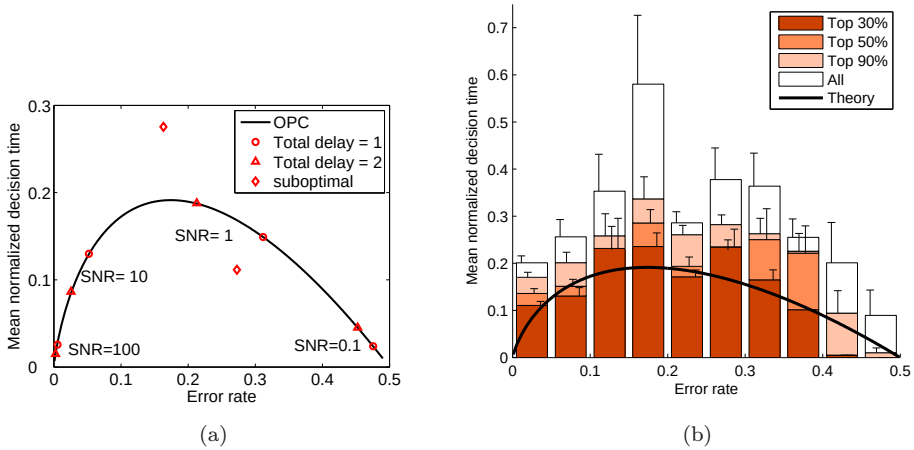


Fig. 4. (a) OPC of Eq. (28) relates mean normalized decision time $\langle DT \rangle / D_{tot}$ to error-rate $p(\text{err})$, triangles and circles mark performances under 4 different task conditions. Moving leftwards, the resulting RRs increase with SNR from 0.51 to 0.60, 0.84 and 0.97 for $D_{tot} = 1$, and from 0.26 to 0.33, 0.45 and 0.49 for $D_{tot} = 2$. Diamonds mark suboptimal performance points resulting from thresholds 25% above and below θ_{op} for SNR=1 and $D_{tot} = 2$; both reduce RR by $\approx 1.3\%$. (b) OPC (thick curve) and data collected from 80 human subjects (histograms) sorted according to total rewards accrued over multiple blocks of trials and conditions. White bars: all subjects; lightest bars: lowest 10% excluded; medium bars: lowest 50% excluded; darkest bars: lowest 70% excluded. Vertical line segments indicate standard errors. From [37, Fig. 1].

The OPC's shape may be intuitively understood by observing that very noisy stimuli ($\eta \approx 0$) contain little information, so that given a priori knowledge that they are equally likely, it is optimal to choose at random without examining them, giving $p(\text{err}) = 0.5$ and $\langle DT \rangle = 0$ (SNR = 0.1 in Fig. 4(a)). As $\eta \rightarrow \infty$ stimuli become so easy to discriminate that both $\langle DT \rangle$ and $p(\text{err})$ approach zero (SNR = 100 in Fig. 4(a)). For intermediate SNRs it is advantageous to accumulate evidence for long enough to make the best possible choice (SNR = 1 and 10 in Fig. 4(a)).

3. Behavioral experiments: a prevalent lack of optimality

Two 2AFC experiments were performed to test the ability of humans to optimize their performance [21, 38]. In the first, 20 subjects viewed motion stimuli [39] and were rewarded for each correct discrimination. The experiment was divided into 7-minute blocks with different response-to-stimulus intervals in each block. In the second, 60 subjects discriminated if the majority of 100 locations on a static visual display were filled with stars or empty [4]. In this case blocks lasted for 4 minutes, and the entire set was repeated under two difficulty conditions. Variation of D_{RSI} and difficulty ($\sim 1/\eta$) produced a wide range of error rates. In both cases subjects were instructed to maximize their total earnings, and unrewarded practice blocks were administered before the tests began. For details, see [21, 38].

Since the OPC is independent of the parameters defining the DD process, and

D_{tot} enters only as the denominator in Eqn. (28), data can be pooled for all subjects and task conditions, with different D_{RSI} and η , and compared with the OPC. $p(\text{err})$'s were computed and $\langle DT \rangle$'s estimated by fitting the DD model to reaction time distributions [21, 38] and the 0 – 50% error rate range was divided into 10 bins and $\langle DT/D_{tot} \rangle$ computed for each bin by averaging over those results and conditions with error rates in that bin. This yields the open (tallest) bars in Fig. 4(b); the shaded bars derive from similar analyses restricted to subgroups of subjects ranked by their total rewards accrued over all blocks of trials and conditions. Clearly, the top 30% group performs near optimally, with performances close to the OPC, but the groups with lower total scores exhibit significantly longer decision times and are typically more accurate [38].

In a related study of perceptive decisions in the presence of prior information on relative reward magnitudes [40], it was found that two monkeys came within 0.5% and 2% respectively of their maximum possible expected rewards by (almost) optimizing the balance between tilting toward the higher reward, and accurately assessing the noisy stimulus. These animals made decisions under a cued-response paradigm that relieves the need for a speed-accuracy tradeoff, but their ability to appropriately modulate noisy “bottom-up” stimuli (moving dot displays with randomly mixed coherences) with “top-down” expectations of rewards is nonetheless impressive.

In the next two subsections we discuss alternative theories from which SATs can be derived, assuming that the DD process of Eqn. (23) does describe evidence accumulation, and in §3.3 we describe some biophysical factors that may prevent optimal evidence accumulation in decisions under the free-response paradigm.

3.1. *A preference for accuracy?*

Taking longer to assess incoming data typically produces more accurate responses ($\langle DT/D_{tot} \rangle$ rises with θ in Eqn. (24), while $p(\text{err})$ declines). The apparent emphasis on accuracy over speed is common [41], and it motivated us to consider two further objective functions. The first postulates a balance of rewards and accuracy:

$$RA = RR - \frac{q p(\text{err})}{D_{tot}}, \quad (29)$$

in which q specifies the relative weight placed on accuracy. Alternatively, for monetary rewards one can suppose that errors are penalized by subtraction from previous winnings:

$$RR_m = \frac{(1 - p(\text{err})) - q p(\text{err})}{\langle DT \rangle + D_{tot}}. \quad (30)$$

These can both be seen as specific instantiations of the competition between accuracy and reward (COBRA) theory proposed in [42, 43].

The functions defining RA and RR_m each involve a free parameter q , and the resulting one-parameter families of OPCs [21] clearly can provide better fits to

experimental data than optimization of reward rate alone (the curves lie above the OPC of Fig 4 for $q > 0$). In the next subsection we compare them with the data of Fig. 4(b), and with predictions of an entirely different theory that also predicts longer-than-optimal decision times.

3.2. Robust decisions in the face of uncertainty?

In the analyses of §2.3 and §3.1 it is assumed that subjects maximize a relevant utility function, given the actual task parameters. However, these parameters are rarely known with accuracy and decisions must be made under uncertainties: e.g., the RR depends on inter-trial delays that may be difficult to estimate. Here we sketch an approach that presumes that the parameters are deterministic but unknown [44, 45]. We focus on bounded uncertainties, in which parameters are assumed to lie within a specified uncertainty set, and apply the maximin strategy to select the decision that maximizes the minimum utility. This guarantees good, albeit suboptimal, performance throughout the uncertainty set; in this sense the behavior is robust.^c We consider uncertainties in delays and SNR, showing that the basic pattern of maximin performance curves for the former agrees well with the experimental data, while that for the latter does not. For details, see [46].

Interval timing studies [47] indicate that estimated intervals are normally distributed around the true duration with a standard deviation proportional to it. This “scalar invariance property” [48] suggests that the size of the uncertainty set for the actual interval increases with its duration. We therefore suppose that subjects work on the assumption that the actual delay lies in a presumed uncertainty set

$$U_p(\alpha_p; \tilde{D}_{tot}) = \left\{ D_{tot} > 0 : \left| D_{tot} - \tilde{D}_{tot} \right| \leq \alpha_p \tilde{D}_{tot} \right\}, \quad (31)$$

whose size is proportional to the nominal delay $\tilde{D}_{tot}$, with a *presumed level of uncertainty* α_p . In place of the optimal threshold given by Eqn. (26), the maximin strategy selects the threshold θ_{MM} that maximizes the worst RR that can occur for $D_{tot} \in U_p(\alpha_p; \tilde{D}_{tot})$. It is proved in [46] that this predicts a one-parameter family of *maximin performance curves* (MMPCs) for uncertainties in total delay that are scaled versions of the OPC (28), specifically:

$$\frac{\langle DT \rangle}{D_{tot}} = \gamma(1 + \alpha_p) \left[\frac{1}{p(\text{err}) \log \left(\frac{1-p(\text{err})}{p(\text{err})} \right)} + \frac{1}{1 - 2p(\text{err})} \right]^{-1}, \quad (32)$$

where $\gamma \equiv \tilde{D}_{tot}/D_{tot}$ is the ratio of the nominal to the actual delay. Recalling the results of Fig. 4(b) and anticipating those of Fig. 5 below, we note that, like the OPCs for the utility functions of §3.1, these curves also predict longer mean decision times than the OPC (28).

^cRobustness here is analogous to a similar notion reported in [24]: maintaining near-optimal performance over a range of neuromodulator levels, cf. §2.2 and Fig. 2.

A similar theory can be developed for SNRs η within a presumed uncertainty set $U_p(\alpha_p, \tilde{\eta}) = \{\eta > 0 : |\eta - \tilde{\eta}| \leq \alpha_p \tilde{\eta}\}$, although the cases of uncertainties in drift rate A and in noise variance σ must be considered separately. This leads to two families of MMPCs that differ in functional forms both from the OPC (28) and MMPC (32), rising to peaks in mean decision time at progressively smaller error rates as the uncertainty α_p increases (large SNRs are more susceptible to poor estimates, and therefore have greater effects at low error rates).

Unlike the maximin strategy, *robust-satisficing* assumes that subjects try to meet a required performance level despite the uncertainties [45]. They therefore attempt to maximize robustness: i.e., the (unknown, and possibly unbounded) level of uncertainty α_p for which a necessarily sub-optimal reward rate RR_{req} can be guaranteed. For both delays and SNRs this leads to one-parameter families of *robust-satisficing performance curves* (RSPCs). However, the RSPCs for uncertainties in SNR predict mean normalized decision times that deviate considerably from the behavioral data of Fig. 4(b), matching it even less well than the analogous MMPCs.

Fig. 5 shows fits of several of the performance curves derived above to subsets of the behavioral data. Specifically, we compare the parameter-free OPC with fits of the two objective functions introduced in §3.1 that penalize errors, and with MMPCs for uncertainties in delays and SNR as well as RSPCs for uncertainties in delays (we exclude RSPCs for uncertainties in SNR for the reason noted above). The figure shows that, while there is little difference in the fits of all the alternative theories to data from the top 30% group (the free parameter in each function merely provides a slightly better fit over the OPC (28)), data from the middle 60% and lowest 10% subgroups exhibit patterns that allow us to distinguish among fit qualities. Overall, we find that MMPCs for uncertainties in delays provide the best fits, with RSPCs for uncertainties in delays second best, and the alternative objective function RA third. Table 1 provides quantitative details of fit qualities in terms of maximum likelihood, showing that the data is over 13 (respectively, 43) times more likely given the MMPC with uncertainty in D_{tot} , than the RSPC with uncertainty in D_{tot} (respectively, the OPC for RA). Further details and discussions appear in [46].

We are currently conducting experiments to test the conjecture, motivated by this analysis, that “poor timers” are more likely to overestimate optimal decision times by (implicitly) applying an MMPC or RSPC policy. In addition to running 2AFC experiments similar to those of [21, 38] (with additional training sessions to better stabilize behavior), we are directly monitoring timing ability [47], and also running a deadlined version of the 2AFC task, which a Bayesian optimality analysis [50] predicts should result in *underestimation* of decision times, due to a desire not to incur penalties for late responses.

3.3. Biophysical constraints?

We end this section by briefly noting a further potential reason for suboptimal performance. The DDM, and the SPRT from which it derives, is formulated under

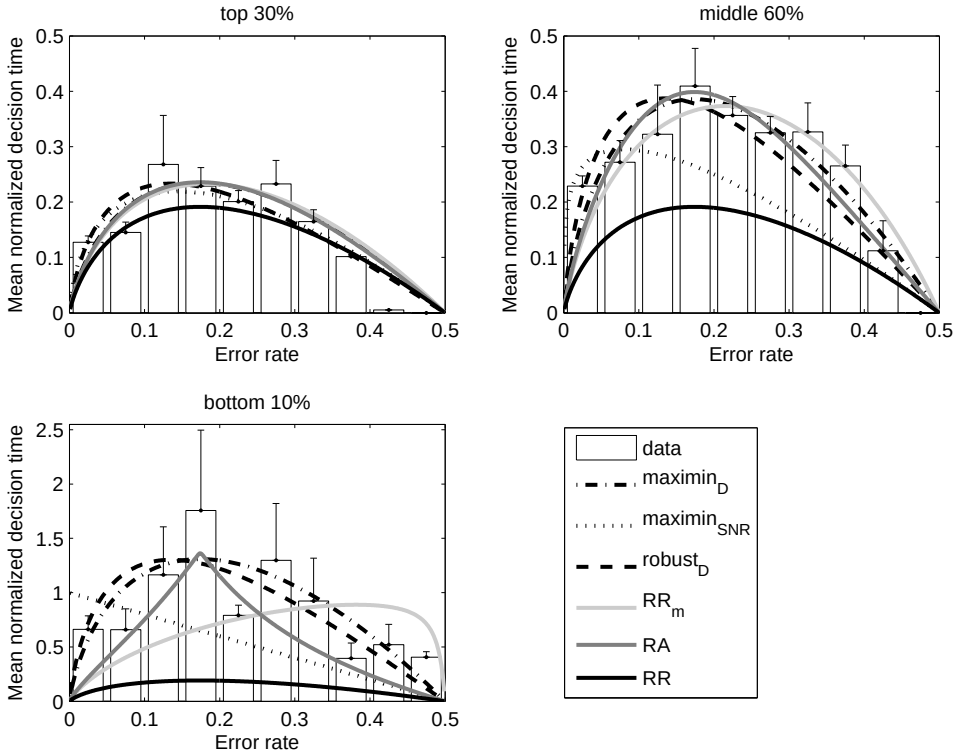


Fig. 5. Comparisons of performance curves with mean normalized decision times from experimental blocks for three groups of subjects sorted by total rewards acquired. Error bars show standard errors of mean normalized decision times. Different curves are identified by line style and gray scale in the key, in which maximin_D and $\text{maximin}_{\text{SNR}}$ refer to MMPCs for uncertainty in total delay (eq. (32)) and noise variance respectively, robust_D refers to RSPCs for uncertainty in total delay, RA and RR_m refer to the objective functions of Eqns. (29) and (30) respectively, and RR refers to the OPC of Eqn. (28). Note the different vertical axis scales in upper and lower panels.

the assumption that evidence can be amassed at a constant rate for as long as required to obtain a decision of *any* specified accuracy [22, 23]. To obtain arbitrarily low error rates (which are necessary for optimal performance as RSIs increase), one must select arbitrarily high thresholds (cf. Eq. (24)), implying that differences between neural activities cannot be bounded. This clearly contradicts the fact that neural activity, measured in terms of short-term firing rates, is bounded both above and below. Additionally, as the derivation of the reduced population models in §2.2 makes clear, even in simplified cases with piecewise-linear f-I functions such as Eqs. (17-22), the final dynamical system is nonlinear. While it can be linearized near specific equilibria, the most relevant region includes a saddle point that lies on a boundary separating the domains of attraction of the low-high and high-low choice states (analogous to the branches of saddle points that bend back towards

Table 1. Likelihood ratios of the data given the MMPC with uncertainty in D compared to other performance curves (rows), for different groups of subjects (columns). The last column is the product of the first three and thus expresses likelihood ratios for all data. The higher the likelihood ratio, the less likely is it that the data could be generated by the corresponding model, in comparison to the MMPC with uncertainty in D . To account for the fact that all curves have one free parameter except the OPC, the bottom row includes in brackets the ratio of the likelihood of the data given the two curves divided by $\exp(1)$, as proposed by [49].

Performance curve	Top 30%	Middle 60%	Bottom 10%	Product
MMPC for SNR	1.13	$> 10^8$	308.38	$> 10^{10}$
RSPC for D	1.99	4.04	1.72	13.82
OPC for RR_m	1.28	38.79	208.47	$> 10^4$
OPC for RA	1.11	8.11	4.83	43.48
OPC for RR	47.34 [17.42]	$> 10^{21}$	$> 10^6$	$> 10^{28}$

lower μ values for the two-population model in Fig. 3).

The reduced one-dimensional system, restricted to the unstable manifold of this equilibrium, behaves locally like an unstable Ornstein-Uhlenbeck process [19, 21, 51]. Suboptimality results from the tendency of solutions, perturbed by noise into the basin of the incorrect attractor early in the decision process, to irretrievably accelerate away from the boundary and cross the incorrect threshold before becoming trapped by the incorrect choice attractor. Such a nonlinear model has been fitted to behavioral data [20], as well as being derived from a biophysical spiking network near a subcritical pitchfork bifurcation point [16], although the fits to specific behavioral data in [20] show little improvement over simpler linear models.

4. Discussion and conclusions

Theoretical neuroscience, or more loosely, “brain science”, has prompted a wealth of mathematical modeling, from computationally intensive simulations of networks of ion-channel, multi-compartment cells, to abstracted connectionist models of interacting brain areas that may be reduced to one- and two-dimensional stochastic dynamical systems [52]. The works surveyed in this article attempt to bridge the gap between computational studies of physiologically-realistic spiking neuron models of neuromodulation [24] and the leaky accumulator, drift-diffusion, and Ornstein-Uhlenbeck models used in cognitive psychology (§2.1). Such low-dimensional models, with appropriate parameterization, can describe behavioral aspects of two-alternative decisions remarkably well [3–5, 7, 13, 36]. They have even recently been extended to model decision making by individuals in groups receiving feedback from other group members [53].

These models are not only fast and simple to simulate; they also admit analytical study and improve understanding of basic mechanisms. For example, we doubt that theoretical developments such as the prediction of optimal speed-accuracy trade-offs described here, could as easily or as often emerge from massive computational

simulations. However, low-dimensional descriptions lack the cellular and synaptic details needed to relate neuromodulator and other effects to behavioral outcomes, and without these it is difficult to introduce parameters representing cognitive control into such models.

In §2.2 we provide a bridge across the dimensional divide by sketching the derivation of low-dimensional population models from a 9,200-dimensional spiking (integrate-and-fire) network. The reduction process preserves the parametric gain modulation due to neurotransmitter levels. Bifurcation diagrams computed for a two-population model, in particular, are helpful in interpreting the reward rate maps of Fig. 2 in terms of the dynamics and firing rates of populations of cells selective for the two stimuli. Moreover, for the computational cost of simulating 500 trials at a single gain condition in the spiking model, it is possible to simulate 500 four-population or 2000 two-population trials at each point on a grid of 900 gain conditions. Additionally, the low-dimensional models can be improved as neurophysiological data becomes available, for example by further adjusting their input-output response curves, or by going beyond the first order balances of mean field theory used here (see, e.g., [54, 55]).

Encouraged by these results, in §2.3 we revisit the simplest decision making model – the drift-diffusion (DD) process – and review the optimal performance curve relating mean decision times and error rates that derives from it. The behavioral experiments that this theory prompted reveal that a majority of human subjects fails to optimize, preferring to make slower, more accurate decisions. In §3 we review this data, proposing several explanations for suboptimality and showing that accounting for uncertainty in timing ability appears to provide the best explanation for this particular data set (§3.2). Assuming uncertainties in estimating delays, both the maximin approach, in which subjects maximize the lowest reward rate that can occur within a presumed uncertainty window, and the robust satisfying approach, in which subjects maximize the uncertainty window under which a required reward rate can be guaranteed, lead to performance curves that match the data well; although the later provides marginally superior fits. Approaches that penalize errors, as often assumed in cognitive psychology [41–43], provide inferior fits.

The question of how this speed-accuracy tradeoff is modulated remains largely open. It is generally agreed that regions in the prefrontal cortex exert cognitive control by modulating the activity of other brain areas in which information is processed and retrieved [56, 57]. In the present context of perceptual decisions based on correct identification of visual motion stimuli, we are especially interested in the interaction, in the oculo-motor pathway, of “bottom-up” stimulus information and “top-down” biases that reflect a subject’s experience and expectations. These appear to meet in the lateral intraparietal area [40, 58]: the region modeled in §2.2. We believe that the development and use of low-dimensional population models derived from biophysically-detailed spiking networks will illuminate the neural mechanisms that make possible the merging of prior and present experiences that is so central

to the human condition.

Acknowledgments

This work was supported by PHS grant MH62196 (Cognitive and Neural Mechanisms of Conflict and Control, Silvio M. Conte Center) and AFOSR grants FA9550-07-1-0537 and FA9550-07-1-0528. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government. The author gratefully acknowledges many contributions of present and former members of the Conte Center modeling group, not all of whom could be cited in this article.

References

- [1] M. Stone, *Psychometrika* **25**, 251 (1960).
- [2] D. Laming, *Information Theory of Choice-Reaction Times* (Academic Press, New York, 1968).
- [3] R. Ratcliff, *Psychol. Rev.* **85**, 59 (1978).
- [4] R. Ratcliff, T. Van Zandt and G. McKoon, *Psychol. Rev.* **106** (2), 261 (1999).
- [5] P. Smith and R. Ratcliff, *Trends in Neurosci.* **27** (3), 161 (2004).
- [6] J. Schall, *Nature Reviews in Neuroscience* **2**, 33 (2001).
- [7] J. Gold and M. Shadlen, *Trends in Cognitive Science* **5** (1), 10 (2001).
- [8] J. Roitman and M. Shadlen, *J. Neurosci.* **22** (21), 9475 (2002).
- [9] R. Ratcliff, A. Cherian and M. Segraves, *J. Neurophysiol.* **90**, 1392 (2003).
- [10] M. Mazurek, J. Roitman, J. Ditterich and M. Shadlen, *Cerebral Cortex* **13**, 891 (2003).
- [11] R. Ratcliff, Y. Hasegawa, R. Hasegawa, P. Smith and M. Segraves, *J. Neurophysiol.* **97**, 1756 (2006).
- [12] A. Hodgkin and A. Huxley, *J. Physiol.* **117**, 500 (1952).
- [13] M. Usher and J. McClelland, *Psychol. Rev.* **108**, 550 (2001).
- [14] D. Rumelhart and J. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (MIT Press, Cambridge, MA., 1986).
- [15] S. Grossberg, *Neural Networks* **1**, 17 (1988).
- [16] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer-Verlag, New York, 1983).
- [17] E. Brown and P. Holmes, *Stochastics and Dynamics* **1**, 159 (2001).
- [18] E. Brown, J. Gao, P. Holmes, R. Bogacz, M. Gilzenrat and J. Cohen, *Int. J. Bifurcation and Chaos* **15**(3), 803 (2005).
- [19] C. Gardiner, *Handbook of Stochastic Methods, Second Edition* (Springer, New York, 1985).
- [20] A. Roxin and A. Ledberg, *PLoS Comput. Biol.* **4**, p. e1000046 (2008).
- [21] R. Bogacz, E. Brown, J. Moehlis, P. Holmes and J. Cohen, *Psychol. Rev.* **113**(4), 700 (2006).
- [22] A. Wald, *Sequential Analysis* (Wiley, New York, 1947).
- [23] A. Wald and J. Wolfowitz, *Ann. Math. Statist.* **19**, 326 (1948).
- [24] P. Eckhoff, K.-F. Wong and P. Holmes, *J. Neurosci.* **29**, 4301 (2009).
- [25] K. F. Wong and X.-J. Wang, *J. Neurosci.* **26**, 1314 (2006).

- [26] C. W. Berridge and B. D. Waterhouse, *Brain Res. Rev.* **42**, 33 (2003).
- [27] G. Aston-Jones and J. D. Cohen, *Annu. Rev. Neurosci.* **28**, 403 (2005).
- [28] S. J. Sara, *Nat. Rev. Neuro.* **10**, 211 (2009).
- [29] P. Eckhoff, Neuromodulation of decision making: Incorporating single-cell physiology in low-dimensional models, PhD thesis, Princeton University, 2009.
- [30] P. Eckhoff, K.-F. Wong and P. Holmes, Dimension reduction and dynamics of a spiking neuron model for decision making under neuromodulation, In review (2009).
- [31] N. Brunel and X.-J. Wang, *J. Comput. Neurosci.* **11**, 63 (2001).
- [32] A. Renart, N. Brunel and X.-J. Wang, Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks, in *Computational Neuroscience: A Comprehensive Approach*, ed. J. Feng (Chapman and Hall/CRC, Boca Raton, 2004) pp. 431–490.
- [33] C. Jones, *Geometric Singular Perturbation Theory*, Lecture Notes in Mathematics, Vol. 1609 (Springer Verlag, Heidelberg, 1994). C.I.M.E. Lectures.
- [34] E. Doedel, A. Champneys, T. Fairgrieve, Y. Kuznetsov, B. Sandstede and X. Wang, *AUTO 97: Continuation and bifurcation software for ordinary differential equations*, tech. rep., Concordia University (1997), Available via FTP from directory /pub/doedel/auto at [ftp.cs.concordia.ca](ftp://ftp.cs.concordia.ca); software available from <http://indy.cs.concordia.ca/auto/> and <http://sourceforge.net/projects/auto2000>.
- [35] J. Busemeyer and J. Townsend, *Psychological Review* **100**, 432 (1993).
- [36] J. Gold and M. Shadlen, *Neuron* **36**, 299 (2002).
- [37] P. Holmes, E. Shea-Brown, J. Moehlis, R. Bogacz, J. Gao, G. Aston-Jones, E. Clayton, J. Rajkowski and J. Cohen, *IEICE Transactions on Fundamentals on Electronics, Communications and Computer Science* **E88-A(10)**, 2496 (2005).
- [38] R. Bogacz, P. Hu, J. Cohen and P. Holmes, *Quarterly J. Exp. Psychol.* In press, published on line Sept 10th, 2009.
- [39] K. Britten, M. Shadlen, W. Newsome and J. Movshon, *Visual Neurosci.* **10**, 1157 (1993).
- [40] S. Feng, P. Holmes, A. Rorie and W. Newsome, *PLoS Comput. Biol.* **5**, p. e1000284 (2009).
- [41] I. Myung and J. Busemeyer, *Amer. J. Psychol.* **102(1)**, 1 (1989).
- [42] W. Maddox and C. Bohil, *J. Exp. Psychol.* **24(6)**, 1459 (1998).
- [43] C. Bohil and W. Maddox, *Memory & Cognition* **31(2)**, 181 (2003).
- [44] A. Ben-Tal and A. Nemirovski, *Mathematical Programming* **112**, 125 (2008).
- [45] Y. Ben-Haim, *Information Gap Decision Theory: Decisions under Severe Uncertainty* (Academic Press, New York, 2006). 2nd Edition.
- [46] M. Zacksenhouse, R. Bogacz and P. Holmes, Robust versus optimal strategies for two-alternative forced choice tasks, In review, (2009).
- [47] C. Buhusi and W. Meck, *Nature* **6**, 755 (2005).
- [48] J. Gibbon, *Psychol. Rev.* **84(3)**, 279 (1977).
- [49] H. Akaike, *J. Econometrics* **16**, 3 (1981).
- [50] P. Frazier and A. Yu, Sequential hypothesis testing under stochastic deadlines, in *Advances in Neural Information Processing Systems 20*, eds. J. Platt, D. Koller, Y. Singer and S. Roweis (MIT Press, Cambridge, MA, 2008) pp. 465–472.
- [51] W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1957). 2 volumes. Second Edition.
- [52] Y. Liu, P. Holmes and J. Cohen, *Neural Computation* **20 (2)**, 345 (2008).
- [53] A. Nedic, D. Tomlin, P. Holmes, D. Prentice and J. Cohen, A simple decision task in a social context: experiments, a model, and analyses of behavioral data, in *Proc. 47th*

- IEEE Conf. on Decision and Control*, 2008. (Session TuB15: Mixed Robot/Human Team Decision Dynamics).
- [54] D. Cai, L. Tao and D. McLaughlin, *Proc. Nat Acad. Sci.* **101**, 14288 (2004).
 - [55] D. Cai, L. Tao, A. Rangan and D. McLaughlin, *Commun. Math. Sci.* **4**, 97 (2006).
 - [56] E. Miller and J. Cohen, *Annu. Rev. Neurosci.* **24**, 167 (2001).
 - [57] M. Botvinick, T. Braver, D. Barch, C. Carter and J. Cohen, *Psychol. Rev.* **108**, 624 (2001).
 - [58] G. Corrado, L. Sugrue, S. Seung and W. Newsome, *J. Exp. Anal. Behav.* **84**, 581 (2005).

NEW TECHNOLOGIES IN THE HUNT FOR NEW PHYSICS

DAVID A. KOSOWER

*Institute de Physique Théorique CEA-Saclay
F-91191 Gif-sur-Yvette cedex
France*

The Large Hadron Collider at CERN will probe a new regime of physics, seeking answers to long-standing puzzles underlying the Standard Model of particle physics. Precision computations of known physics, and in particular of scattering processes in quantum chromodynamics, will play an important role in understanding backgrounds to new physics. The required calculations are complex and difficult. Recent years have seen the emergence and maturation of new “on-shell” technologies, based on unitarity and factorization with a pinch of twistor string theory, for performing calculations of loop amplitudes in gauge theories.

Keywords: On-Shell Methods; Perturbative QCD

1. Introduction

The start-up of the Large Hadron Collider (LHC), now underway at CERN, opens a new path to physics discoveries at the small-distance frontier of particle physics. The experimental program will peer beyond the known physics of the $SU(3) \times SU(2) \times U(1)$ Standard Model. Finding the remaining undiscovered particle of the Standard Model, the Higgs boson — or whatever new physics substitutes for it — will be the first challenge for the primary particle-physics detectors, ATLAS and CMS. The even harder challenges of finding physics completely beyond the Standard Model await in turn. The experimental program also challenges theorists to provide tools for assisting data analysis, and in particular for providing precision predictions for the known physics of the Standard Model that will show up as backgrounds to new-physics searches.

The general-purpose detectors at the LHC consist of several subsystems: precision vertex detectors; charged-particle tracking; electromagnetic calorimetry; hadron calorimetry; and muon chambers. Charged leptons — electrons and muons — provide the cleanest signals. If there is new physics that can be detected solely using charged leptons, it will be relatively easy to find.

The two other final-state objects that show up in the detector are narrowly-collimated streams of hadrons, known as ‘jets’; and missing energy, corresponding to the presence of one or more neutrinos (or other, as yet unknown, neutral weakly-interacting particles). Jets are the manifestation of colored-parton scattering and

production. At very short times after the collision, shorter than the time for light to cross a proton, the scattering can be thought of perturbatively, that is in terms of quarks and gluons. The basic process is that of quark or gluon scattering, or annihilation to produce a ‘final’ state such as a quark recoiling against an electroweak vector boson. As is true in electrodynamics, radiation is emitted when charged particles are accelerated as is inevitable in a scattering process. In quantum chromodynamics (QCD), additional gluons are produced during the short-distance scattering. The difference is that the production rate is much larger, because the strong coupling is much larger. In addition, the gluons themselves are charged, and so a cascade of gluons will be produced. At timescales of order the proton light-crossing time, these partons become confined inside hadrons and hadronic resonances, and it is these hadrons (or their weak decay products) which are actually seen inside the detector in the form of jets.

The two general-purpose collider detectors at CERN are able to collect nearly all hadrons; only those traveling down the beam pipe or very nearly so can escape detection. This is necessary in order to measure accurately the net momentum transverse to the beam pipe. Measuring the energy of each visible particle, weighted by its direction and then projected onto the plane transverse to the beam, gives a quantity known as the transverse energy E_T . An imbalance in transverse energy — so-called missing transverse energy — signals the presence of weakly-interacting neutral particles, which accordingly leave no track nor energy deposit in the detector. Missing transverse energy will play a crucial role in many new-physics searches at the LHC. Calibrating and developing a detailed understanding of detector response will take time to accomplish; and even once accomplished, discoveries and measurements relying on jets will be more difficult. This difficulty is due to the larger backgrounds, which in turn reflect the larger partonic radiation of colored particles accompanying a short-distance QCD process. Nonetheless, the use of channels involving jets is essential because of their much-larger rates. The difficulty of finding new physics signals using final states involving only jets means that the primary discovery channels will be those involving both jets and at least one lepton, or jets and missing transverse energy.

Part of the difficulty arises from Standard-Model processes which produce the same final states. The signals can only be extracted by a combination of experimental cuts and subtraction of computed backgrounds. This in turn requires a detailed quantitative understanding of these backgrounds, as any new physics signal the experiments are seeking will necessarily have total rates below (often far below) Standard-Model signals, for example production of electroweak vector bosons in association with multiple jets.

One might not expect processes involving the strong interactions to be calculable at all in a perturbative expansion. Here, the asymptotic freedom discovered by Gross, Politzer, and Wilczek (which earned them the 2004 Nobel Prize in physics) comes to our rescue. While at long distances the strong interactions indeed cannot be treated perturbatively, at short distances, they can be, so that a differential cross

section will have the following form,

$$\sigma_n = \hat{\sigma}^{\text{LO}} \alpha_s^n(\mu) + \hat{\sigma}^{\text{NLO}}(\mu) \alpha_s^{n+1}(\mu) + \hat{\sigma}^{\text{NNLO}}(\mu) \alpha_s^{n+2}(\mu) + \dots \quad (1)$$

The argument of the strong coupling α_s is the renormalization scale, the scale at which the coupling is to be evaluated. The long-distance conversion of partons to hadrons will produce only small corrections to this expansion, so long as we consider sufficiently-inclusive observables. Jets defined with an infrared-safe algorithm fall into this class of observables.

Ordinarily in perturbative quantum field theory, or for that matter in any framework where a perturbative approach is valid, one expects the leading term in the computation of an observable quantity to provide a reasonable quantitative approximation to the physical value. Higher-order terms would then give increasingly precise predictions (up to the point at which the asymptotic nature of an expansion blocks further increases in precision). This is, for example, the situation in quantum electrodynamics for the anomalous magnetic moment of the electron.

The situation in quantum chromodynamics is different. In order to define the coupling, we must introduce a renormalization scale μ . The scale is unphysical — no physical observable can depend on it — but a fixed-order prediction in perturbation theory will depend on it. The dependence will be of order the first omitted term in the expansion. Qualitative arguments tell us that a sensible choice, minimizing the missing terms, is to take the renormalization scale to be of the same order as the characteristic energy scale in the process under study. This might be, for example, the jet transverse energy E_T in studying inclusive-jet production. This leaves an ambiguity in the value of the scale, and hence the value of the coupling one should use. Equivalently, the theoretical prediction is actually given by a band with fuzzy edges, rather than a sharp prediction. The bands are typically quoted varying the renormalization scale up and down by a factor of two from the ‘characteristic’ scale.

If the band were nonetheless very narrow, much narrower than experimental statistical and systematic errors, this dependence would be mostly a theoretical curiosity. However, in QCD the dependence of the coupling on the renormalization scale is significant. Accordingly, the leading-order (LO) prediction has substantial scale dependence even for processes involving the production of only one or two jets. The dependence becomes even stronger for processes in which more jets are produced. Furthermore, the ‘natural’ scale is not always obvious in processes with several jets, in which several different physical scales come into play.

Accordingly, one needs to go at least to the next order to obtain quantitatively reliable predictions. Next-to-leading order (NLO) calculations include virtual corrections to the basic partonic process. The loop amplitudes which furnish these virtual corrections also depend on the renormalization scale. As indicated in eq. (1), this dependence is different from that of the coupling α_s . Indeed, this additional dependence compensates partially for that of the coupling, and pushes the remaining scale dependence to one higher order in α_s . In practice, one finds that NLO predictions typically reduce the scale dependence to 10–20% from a factor of 2 or more in

multi-jet calculations. To obtain reliable error estimates on theoretical predictions, or to obtain predictions with scale dependence reduced even further (to 1%), one would need to go to yet higher order in the perturbative expansion.

The virtual corrections that enter into NLO calculations arise from the interference of one-loop amplitudes with tree-level amplitudes. Both are traditionally computed with Feynman-diagram techniques, which have proven both versatile and powerful in over a half-century of intensive use. At one loop, the diagrams must be combined with reduction of the tensor integrals that emerge [1], and reduction of higher-point loop integrals themselves [2]. For the amplitudes of interest to us, these techniques produce huge intermediate expressions. In the case of six-jet production, for example, requiring eight-point amplitudes, there are several million diagrams, leading to of order twenty billion terms even before integral reductions.

Yet the answers often turn out to be very simple, expressible in no more than hundreds or thousands of terms — and for some helicity amplitudes, in a single line. The vast size of the intermediate expressions is a reflection of the gauge non-invariant, redundant parts of the calculation. These cancel in the final answer, but with conventional diagrammatic techniques are present in intermediate stages. The factorial growth of the number of diagrams with increasing number of external legs also plays a role in the size of intermediate expressions.

If, following a phrase attributed to Feynman, we believe that *simple results should have a simple derivation*, we should seek calculational methods that remove the gauge non-invariance at all stages of a calculation, and make it as simple as possible at each stage. On-shell methods accomplish this goal.

2. On-Shell Methods

The differential cross sections we require for experimental predictions are built out of amplitudes, which are functions of the external momenta k_i , external particle wavefunctions (polarization vectors ϵ_i for gluons, and spinor wavefunctions for fermions), and color indices a_i . For simplicity, I shall focus on gluon external states. At tree level, it is useful to decompose the amplitude in a basis of color traces,

$$\mathcal{A}_n(k_1, \epsilon_1, a_1; k_2, \epsilon_2, a_2; \dots; k_n, \epsilon_n, a_n) = \sum_{\sigma \in S_n/Z_n} \text{Tr}(T^{a_{\sigma(1)}} T^{a_{\sigma(2)}} \dots T^{a_{\sigma(n)}}) A_n(k_1, \epsilon_1; k_2, \epsilon_2; \dots; k_n, \epsilon_n) \quad (2)$$

and thereby separate out the color dependence. (The sum is taken over non-cyclic permutations of the labels $1, \dots, n$.) A similar decomposition holds for one-loop amplitudes. The objects we need to compute are the color-ordered amplitudes A_n , whose arguments are cyclicly-consecutive external legs. We make use of the spinor-helicity formalism [3] (building on earlier forms for polarization vectors [4]) to give numerically-computable forms for the polarization vectors and compact forms for the amplitudes. All Lorentz invariants can also be expressed in terms of spinor products.

The basic idea behind on-shell methods is to use only information from physical, that is on-shell and transverse, intermediate states. Both tree and loop amplitudes do of course contain off-shell internal states in a diagrammatic expansion; but we extract all information about the off-shell behavior from quantities in which these states are placed on shell. The basic approach turns general properties of amplitudes into explicit calculational techniques. In particular,

- The factorization of amplitudes, when sums of several external momenta go on shell, gives rise to on-shell recursion relations both at tree level and at one loop;
- The unitarity of the S -matrix gives rise to the unitarity method of evaluating loop amplitudes via their cuts, both at one loop and at higher loops;
- The existence of an underlying field theory provides a representation of loop amplitudes in terms of Feynman integrals, which makes the unitarity approach practical.

The methods as applied to loop amplitudes can be summarized in the following equation,

$$\text{Amplitude} = \sum_{j \in \text{Basis}} c_j \text{Integral}_j + \text{Rational} . \quad (3)$$

The sum runs over the integrals in a basis. The coefficients c_j are rational functions of the spinor products formed out of spinors corresponding to the external momenta, as are the additional rational terms. The coefficients are computed using the four-dimensional unitarity method. The additional rational terms may be computed either using loop-level on-shell recursion relations, or using D -dimensional unitarity. The calculation yields loop amplitudes evaluated using a dimensional regulator with $D = 4 - 2\epsilon$, the standard regulator for modern QCD calculations. The calculation of the coefficients and of the rational part can be performed analytically, or as in more recent work, numerically. A numerical approach makes it possible to develop general-purpose codes capable of computing amplitudes for a wide range of processes, comparable to the codes [5, 6] used for tree-level matrix elements.

2.1. Integrals

The basis of integrals must be sufficient to express all of the original integrals present in a Feynman-diagram calculation. Each of these integrals must be expressible as a linear combination of basis integrals, with coefficients that are rational functions of spinor products and covariant tensors built out of the external momenta and the metric tensor $\eta^{\mu\nu}$. For analytic calculations, the basis can be overcomplete, so that this representation is not unique. For direct numerical calculations in the unitarity method, it must be a minimal basis, so that the representation is unique. The integral reduction techniques mentioned earlier [1, 2] and their extensions to D dimensions [7] show that at one loop, the basis for computations where all virtual particles are massless consists of box, triangle, and bubble integrals. The external legs correspond either to the original external momenta of the amplitude we are

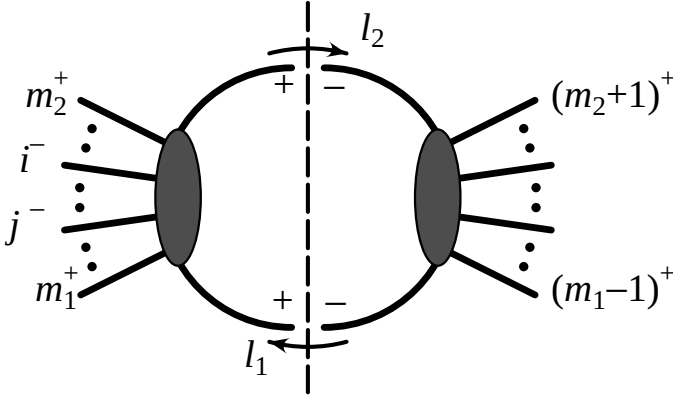


Fig. 1. The basic unitarity cut of a one-loop amplitude, showing how it is built out of two tree amplitudes.

calculating, or to sums of some of these momenta. Accordingly, while the internal lines of the integrals are massless, the external lines may be massive. This basis is in fact minimal, so it is suitable for numerical calculations. This gives us a total of ten integrals: six boxes, three triangles, and one bubble. (One of the box integrals, with all external legs massless, appears only in computations of the four-point amplitude.) These integrals have known analytic forms, computed using a dimensional regulator for both infrared and ultraviolet singularities.

2.2. Unitarity

The conservation of probability implies that the scattering matrix in a quantum theory is unitary. This implies that the discontinuity of the nontrivial transition matrix is given by its square,

$$\text{Disc } T = T^\dagger T. \quad (4)$$

As was already understood in the 1960s, the equation can be used to obtain the absorptive part of a one-loop diagram; up to the ambiguity of subtractions, the dispersive part could then be obtained by performing a dispersion integral. What was lacking was an understanding of how to systematically compute processes with more channels than a $2 \rightarrow 2$ process; of how to handle massless particles; and of how to combine the use of unitarity with field theory. Making use of the existence of an underlying field theory is what introduces an integral basis, avoiding the need for performing dispersion integrals. This makes the approach practical, and is essential to performing the cut computations in a purely numerical way.

The modern unitarity method [8], whose development started in the mid-1990s, overcame these deficiencies. Recent developments, relying in part on ideas emerging from twistor string theory [9], have made it possible to develop general-purpose efficient programs for computing one-loop amplitudes.

If we perform a simple unitarity cut on a one-loop amplitude, we obtain a phase space integral of the product of two on-shell tree-level amplitudes, as shown in fig. 1. Each amplitude is thus gauge invariant, and can be computed using efficient techniques for on-shell tree-level amplitudes. In a given channel, we can reverse this procedure, sewing together two tree-level amplitudes. We do this by first rewriting the phase-space integral representing the cut by a loop integral with delta functions forcing the cut legs on shell. We then undo the cut, replacing these delta functions with propagators for the cut legs. The two tree amplitudes within the integrand are still evaluated on shell; the difference between their off-shell and on-shell values leads only to contributions that have no cut in the given channel. The key to ensuring gauge invariance (and thus simple expressions) at each stage of the computation is the use of complete tree amplitudes, rather than just individual tree diagrams.

Having sewn together two tree amplitudes, we must perform algebra to identify the coefficients of the basis integrals, retaining only those integrals which have a cut in the given channel. In order to obtain the complete answer for all cut-containing terms in the amplitude, we must scan over all channels corresponding to cyclicly-ordered Lorentz invariants. In some cases, integrals may have cuts in more than one channel. For these channels, we may take the coefficients from the cut in either channel (or use the computation in alternate channels as a cross-check on the first). In this sense, we must “merge” results from different channels rather than simply summing over channels, as indicated by the quotes around the sum,

$$A^{1\text{-loop}} = \text{“} \sum_{\text{cuts } K^2} \text{”} \int \frac{d^{4-2\epsilon} \ell}{(2\pi)^{4-2\epsilon}} \frac{i}{\ell^2} A_L^{\text{tree}} \frac{i}{(\ell - K)^2} A_R^{\text{tree}}. \quad (5)$$

The merging runs over all invariants built out of cyclicly-consecutive sets of momenta, $K^2 = s_{12}, s_{23}, \dots, s_{123}, \dots$; ϵ denotes the dimensional regulator.

Cutting two propagators in an amplitude corresponds to one of the terms in the above computation. This selects terms in the master equation (3) which have those two propagators present in the integral function. There is no need to limit ourselves to two propagators, however. We can isolate a smaller set of contributions by requiring three specified propagators to be present, for example [10]. This would isolate a unique triangle integral, and a smaller set of box integrals. It would exclude bubble integrals entirely. We can achieve this by cutting three lines instead of two. Such a cut is known as a generalized unitarity cut.

At one loop, we can go as far as cutting four lines simultaneously, as pointed out by Britto, Cachazo, and Feng [11]. Such a quadruple cut freezes the loop momentum entirely, allowing us to solve for all cut loop momenta, and requiring no further algebra to isolate the coefficient of a single box integral. Indeed, the cut — and hence the coefficient of any given box integral — is given simply by a product of four tree amplitudes, summed over the two solutions to the cut equations,

$$\text{coeff} = \frac{1}{2} \sum_{\text{solutions}} A_1 A_2 A_3 A_4, \quad (6)$$

where the cut equations are,

$$\ell^2 = 0, \quad (\ell - K_1)^2 = 0, \quad (\ell - K_1 - K_2)^2 = 0, \quad (\ell + K_4)^2 = 0. \quad (7)$$

All tensor reductions are performed implicitly by the cutting and summation, eliminating another source of increase in the size of intermediate expressions when using traditional diagrammatic techniques. In general, the solutions to the quadruple-cut equations are complex. The use of complex momenta, again inspired by twistor string theory, is essential to treating massless external legs on a uniform footing with massive ones.

In simpler integrals, cutting all internal lines does not fix the loop momentum completely. For example, cutting all three internal lines in a triangle integral leaves one remaining degree of freedom for the loop integral. In order to isolate the triangle coefficient from those of boxes sharing the same triple cut, we can proceed in two complementary ways. The first, due to Forde [12], uses a carefully-chosen parametrization of the loop momenta, and observes that the coefficient is a contour integral extracting a residue at infinity,

$$\text{coeff} = \frac{1}{2\pi i} \oint \frac{dt}{t} [A_1 A_2 A_3](t), \quad (8)$$

of the product of three tree amplitudes isolated by the cuts. An alternate approach, due to Ossola, Papadopoulos, and Pittau (OPP) [13] again makes use of a parametrization of the cut loop momentum, but uses it to subtract the contributions of the box integrals from the expression, leaving behind the triangle coefficient. The computation of the coefficient of bubble integrals is somewhat more complicated but can be performed in a similar manner in either of these approaches.

These generalized-unitarity techniques, using complex momenta, along with reduction-subtraction techniques [13], can and have been used to develop general-purpose programs for computing one-loop amplitudes, where all the algebra and contour integration are performed numerically; only the integrals are input as analytic expressions.

At higher loops, we do not yet know the form of a complete basis; nor do we have known analytic expressions for all loop integrals. Nonetheless, generalized unitarity techniques can and have been applied successfully in a variety of computations [14–16] both in gauge and gravity theories. The idea is to apply multiple cuts to the amplitude in order to break it apart into a product of tree amplitudes. One chooses sets of cuts so as to isolate sets of possible integrals. Upon reassembling the cuts, one performs algebra if needed to separate the contributions sharing the same cuts. To obtain explicit numerical or analytic expressions for the amplitude, one must also evaluate the integral functions. This is a much more challenging task than at one loop, but both reduction technologies, primarily based on integration-by-parts identities [17, 18] and a variety of evaluation technologies, based on differential equations [19], Mellin–Barnes representations [20], and sector decomposition combined with contour deformation [21] have been developed over the last decade.

2.3. Factorization and On-Shell Recursion Relations

When a sum of external momenta goes on shell, an amplitude will factorize on the pole, for example,

$$A_n(\dots, a, b, c, \dots) \rightarrow \sum_{\lambda=\pm} A_4(a, b, c, K^\lambda) \frac{i}{(k_a + k_b + k_c)^2} A_{n-2}(-K^{-\lambda}, \dots), \quad (9)$$

with $K = -(k_a + k_b + k_c)$, when $(k_a + k_b + k_c)^2 \rightarrow 0$. The sum over λ is over the helicities of the intermediate leg.

We can use this factorization to derive an equation for the amplitude, following Britto, Cachazo, Feng, and Witten [22]. To do so, we introduce a shift of the spinors associated with two external legs, labeled by j and l ,

$$\begin{aligned} |j^-\rangle &\rightarrow |j^-\rangle - z|l^-\rangle, \\ |l^+\rangle &\rightarrow |l^+\rangle + z|j^+\rangle, \end{aligned} \quad (10)$$

where z is a complex parameter. This shifts the two external momenta by a complex vector,

$$\begin{aligned} \hat{k}_j^\mu &\rightarrow k_j^\mu(z) = k_j^\mu - \frac{z}{2} \langle j^- | \gamma^\mu | l^- \rangle, \\ \hat{k}_l^\mu &\rightarrow k_l^\mu(z) = k_l^\mu + \frac{z}{2} \langle j^- | \gamma^\mu | l^- \rangle, \end{aligned} \quad (11)$$

and in turn defines a z -dependent continuation $A_n(z)$ of the amplitude A_n . The shifted momenta are complex, but still satisfy $k^2 = 0$. Also, their sum is unchanged, so the resulting amplitude is still an on-shell amplitude for a momentum-conserving configuration. We will assume that $A(z) \rightarrow 0$ as $z \rightarrow \infty$; for appropriate helicities of legs j and l , this can be proven using a diagrammatic argument.

Now consider the contour integral,

$$\frac{1}{2\pi i} \oint_C \frac{dz}{z} A(z), \quad (12)$$

where C is a circle at infinity. By assumption on $A(z)$, this integral vanishes. We can evaluate it using residues; there is one residue from the pole at $z = 0$, and others from the poles of $A(z)$. We obtain the relation,

$$A(0) = - \sum_{\text{poles } \alpha} \text{Res}_{z=z_\alpha} \frac{A(z)}{z}, \quad (13)$$

for the original amplitude.

Where are the other poles of $A(z)$? The singularities of the tree-level amplitude arise precisely from internal momenta going on shell, that is from zeros of z -shifted propagator denominators. In order to have z dependence at all in a propagator denominator, the sum of momenta flowing through it must contain only one of the two shifted momenta $\hat{k}_j$ and $\hat{k}_l$. We should thus split the external momenta into two sets, each containing one of these two shifted momenta. We will need to sum over all ways of partitioning the external momenta into two such sets, while of course

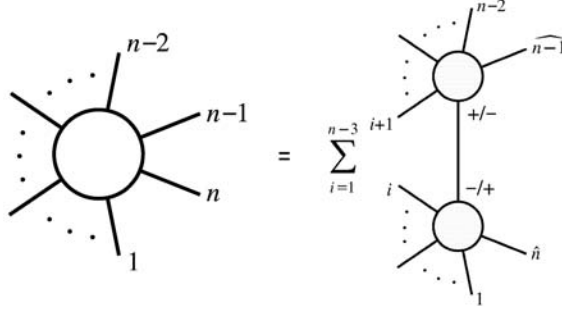


Fig. 2. Expression for the n -point tree amplitude obtained using on-shell recursion, with shifted legs $n-1$ and n .

maintaining the cyclic adjacency within each set. The sum over residues will take the form,

$$\sum_{\text{partitions } P} \text{shifted legs on opposite sides.} \quad (14)$$

For each partition, we will have the sum of the momenta $\hat{K}(z)$ (from either side of the partition) flowing through the propagator. The square of this momentum is linear in z because the shift vector itself has square zero,

$$K_{a\dots j\dots b}^2(z) = K_{a\dots b}^2 - z \langle j^- | K_{a\dots b} | l^- \rangle, \quad (15)$$

where one side of the partition contains the cyclicly-consecutive set of momenta $\{k_a, \dots, k_j, \dots, k_b\}$. The location of the pole is given by setting this expression to zero,

$$z_{ab} = \frac{K_{a\dots b}^2}{\langle j^- | K_{a\dots b} | l^- \rangle}. \quad (16)$$

To compute the value of the residue, we rely on factorization: on the pole, the amplitude will factorize exactly into the product of two on-shell amplitudes, A_L and A_R . The required residue is then,

$$\text{Res}_{z=z_{ab}} \frac{f(z)}{z K_{a\dots b}^2(z)} = A_L(z_{ab}) \times \frac{i}{K_{a\dots b}^2} \times A_R(z_{ab}). \quad (17)$$

The amplitude itself is given by a sum over such residues, illustrated pictorially in fig. 2 for the special choice $j = n-1$ and $l = n$. Because this expresses the amplitude in terms of lower-point on-shell amplitudes, it provides a recursion relation, which ultimately terminates in three-point amplitudes.

The three-gluon amplitude is linear in the momenta; on shell, however, all invariants vanish,

$$k_3^2 = 0 \Rightarrow 0 = (k_1 + k_2)^2 = 2k_1 \cdot k_2, \quad (18)$$

and likewise for the other two products. Accordingly, all spinor products for ordinary momenta vanish, and so does the three-point amplitude. This seems to pose a

problem; how can we obtain a non-vanishing result for any amplitude, when all of them reduce to three-point amplitudes? The resolution again relies on complex momenta; for complex momenta, only one spinor product of the pair $\langle 1\,2 \rangle$ and $[1\,2]$ will vanish, so that the Lorentz invariants still vanish, but we obtain non-vanishing expressions for the amplitudes,

$$A_3(1^-, 2^-, 3^+) = i \frac{\langle 1\,2 \rangle^3}{\langle 2\,3 \rangle \langle 3\,1 \rangle}, \quad A_3(1^+, 2^+, 3^-) = -i \frac{[1\,2]^3}{[2\,3] [3\,1]}. \quad (19)$$

All gluon amplitudes can be built up from these three-point amplitudes.

One can use on-shell recursion techniques in a similar way to obtain expressions for the rational terms in one-loop amplitudes, once the cut-containing parts are known from a unitarity calculation.

2.4. Numerical Methods

As mentioned above, the unitarity and on-shell recursion techniques can be used numerically. Expressions for the integrals are known analytically in terms of logarithms and dilogarithms; these can be evaluated numerically. Tree amplitudes that enter into coefficients can be evaluated numerically, using on-shell recursion relations or older (but more efficient) techniques such as the Berends–Giele recursion relations [23]. The box coefficients are given by products of tree amplitudes, and thus can also be evaluated numerically. Both the subtraction [13] and contour-integral [12] techniques for the evaluation of triangle and bubble coefficients can be carried out numerically. It turns out to be advantageous to combine the two techniques. This improves the numerical stability of the evaluation.

A number of groups are developing numerical codes based on numerical application of the unitarity methods: CutTools+HELAC-1LOOP (OPP and their collaborators [24]); Ellis et al. [25]; Giele et al. [26]; and the BLACKHAT collaboration [27], which also makes use of loop-level on-shell recursion for parts of the evaluation. An example of a result obtained using BLACKHAT along with SHERPA [6] for the real-emission contributions and phase-space integration is shown in fig. 3. The example compares the third-jet E_T spectrum in $W + 3$ -jet events at the Tevatron against CDF data [28], showing good agreement. It shows the reduction in scale dependence, and hence the improvement in theoretical reliability, of the NLO prediction.

A numerical approach makes it possible to construct a general-purpose program, which can address the dozens of processes for which experimenters can make good use of an NLO prediction. Such an approach is probably necessary to move beyond earlier bespoke calculations to industrial handling of large number of processes.

3. $\mathcal{N} = 4$ Supersymmetry: A Theoretical Laboratory

In developing new calculational techniques, it is useful to have a theoretical laboratory in which to try them out. A gauge theory with the simplest possible amplitudes

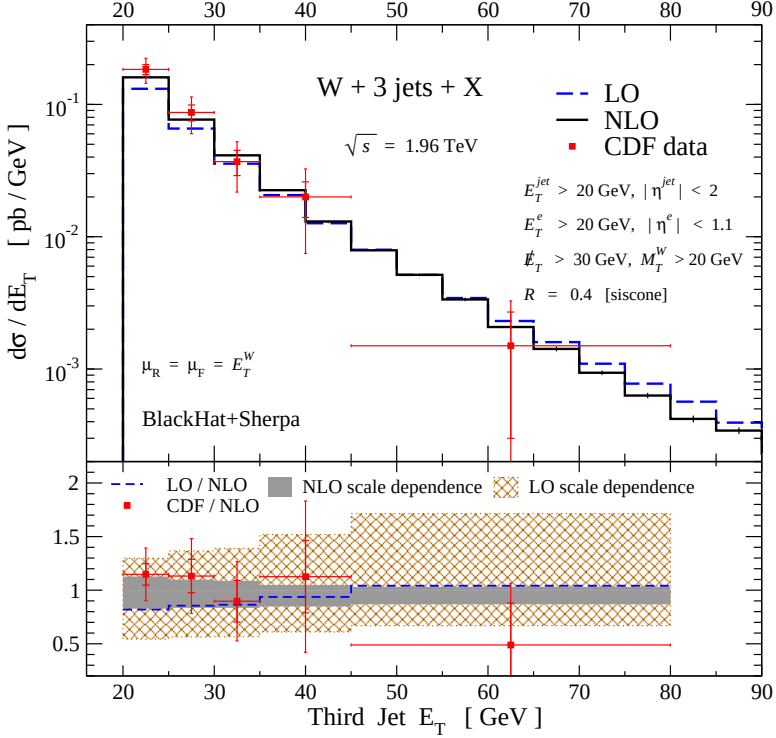


Fig. 3. The measured cross section, $d\sigma(p\bar{p} \rightarrow e\nu + \geq 3 \text{ jets})/dE_T^{\text{3rd-jet}}$, for inclusive $W + 3\text{-jet}$ production, compared to the complete NLO prediction. In the upper panels the NLO distribution is the solid (black) histogram, and CDF data points are the (red) points, whose inner and outer error bars respectively denote the statistical and total uncertainties (excluding the luminosity error) on the measurements added in quadrature. The LO predictions are shown as dashed (blue) lines. The thin vertical lines in the center of each bin (where visible) give the numerical integration errors for that bin. Each lower panel shows the distribution normalized to the complete NLO prediction, using the CDF experimental bins (that is, averaging over bins in the upper panel). The scale-dependence bands are shaded (gray) for NLO and cross-hatched (brown) for LO.

would provide an ideal such laboratory. The maximally supersymmetric gauge theory, which contains four Majorana fermions (“gluinos”) and six real scalars, all transforming in the adjoint representation of the gauge group, is such a theory. Its Lagrangian is much more complicated than that of pure non-Abelian gauge theory. Its amplitudes, however, are much simpler, and most amplitude calculations to date beyond one loop have been done in this theory.

This simplicity starts at tree level, where the supersymmetry Ward identities [29] show that two patterns of gluon helicity amplitudes vanish,

$$\begin{aligned} A_n(1^+, 2^+, \dots, n^+) &= 0, \\ A_n(1^-, 2^+, \dots, n^+) &= 0. \end{aligned} \quad (20)$$

At tree level, $\mathcal{N} = 4$ amplitudes are identical to those in pure gauge theory, so the

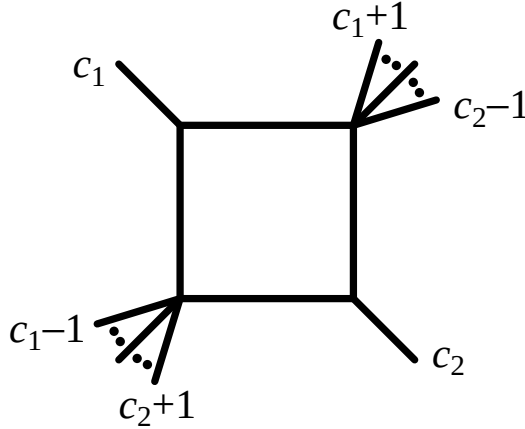


Fig. 4. The easy two-mass box integral $I_4^{2me}(c_1, c_2)$ that appears in the expression for the one-loop MHV amplitude.

simplicity carries over to the latter theory as well.

The simplest non-vanishing amplitudes are those with two opposite-helicity gluons, known as maximally helicity-violating (MHV) amplitudes,

$$A_n^{\text{tree}}(1^+, \dots, m_1^-, (m_1 + 1)^+, \dots, m_2^-, (m_2 + 1)^+, \dots, n^+) = i \frac{\langle m_1 m_2 \rangle^4}{\langle 1 2 \rangle \langle 2 3 \rangle \cdots \langle (n-1) n \rangle \langle n 1 \rangle} \quad (21)$$

whose simplicity was first guessed by Parke and Taylor [30], with the spinor form first written down by Mangano, Parke, and Xu [31]. They were first proven using the Berends–Giele recursion relations [23].

Eqs. (20) hold to all orders in perturbation theory, and the simplicity of MHV amplitudes appears to hold as well. The one-loop MHV amplitude in the $\mathcal{N} = 4$ theory was computed during the original development of the unitarity method [8],

$$A_n^{1\text{-loop}}(1^+, \dots, m_1^-, \dots, m_2^-, \dots, n^+) = -A_n^{\text{tree}}(1^+, \dots, m_1^-, \dots, m_2^-, \dots, n^+) \quad (22)$$

$$\times \frac{1}{2} \sum_{c_1, c_2} (st - m_1^2 m_2^2) I_4^{2me}(c_1, c_2) \quad (23)$$

where I_4^{2me} is the box integral with two massless $(c_{1,2})$ and two massive external legs, at opposite corners, shown in fig. 4; s and t are the invariants of the box, and $m_{1,2}^2$ the masses of the two massive legs. (The massive legs are as above sums of massless external momenta.)

The two-loop computation of the corresponding four-point amplitude, carried out with Anastasiou, Bern, and Dixon [15] completing earlier work of Bern, Rozowsky, and Yan [14], revealed a surprise: the amplitude can be written in terms of

the one-loop amplitude,

$$\frac{1}{A_4^{\text{tree}}} \left(A_4^{1\text{-loop}}(s, t; \epsilon) \right)^2 + f^{(2)}(\epsilon) A_4^{1\text{-loop}}(s, t; 2\epsilon) - \frac{1}{2} \zeta_2^2 A_4^{\text{tree}} + \mathcal{O}(\epsilon). \quad (24)$$

where the amplitude in the second term is evaluated at a different value of the dimensional regulator ϵ in order to ensure that all terms have the same dimension.

The expressibility in terms of the one-loop amplitude was conjectured by Bern, Dixon, and Smirnov [32] to generalize to all loop orders in a specific exponentiated form,

$$\ln \frac{A_n^{\text{all-loop}}}{A_n^{\text{tree}}} = \frac{1}{A_n^{\text{tree}}} \sum_{L=1}^{\infty} a^L \left(f^{(L)}(\epsilon) A_n^{1\text{-loop}}(L\epsilon) + C^{(L)} + \mathcal{O}(\epsilon) \right) \quad (25)$$

and the generalization indeed holds for the four- and five-point amplitudes, though there are now known to be corrections to this formula for amplitudes with six or more external legs.

Remarkably, Alday and Maldacena [33] were able to perform the strong-coupling ($\alpha_s N_c \rightarrow \infty$) computation of the four-point amplitude. They used string theory and relied on the duality between string theory in an anti-de Sitter background and the $\mathcal{N} = 4$ gauge theory. Their result,

$$\begin{aligned} \ln \frac{A_4}{A_4^{\text{tree}}} = & -\frac{1}{8\epsilon^2} f^{(-2)} \left(\frac{\lambda_4 \mu^{2\epsilon}}{s^\epsilon} \right) - \frac{1}{4\epsilon} g^{(-1)} \left(\frac{\lambda_4 \mu^{2\epsilon}}{s^\epsilon} \right) + (s \rightarrow t) \\ & + \frac{f(\lambda)}{8} [\ln^2 s/t + 4\pi^2/3] + C(\lambda), \end{aligned} \quad (26)$$

also agrees with the exponentiation ansatz of eq. (25).

Amplitudes in $\mathcal{N} = 4$ gauge theory are simpler than those in QCD because the theory has greater symmetry. In particular, it has a configuration-space superconformal symmetry. This symmetry is exact at tree level; while it is broken by infrared singularities at loop level, in a sense it is still present in the finite terms.

It also has another superconformal symmetry, in momentum space. This symmetry, elucidated by Drummond, Henn, Korchemsky, Sokatchev, and Smirnov [34, 35] (see also ref. [36]), is called dual superconformal symmetry. Tree-level amplitudes have the full dual superconformal symmetry, while one-loop amplitudes are dual conformal invariant up to a known dual conformal anomaly. The dependence on loop integrals breaks the superconformal part of the symmetry completely. This symmetry is manifest in a completely different representation for loop amplitudes, in the form of momentum-space Wilson loops. It has been used to prove the exponentiation hypothesis (and hence the link to the Alday–Maldacena result) for four- and five-point amplitudes.

The origins of this symmetry and its full implication for the amplitudes in the theory are not yet understood, and remain an active area for future research. The hint of additional structure in gauge theories is a tantalizing invitation to further calculations.

Acknowledgments

It is a pleasure to thank Zvi Bern, Lance Dixon, and Dave Dunbar for extensive collaboration over many years of developing on-shell methods. I also wish to thank younger and more recent collaborators who have worked on developing the BlackHat program library: Carola Berger, Fernando Febres Cordero, Darren Forde, Tanju Gleisberg, Harald Ita, and Daniel Maître. The author's work is supported by the European Research Council under Advanced Investigator Grant ERC-AdG-228301.

References

- [1] L. M. Brown and R. P. Feynman, *Phys. Rev.* **85**, 231 (1952);
L.M. Brown, *Nuovo Cim.* **21**, 3878 (1961);
G. Passarino and M. J. G. Veltman, *Nucl. Phys. B* **160**, 151 (1979).
- [2] B. Petersson, *J. Math. Phys.* **6**, 1955 (1965);
G. Källén and J.S. Toll, *J. Math. Phys.* **6**, 299 (1965);
D. B. Melrose, *Nuovo Cim.* **40**, 181 (1965).
- [3] Z. Xu, D. H. Zhang and L. Chang, TUTP-84/3-TSINGHUA;
Z. Xu, D. H. Zhang and L. Chang, *Nucl. Phys. B* **291**, 392 (1987).
- [4] F. A. Berends, R. Kleiss, P. De Causmaecker, R. Gastmans and T. T. Wu, *Phys. Lett. B* **103**, 124 (1981);
P. De Causmaecker, R. Gastmans, W. Troost and T. T. Wu, *Nucl. Phys. B* **206**, 53 (1982);
R. Kleiss and W. J. Stirling, *Nucl. Phys. B* **262**, 235 (1985);
J. F. Gunion and Z. Kunszt, *Phys. Lett. B* **161**, 333 (1985).
- [5] T. Stelzer and W. F. Long, *Comput. Phys. Commun.* **81**, 357 (1994) [[hep-ph/9401258](#)];
A. Pukhov *et al.*, [hep-ph/9908288](#);
A. Kanaki and C. G. Papadopoulos, *Comput. Phys. Commun.* **132**, 306 (2000) [[hep-ph/0002082](#)];
M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, *JHEP* **0307**, 001 (2003) [[hep-ph/0206293](#)];
F. Krauss, R. Kuhn and G. Soff, *JHEP* **0202**, 044 (2002) [[hep-ph/0109036](#)].
- [6] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert and J. Winter, *JHEP* **0902**, 007 (2009) [[0811.4622 \[hep-ph\]](#)].
- [7] W. L. van Neerven and J. A. M. Vermaseren, *Phys. Lett. B* **137**, 241 (1984);
A. I. Davydychev, *Phys. Lett. B* **263**, 107 (1991);
Z. Bern, L. J. Dixon and D. A. Kosower, *Phys. Lett. B* **302**, 299 (1993) [Erratum-ibid. *B* **318**, 649 (1993)] [[hep-ph/9212308](#)];
J. Fleischer, F. Jegerlehner and O. V. Tarasov, *Nucl. Phys. B* **566**, 423 (2000) [[hep-ph/9907327](#)];
T. Binoth, J. P. Guillet and G. Heinrich, *Nucl. Phys. B* **572**, 361 (2000) [[hep-ph/9911342](#)];
G. Duplancić and B. Nizić, *Eur. Phys. J. C* **35**, 105 (2004) [[hep-ph/0303184](#)];
A. Denner and S. Dittmaier, *Nucl. Phys. B* **734**, 62 (2006) [[arXiv:hep-ph/0509141](#)].
- [8] Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, *Nucl. Phys. B* **425**, 217 (1994) [[arXiv:hep-ph/9403226](#)];
Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, *Nucl. Phys. B* **435**, 59 (1995) [[arXiv:hep-ph/9409265](#)].
- [9] V. P. Nair, *Phys. Lett. B* **214**:215 (1988);
E. Witten, *Commun. Math. Phys.* **252**, 189 (2004) [[arXiv:hep-th/0312171](#)].

- [10] Z. Bern, L. J. Dixon and D. A. Kosower, Nucl. Phys. B **513**, 3 (1998) [hep-ph/9708239].
- [11] R. Britto, F. Cachazo and B. Feng, Nucl. Phys. B **725**, 275 (2005) [hep-th/0412103].
- [12] D. Forde, Phys. Rev. D **75**, 125019 (2007) [0704.1835 [hep-ph]].
- [13] G. Ossola, C. G. Papadopoulos and R. Pittau, Nucl. Phys. B **763**, 147 (2007) [hep-ph/0609007].
- [14] Z. Bern, J. S. Rozowsky and B. Yan, Phys. Lett. B **401**, 273 (1997) [arXiv:hep-ph/9702424].
- [15] C. Anastasiou, Z. Bern, L. J. Dixon and D. A. Kosower, Phys. Rev. Lett. **91**, 251602 (2003) [arXiv:hep-th/0309040].
- [16] Z. Bern, L. J. Dixon and D. A. Kosower, JHEP **0001**, 027 (2000) [arXiv:hep-ph/0001001];
 Z. Bern, A. De Freitas and L. J. Dixon, JHEP **0109**, 037 (2001) [arXiv:hep-ph/0109078];
 Z. Bern, A. De Freitas, L. J. Dixon and H. L. Wong, Phys. Rev. D **66**, 085002 (2002) [arXiv:hep-ph/0202271];
 Z. Bern, L. J. Dixon and D. A. Kosower, JHEP **0408**, 012 (2004) [arXiv:hep-ph/0404293];
 A. De Freitas and Z. Bern, JHEP **0409**, 039 (2004) [arXiv:hep-ph/0409007];
 Z. Bern, M. Czakon, D. A. Kosower, R. Roiban and V. A. Smirnov, Phys. Rev. Lett. **97**, 181601 (2006) [arXiv:hep-th/0604074];
 Z. Bern, M. Czakon, L. J. Dixon, D. A. Kosower and V. A. Smirnov, Phys. Rev. D **75**, 085010 (2007) [arXiv:hep-th/0610248];
 Z. Bern, J. J. Carrasco, L. J. Dixon, H. Johansson, D. A. Kosower and R. Roiban, Phys. Rev. Lett. **98**, 161303 (2007) [arXiv:hep-th/0702112];
 Z. Bern, L. J. Dixon, D. A. Kosower, R. Roiban, M. Spradlin, C. Vergu and A. Volovich, Phys. Rev. D **78**, 045007 (2008) [arXiv:0803.1465 [hep-th]];
 F. Cachazo, M. Spradlin and A. Volovich, Phys. Rev. D **78**, 105022 (2008) [arXiv:0805.4832 [hep-th]];
 Z. Bern, J. J. Carrasco, L. J. Dixon, H. Johansson and R. Roiban, Phys. Rev. Lett. **103**, 081301 (2009) [arXiv:0905.2326 [hep-th]].
- [17] F. V. Tkachov, Phys. Lett. B **100**, 65 (1981);
 K. G. Chetyrkin and F. V. Tkachov, Nucl. Phys. B **192**, 159 (1981).
- [18] S. Laporta, Phys. Lett. B **504**, 188 (2001) [hep-ph/0102032];
 S. Laporta, Int. J. Mod. Phys. A **15**, 5087 (2000) [hep-ph/0102033];
 C. Anastasiou and A. Lazopoulos, JHEP **0407**, 046 (2004) [arXiv:hep-ph/0404258];
 A. V. Smirnov, JHEP **0810**, 107 (2008) [arXiv:0807.3243 [hep-ph]].
- [19] E. Remiddi, Nuovo Cim. A **110**, 1435 (1997) [arXiv:hep-th/9711188];
 M. Caffo, H. Czyz, S. Laporta and E. Remiddi, Nuovo Cim. A **111**, 365 (1998) [arXiv:hep-th/9805118];
 T. Gehrmann and E. Remiddi, Nucl. Phys. B **580**, 485 (2000) [arXiv:hep-ph/9912329].
- [20] V. A. Smirnov, Phys. Lett. B **460**, 397 (1999) [arXiv:hep-ph/9905323];
 J. B. Tausk, Phys. Lett. B **469**, 225 (1999) [arXiv:hep-ph/9909506];
 M. Czakon, Comput. Phys. Commun. **175**, 559 (2006) [arXiv:hep-ph/0511200];
 A. V. Smirnov and V. A. Smirnov, Eur. Phys. J. C **62**, 445 (2009) [arXiv:0901.0386 [hep-ph]].
- [21] C. Anastasiou, S. Beerli and A. Daleo, JHEP **0705**, 071 (2007) [arXiv:hep-ph/0703282].
- [22] R. Britto, F. Cachazo, B. Feng and E. Witten, Phys. Rev. Lett. **94**, 181602 (2005) [hep-th/0501052].

- [23] F. A. Berends and W. T. Giele, Nucl. Phys. B **306**, 759 (1988).
- [24] G. Ossola, C. G. Papadopoulos and R. Pittau, JHEP **0803**, 042 (2008) [arXiv:0711.3596 [hep-ph]];
P. Mastrolia, G. Ossola, C. G. Papadopoulos and R. Pittau, JHEP **0806**, 030 (2008) [arXiv:0803.3964 [hep-ph]];
A. van Hameren, C. G. Papadopoulos and R. Pittau, JHEP **0909**, 106 (2009) [arXiv:0903.4665 [hep-ph]];
G. Bevilacqua, M. Czakon, C. G. Papadopoulos, R. Pittau and M. Worek, JHEP **0909**, 109 (2009) [arXiv:0907.4723 [hep-ph]].
- [25] R. K. Ellis, W. T. Giele and Z. Kunszt, JHEP **0803**, 003 (2008) [arXiv:0708.2398 [hep-ph]];
R. K. Ellis, W. T. Giele, Z. Kunszt, K. Melnikov and G. Zanderighi, JHEP **0901**, 012 (2009) [arXiv:0810.2762 [hep-ph]];
A. Lazopoulos, arXiv:0812.2998 [hep-ph];
R. K. Ellis, K. Melnikov and G. Zanderighi, JHEP **0904**, 077 (2009) [arXiv:0901.4101 [hep-ph]];
R. Keith Ellis, K. Melnikov and G. Zanderighi, arXiv:0906.1445 [hep-ph].
- [26] W. T. Giele, Z. Kunszt and K. Melnikov, JHEP **0804**, 049 (2008) [arXiv:0801.2237 [hep-ph]];
W. T. Giele and G. Zanderighi, JHEP **0806**, 038 (2008) [arXiv:0805.2152 [hep-ph]];
W. T. Giele, Z. Kunszt, and J. Winter, presentation at RADCOR 2009, Ascona, Switzerland.
- [27] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, H. Ita, D. A. Kosower, D. Maître, Phys. Rev. D **78**, 036003 (2008) [arXiv:0803.4180 [hep-ph]];
C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, D. Maître, Phys. Rev. Lett. **102**, 222001 (2009) [arXiv:0902.2760 [hep-ph]];
C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, D. Maître, arXiv:0907.1984 [hep-ph].
- [28] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. D **77**, 011108 (2008) [0711.4044 [hep-ex]].
- [29] M. T. Grisaru, H. N. Pendleton and P. van Nieuwenhuizen, Phys. Rev. D **15**, 996 (1977);
M. T. Grisaru and H. N. Pendleton, Nucl. Phys. B **124**, 81 (1977);
S. J. Parke and T. R. Taylor, Phys. Lett. B **157**, 81 (1985) [Erratum-ibid. **174B**, 465 (1986)];
Z. Kunszt, Nucl. Phys. B **271**, 333 (1986).
- [30] S. J. Parke and T. R. Taylor, Phys. Rev. Lett. **56**, 2459 (1986).
- [31] M. L. Mangano, S. J. Parke and Z. Xu, Nucl. Phys. B **298**, 653 (1988).
- [32] Z. Bern, L. J. Dixon and V. A. Smirnov, Phys. Rev. D **72**, 085001 (2005) [arXiv:hep-th/0505205].
- [33] L. F. Alday and J. M. Maldacena, JHEP **0706**, 064 (2007) [arXiv:0705.0303 [hep-th]].
- [34] J. M. Drummond, J. Henn, V. A. Smirnov and E. Sokatchev, JHEP **0701**, 064 (2007) [arXiv:hep-th/0607160].
- [35] J. M. Drummond, G. P. Korchemsky and E. Sokatchev, Nucl. Phys. B **795**, 385 (2008) [arXiv:0707.0243 [hep-th]];
J. M. Drummond, J. Henn, G. P. Korchemsky and E. Sokatchev, Nucl. Phys. B **795**, 52 (2008) [arXiv:0709.2368 [hep-th]];
J. M. Drummond, J. Henn, G. P. Korchemsky and E. Sokatchev, Phys. Lett. B **662**, 456 (2008) [arXiv:0712.4138 [hep-th]];

- J. M. Drummond, J. Henn, G. P. Korchemsky and E. Sokatchev, Nucl. Phys. B **815**, 142 (2009) [arXiv:0803.1466 [hep-th]];
- J. M. Drummond, J. Henn, G. P. Korchemsky and E. Sokatchev, arXiv:0807.1095 [hep-th];
- G. P. Korchemsky and E. Sokatchev, arXiv:0906.1737 [hep-th].
- [36] A. Brandhuber, P. Heslop and G. Travaglini, Nucl. Phys. B **794**, 231 (2008) [arXiv:0707.1153 [hep-th]];
- C. Anastasiou, A. Brandhuber, P. Heslop, V. V. Khoze, B. Spence and G. Travaglini, JHEP **0905**, 115 (2009) [arXiv:0902.2245 [hep-th]];
- A. Brandhuber, P. Heslop and G. Travaglini, arXiv:0906.3552 [hep-th].

OPERATOR ALGEBRAS AND NONCOMMUTATIVE GEOMETRIC ASPECTS IN CONFORMAL FIELD THEORY*

ROBERTO LONGO

Università di Roma Tor Vergata

E-mail: longo@mat.uniroma2.it

The Operator Algebraic approach to Conformal Field Theory has been particularly fruitful in recent years (leading for example to the classification of all local conformal nets on the circle with central charge $c < 1$, jointly with Y. Kawahigashi). On the other hand the Operator Algebraic viewpoint offers a natural perspective for a Noncommutative Geometric context within Conformal Field Theory. One basic point here is to uncover the relevant structures. In this talk I will explain some of the basic steps in this "Noncommutative Geometrization program" up to the recent construction of a spectral triple associated with certain Ramond representations of the Supersymmetric Virasoro net. So Alain Connes framework enters into play. This is a joint work with S. Carpi, Y. Kawahigashi, and R. Hillier.

*The author chose not to contribute to this proceedings volume.

KINETIC TRANSPORT IN CRYSTALS

JENS MARKLOF

School of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.

E-mail: j.marklof@bristol.ac.uk

One of the central challenges in kinetic theory is the derivation of macroscopic evolution equations—describing, for example, the dynamics of an electron gas—from the underlying fundamental microscopic laws of classical or quantum mechanics. An iconic mathematical model in this research area is the Lorentz gas, which describes an ensemble of non-interacting point particles in an infinite array of spherical scatterers. In the case of a disordered scatterer configuration, the classical results by Gallavotti, Spohn and Boldrighini-Bunimovich-Sinai show that the time evolution of a macroscopic particle cloud is governed, in the limit of small scatterer density (Boltzmann-Grad limit), by the linear Boltzmann equation. In this lecture I will discuss the recent discovery that for a periodic configuration of scatterers the linear Boltzmann equation fails, and the random flight process that emerges in the Boltzmann-Grad limit is substantially more complicated. The key ingredient in the description of the limiting stochastic process is the renormalization dynamics on the space of lattices, a powerful technique that has recently been successfully applied also to other open problems in mathematical physics, including KAM theory and quantum chaos. This lecture is based on joint work with Andreas Strömbergsson, Uppsala.

1. Introduction

An important cornerstone in mathematical physics is the problem of deriving macroscopic evolution equations from first principles, i.e., the microscopic laws of motion governed by quantum theory, or (to simplify) Newton’s laws of classical mechanics. The subject has its origin in Boltzmann’s revolutionary vision formulated more than a century ago, and it is perhaps surprising that still today there is no complete understanding of his most fundamental model, a dilute gas of hard spheres: Lanford’s seminal work [29] establishes the validity of the Boltzmann equation only for times that are a fraction of the mean collision time. Even simpler models, such as a gas of non-interacting particles in a fixed array of scatterers studied first by Lorentz in 1905 [31] (the *Lorentz gas*), are difficult to analyze and lead to new and unexpected macroscopic phenomena. This lecture will focus on the Lorentz gas and report on recent joint work with Andreas Strömbergsson [37], [38], [39], [40] in the case of a crystal with periodic scatterer configuration. The exciting aspect of our findings is that the kinetic transport equations that emerge in the limit of small scatter size (the *Boltzmann-Grad limit*) are new, and distinctly different from the answer for disordered scatterers.

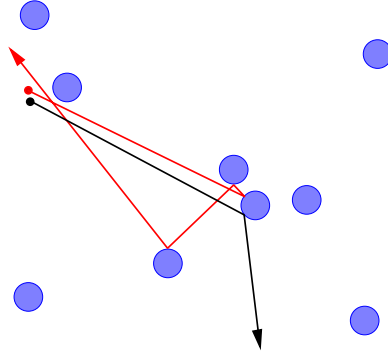


Fig. 1. The Lorentz gas for a random scatterer configuration.

In his original work, Boltzmann considered the case of a dilute gas of hard spheres with elastic collisions; *dilute* refers to the limit of small particle density, i.e., the ratio of particle radius and mean separation tends to zero. In the case of a gas of non-interacting particles in an array of fixed scatterers—the Lorentz gas—the dynamics can be reduced to a one-particle motion (Fig. 1). We denote by $\mathbf{q}(t) \in \mathbb{R}^d$ the position and by $\mathbf{v}(t) \in \mathbb{R}^d$ the velocity of our particle. The correct kinetic limit is then obtained by taking the scatterer radius $\rho \rightarrow 0$. A simple dimensional argument shows that in this limit the mean free path length should scale like $\rho^{-(d-1)}$, i.e., the inverse of the total scattering cross section of an individual scatterer. This suggests to rescale the length units by introducing the macroscopic coordinates

$$(\mathbf{Q}(t), \mathbf{V}(t)) = (\rho^{d-1} \mathbf{q}(\rho^{-(d-1)}t), \mathbf{v}(\rho^{-(d-1)}t)). \quad (1)$$

This rescaling of length and time is commonly referred to as the *Boltzmann-Grad scaling*, and the corresponding limit $\rho \rightarrow 0$ as the *Boltzmann-Grad limit*. The time evolution of a particle with initial data $(\mathbf{Q}, \mathbf{V})$ is then described by the billiard flow

$$(\mathbf{Q}(t), \mathbf{V}(t)) = \Phi_\rho^t(\mathbf{Q}, \mathbf{V}). \quad (2)$$

Since the speed of our particle is a constant of motion we may assume without loss of generality that $\|\mathbf{V}\| = 1$. For notational reasons it is convenient to extend the dynamics to the inside of each scatterer trivially, i.e., set $\Phi_\rho^t = \text{id}$ whenever $\mathbf{Q}$ is inside the scatterer. That is, the relevant phase space is now the unit tangent bundle of $\mathbb{R}^d$, which will be denoted by $T^1(\mathbb{R}^d)$.

The time evolution of an initial particle density $f \in L^1(T^1(\mathbb{R}^d))$ is

$$f_t = L_\rho^t f \quad (3)$$

where L_ρ^t is the Liouville operator defined by

$$[L_\rho^t f](\mathbf{Q}, \mathbf{V}) := f(\Phi_\rho^{-t}(\mathbf{Q}, \mathbf{V})). \quad (4)$$

Following Boltzmann's arguments, Lorentz concluded in his 1905 paper that the macroscopic time evolution of a particle cloud should, in the limit $\rho \rightarrow 0$, be gov-

erned by the *linear Boltzmann equation* (today also referred to as *kinetic Lorentz equation*),

$$\left[\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla_{\mathbf{Q}} \right] f_t(\mathbf{Q}, \mathbf{V}) = \int_{\mathbb{S}_1^{d-1}} [f_t(\mathbf{Q}, \mathbf{V}_0) - f_t(\mathbf{Q}, \mathbf{V})] \sigma(\mathbf{V}_0, \mathbf{V}) d\mathbf{V}_0, \quad (5)$$

where the collision kernel $\sigma(\mathbf{V}_0, \mathbf{V})$ is the differential cross section of the individual scatterer. In the case of elastic scattering at hard spheres we have $\sigma(\mathbf{V}_0, \mathbf{V}) = \frac{1}{4} \|\mathbf{V}_0 - \mathbf{V}\|^{3-d}$. The linear Boltzmann equation describes a random flight process, where a particle moves freely with constant velocity $\mathbf{V}_0$ for time $t \geq \xi$ with probability $e^{-\nu_d \xi}$, where ν_d is the volume of the $(d-1)$ -dimensional unit ball (the total scattering cross section), and is then scattered to velocity $\mathbf{V}$ with probability $\sigma(\mathbf{V}_0, \mathbf{V})$, again flies with constant velocity $\mathbf{V}$ for time $t \geq \xi$ with probability $e^{-\nu_d \xi}$, and so on. The crucial observation is that each scattering event is independent of the previous one. Thus the process that generates the paths of our random flight is Markovian. As we will see, this is different in the case of a periodic scatterer configuration.

The validity of the linear Boltzmann equation was first established rigorously for a random, Poisson distributed scatterer configuration by Gallavotti [23]. His results were generalized by Spohn [47] to more general random scatterer configurations and scattering potentials. In 1983, Boldrighini, Bunimovich and Sinai [6] proved convergence for almost every scatterer configuration drawn from a Poisson distribution. As Spohn's work shows, the details of the randomness of the scatterer positions is not so essential, and as long as there are no strong correlations, all of the above results should remain valid.

The linear Boltzmann equation has numerous important applications, e.g., to neutron transport and radiative transfer, and it is thus natural to ask under which circumstances it may fail to provide an accurate description.

2. The periodic Lorentz gas

Given a Euclidean lattice $\mathcal{L} \subset \mathbb{R}^d$ of covolume one (e.g., $\mathcal{L} = \mathbb{Z}^d$) the *periodic Lorentz gas* is defined as the dynamics of a cloud of non-interacting point particles in an array of identical scatterers that are placed at the vertices of the lattice $\mathcal{L}$ (Fig. 2). The periodic Lorentz gas has served as a fundamental model in the understanding of chaotic diffusion, which emerges in the long-time limit with *fixed* scatterer size. In their pioneering work, Bunimovich and Sinai [8] proved a central limit theorem for the long-time dynamics of a particle cloud in two dimensions. That is, the long-time evolution converges, in the appropriate scaling limit, to a solution of the heat equation. More refined statistical properties that show that the dynamics in fact converges to Brownian motion, have recently been established in the work of Melbourne and Nichol [41], [42], and the recent paper by Dolgopyat, Szasz and Varju [18]. All of the above results assume that the Lorentz gas has *finite horizon*, i.e., there are no infinitely long free flight paths. Without this assumption, a central limit theorem still holds, but the diffusion constant is no longer linear in time

t but diverges as $t \log t$. This was observed by Bleher [3] and recently established rigorously by Szasz and Varju [48]. It is interesting that none of these results have so far been extended to dimension $d \geq 3$. The arguments given by Chernov [12] and by Balint and Toth [1] require non-trivial hypotheses that are difficult to establish.

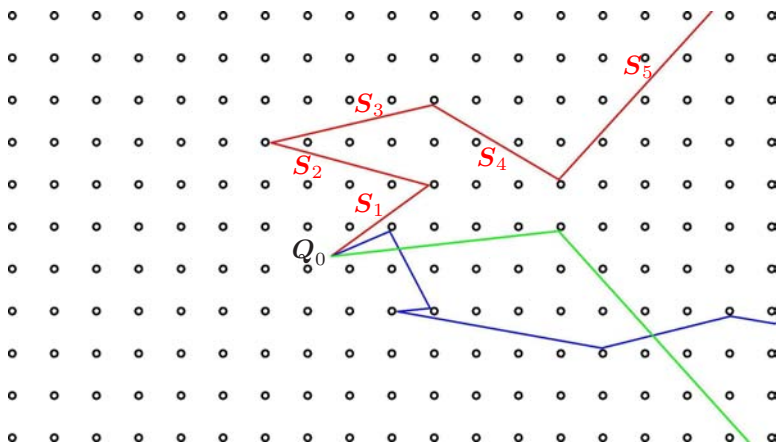


Fig. 2. The Lorentz gas for a periodic scatterer configuration, with three distinct particle trajectories starting at the point Q_0 .

We now return to the question of the existence of kinetic transport equations in the Boltzmann-Grad limit discussed in the previous section. Recall that we are interested in the dynamics of a particle cloud

$$f_t(Q, V) = [L_\rho^t f](Q, V) = f(\Phi_\rho^{-t}(Q, V)) \quad (6)$$

in the macroscopic coordinates (1). The estimates by Bourgain, Golse and Wennberg [7], [24] on the distribution of free path lengths already imply that the linear Boltzmann equation does not hold in the periodic set-up; this was pointed out recently by Golse [26].

The first key result of our joint work with Strömbergsson is the proof of the existence of a limiting random process for the periodic Lorentz gas [38, Section 1].

Theorem 1. *Fix a Euclidean lattice $\mathcal{L}$. For every $t > 0$ there exists a linear operator $L^t : L^1(T^1(\mathbb{R}^d)) \rightarrow L^1(T^1(\mathbb{R}^d))$, such that for every $f \in L^1(T^1(\mathbb{R}^d))$ and any set $\mathcal{A} \subset T^1(\mathbb{R}^d)$ with boundary of Lebesgue measure zero,*

$$\lim_{\rho \rightarrow 0} \int_{\mathcal{A}} [L_\rho^t f](Q, V) dQ dV = \int_{\mathcal{A}} [L^t f](Q, V) dQ dV. \quad (7)$$

The operator L^t thus describes the macroscopic diffusion of the Lorentz gas in the Boltzmann-Grad limit $\rho \rightarrow 0$. As we shall see however, the family $\{L^t\}_{t \geq 0}$ does not form a semigroup, i.e.,

$$L^s L^t \neq L^{s+t}. \quad (8)$$

This is perhaps surprising since $\{L_\rho^t\}_{t \geq 0}$ is indeed a semigroup for every fixed $\rho > 0$. What is more, in the case of the random scatterer configuration the corresponding limiting operators L^t also form a semigroup—after all, $f_t := L^t f$ is a solution of the linear Boltzmann equation. The reason for the failure of the semigroup property in the periodic setting stems from additional correlations in the lattice, which are lost in the macroscopic scaling limit. To keep track of this data, we consider extended phase space coordinates $(\mathbf{Q}, \mathbf{V}, \xi, \mathbf{V}_+)$ where $(\mathbf{Q}, \mathbf{V}) \in \mathbb{T}^1(\mathbb{R}^d)$ is the usual position and momentum, $\xi \in \mathbb{R}_+$ the flight time until the next collision, and $\mathbf{V}_+ \in S_1^{d-1}$ the velocity after the next collision. On the microscopic level, the system is now over-determined (ξ and $\mathbf{V}_+$ are functions of $\mathbf{Q}$ and $\mathbf{V}$), but on the macroscopic scale the extra variables are needed. We prove in [38, Section 6] that the particle density $f_t(\mathbf{Q}, \mathbf{V}, \xi, \mathbf{V}_+)$ indeed satisfies a *generalized* linear Boltzmann equation

$$\left[\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla_{\mathbf{Q}} - \frac{\partial}{\partial \xi} \right] f_t(\mathbf{Q}, \mathbf{V}, \xi, \mathbf{V}_+) = \int_{S_1^{d-1}} f_t(\mathbf{Q}, \mathbf{V}_0, 0, \mathbf{V}) p_0(\mathbf{V}_0, \mathbf{V}, \xi, \mathbf{V}_+) d\mathbf{V}_0. \quad (9)$$

The left hand side again corresponds to free transport (note that ξ is decreasing linearly with t). The right hand side involves a new collision kernel $p_0(\mathbf{V}_0, \mathbf{V}, \xi, \mathbf{V}_+)$, given by

$$p_0(\mathbf{V}_0, \mathbf{V}, \xi, \mathbf{V}_+) = \sigma(\mathbf{V}, \mathbf{V}_+) \Phi_0(\xi, \mathbf{b}(\mathbf{V}, \mathbf{V}_+), -\mathbf{s}(\mathbf{V}, \mathbf{V}_0)) \quad (10)$$

where $\sigma(\mathbf{V}, \mathbf{V}_+)$ is the differential cross section and $\Phi_0(\xi, \mathbf{b}(\mathbf{V}, \mathbf{V}_+), -\mathbf{s}(\mathbf{V}, \mathbf{V}_0))$ the transition probability density to leave a scatterer with exit parameter $\mathbf{s}(\mathbf{V}, \mathbf{V}_0)$ and hit the next scatterer at time ξ with impact parameter $\mathbf{b}(\mathbf{V}, \mathbf{V}_+)$; cf. Fig. 3.

In dimension $d = 2$ we have the following explicit formula for the transition probability [39]:

$$\Phi_0(\xi, w, z) = \frac{6}{\pi^2} \Upsilon \left(1 + \frac{\xi^{-1} - \max(|w|, |z|) - 1}{|w + z|} \right) \quad (11)$$

with

$$\Upsilon(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 < x < 1 \\ 1 & \text{if } 1 \leq x, \end{cases} \quad (12)$$

The same formula has recently been found independently by Caglioti and Golse [11] and by Bykovskii and Ustinov [9], using different methods based on continued fractions.

Our formulas for dimension $d \geq 3$ are not as explicit and substantially more involved, see [37], [38] for details, and [40] for asymptotic tail estimates. The formulas imply in particular that the collision kernel p_0 (and thus the limiting process) is *independent* of the lattice $\mathcal{L}$ on which the scatterers are positioned. Hence any microscopic preference for certain directions completely disappears in the Boltzmann-Grad limit.

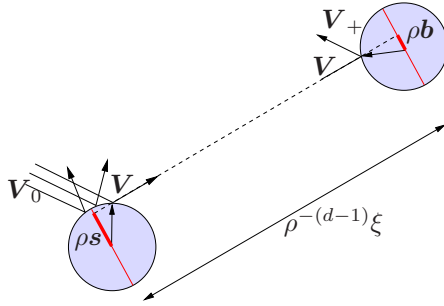


Fig. 3. Two consecutive collisions in the Lorentz gas.

The operators L^t in Theorem 1 we were originally interested in can be recovered by integrating over the auxiliary variables ξ and $\mathbf{V}_+$,

$$[L^t g](\mathbf{Q}, \mathbf{V}) := \int_0^\infty \int_{\mathbb{S}_1^{d-1}} f_t(\mathbf{Q}, \mathbf{V}, \xi, \mathbf{V}_+) d\mathbf{V}_+ d\xi \quad (13)$$

where $f_t(\mathbf{Q}, \mathbf{V}, \xi, \mathbf{V}_+)$ is a solution of the generalized linear Boltzmann equation subject to the initial condition

$$\lim_{t \rightarrow 0} f_t(\mathbf{Q}, \mathbf{V}, \xi, \mathbf{V}_+) = g(\mathbf{Q}, \mathbf{V}) p(\mathbf{V}, \xi, \mathbf{V}_+) \quad (14)$$

with

$$p(\mathbf{V}, \xi, \mathbf{V}_+) := \int_\xi^\infty \int_{\mathbb{S}_1^{d-1}} \sigma(\mathbf{V}_0, \mathbf{V}) p_0(\mathbf{V}_0, \mathbf{V}, \xi', \mathbf{V}_+) d\mathbf{V}_0 d\xi'; \quad (15)$$

the latter is a stationary solution of the generalized linear Boltzmann equation.

3. Why “a generalization” of the linear Boltzmann equation?

The reason why (9) is indeed a generalization of the linear Boltzmann equation is the following. As mentioned in the introduction, the linear Boltzmann equation corresponds to a random flight process where the time ξ until the next collision has probability density $\nu_d e^{-\nu_d \xi}$, where ν_d is the volume of the $(d-1)$ -dimensional unit ball. Furthermore, the probability to exit with parameter $\mathbf{s}(\mathbf{V}, \mathbf{V}_0)$ and hit the next scatterer with impact parameter $\mathbf{b}(\mathbf{V}, \mathbf{V}_+)$ should be uncorrelated and independent of ξ . We have thus

$$\Phi_0(\xi, \mathbf{b}(\mathbf{V}, \mathbf{V}_+), -\mathbf{s}(\mathbf{V}, \mathbf{V}_0)) = e^{-\nu_d \xi}. \quad (16)$$

Substituting in the above the transition density for the random (rather than periodic) scatterer configuration, we obtain

$$p_0(\mathbf{V}_0, \mathbf{V}, \xi, \mathbf{V}_+) = \sigma(\mathbf{V}, \mathbf{V}_+) e^{-\nu_d \xi} = p(\mathbf{V}, \xi, \mathbf{V}_+) \quad (17)$$

and

$$f_t(\mathbf{Q}, \mathbf{V}, \xi, \mathbf{V}_+) = g_t(\mathbf{Q}, \mathbf{V}) \sigma(\mathbf{V}, \mathbf{V}_+) e^{-\nu_d \xi}. \quad (18)$$

It is now straightforward to see that (9) yields the classical linear Boltzmann equation for $g_t(\mathbf{Q}, \mathbf{V})$.

4. Joint distribution of path segments

The following theorem is the central result in our investigation [38]. It shows that the limiting random flight process exists, and that the consecutive path segments $\mathbf{S}_1, \mathbf{S}_2, \dots$ are generated by a Markov process with memory two. That is, the probability for the n th leg of our path only depends on the previous two path segments.

We will use the notation $\widehat{\mathbf{S}} := \|\mathbf{S}\|^{-1}\mathbf{S}$. We set

$$\mathcal{B}_n := \{(\mathbf{S}_1, \dots, \mathbf{S}_n) \in (\mathbb{R}^d \setminus \{\mathbf{0}\})^n : \widehat{\mathbf{S}}_{j+1} \neq \widehat{\mathbf{S}}_j \ (j = 1, \dots, n-1)\}. \quad (19)$$

Theorem 2. *Fix a lattice $\mathcal{L}$ and let Λ be a Borel probability measure on $T^1(\mathbb{R}^d)$ which is absolutely continuous with respect to Lebesgue measure. Then, for each $n \in \mathbb{Z}_{>0}$, and for any set $\mathcal{A} \subset \mathbb{R}^d \times \mathbb{R}^{nd}$ with boundary of Lebesgue measure zero,*

$$\begin{aligned} \lim_{\rho \rightarrow 0} \Lambda(\{(\mathbf{Q}_0, \mathbf{V}_0) \in T^1(\rho^{d-1}\mathcal{K}_\rho) : (\mathbf{Q}_0, \mathbf{S}_1(\mathbf{Q}_0, \mathbf{V}_0; \rho), \dots, \mathbf{S}_n(\mathbf{Q}_0, \mathbf{V}_0; \rho)) \in \mathcal{A}\}) \\ = \int_{\mathcal{A}} P^{(n)}(\mathbf{S}_1, \dots, \mathbf{S}_n) \Lambda'(\mathbf{Q}_0, \widehat{\mathbf{S}}_1) d\mathbf{Q}_0 d\mathbf{S}_1 \cdots d\mathbf{S}_n, \end{aligned} \quad (20)$$

and where Λ' is the Radon-Nikodym derivative of Λ with respect to Lebesgue measure. Furthermore, there is a function $\Psi : \mathcal{B}_3 \rightarrow \mathbb{R}_{\geq 0}$ such that

$$P^{(n)}(\mathbf{S}_1, \dots, \mathbf{S}_n) = P^{(2)}(\mathbf{S}_1, \mathbf{S}_2) \prod_{j=3}^n \Psi(\mathbf{S}_{j-2}, \mathbf{S}_{j-1}, \mathbf{S}_j) \quad (21)$$

for all $n \geq 3$ and all $(\mathbf{S}_1, \dots, \mathbf{S}_n) \in \mathcal{B}_n$.

We in fact also prove a refined version of this theorem, where the initial position $\mathbf{Q}_0$ is not random but fixed (on the microscopic scale); see [38, Theorem 1.1]. Furthermore, the limiting distributions $P^{(n)}$ and Ψ are *independent* of $\mathcal{L}$ and Λ .

We define the probability measure corresponding to (21) by

$$\mu_\Lambda^{(n)}(\mathcal{A}) := \int_{\mathcal{A}} P^{(n)}(\mathbf{S}_1, \dots, \mathbf{S}_n) \Lambda'(\mathbf{Q}_0, \widehat{\mathbf{S}}_1) d\mathbf{Q}_0 d\mathbf{S}_1 \cdots d\mathbf{S}_n. \quad (22)$$

Note in particular that $\mu_\Lambda^{(n+1)}(\mathcal{A} \times \mathbb{R}^d) = \mu_\Lambda^{(n)}(\mathcal{A})$.

5. A limiting random flight process

In Theorem 2 we have identified a Markov process with memory two that describes the limiting distribution of billiard paths with random initial data $(\mathbf{Q}_0, \mathbf{V}_0)$. Let us denote by

$$\{\Xi(t) : t \in \mathbb{R}_{>0}\}, \quad (23)$$

the continuous-time stochastic process that is obtained by moving with unit speed along the random paths $\mathbf{S}_1, \mathbf{S}_2, \dots$ of the above Markov process with memory two. This random flight process is fully specified by the probability

$$\mathbb{P}_\Lambda(\Xi(t_1) \in \mathcal{D}_1, \dots, \Xi(t_M) \in \mathcal{D}_M) \quad (24)$$

that $\Xi(t)$ visits the sets $\mathcal{D}_1, \dots, \mathcal{D}_M \subset \mathbb{T}^1(\mathbb{R}^d)$ at times $t = t_1, \dots, t_M$, with M arbitrarily large. To give a precise definition of (23) set $T_0 := 0$, $T_n := \sum_{j=1}^n \|\mathbf{S}_j\|$, and define the probability that $\Xi(t)$ is in the set $\mathcal{D}_1$ at time t_1 after exactly n_1 hits, in the set $\mathcal{D}_2$ at time t_2 after exactly n_2 hits, etc., by

$$\begin{aligned} & \mathbb{P}_\Lambda^{(\mathbf{n})}(\Xi(t_1) \in \mathcal{D}_1, \dots, \Xi(t_M) \in \mathcal{D}_M, T_{n_1} \leq t_1 < T_{n_1+1}, \dots, T_{n_M} \leq t_M < T_{n_M+1}) \\ & := \mu_\Lambda^{(n+1)}(\{(\mathbf{S}_1, \dots, \mathbf{S}_{n+1}) : \Xi_{n_j}(t_j) \in \mathcal{D}_j, T_{n_j} \leq t_j < T_{n_j+1} \ (j = 1, \dots, M)\}) \end{aligned} \quad (25)$$

with $\mathbf{n} := (n_1, \dots, n_M)$, $n := \max(n_1, \dots, n_M)$, and

$$\Xi_n(t) := \left(\sum_{j=1}^n \mathbf{S}_j + (t - T_n) \hat{\mathbf{S}}_{n+1}, \hat{\mathbf{S}}_{n+1} \right). \quad (26)$$

The formal definition of (23) is thus

$$\begin{aligned} & \mathbb{P}_\Lambda(\Xi(t_1) \in \mathcal{D}_1, \dots, \Xi(t_M) \in \mathcal{D}_M) \\ & := \sum_{\mathbf{n} \in \mathbb{Z}_{\geq 0}^M} \mathbb{P}_\Lambda^{(\mathbf{n})}(\Xi(t_1) \in \mathcal{D}_1, \dots, \Xi(t_M) \in \mathcal{D}_M \\ & \quad \text{and } T_{n_1} \leq t_1 < T_{n_1+1}, \dots, T_{n_M} \leq t_M < T_{n_M+1}). \end{aligned} \quad (27)$$

The following theorem shows that the Lorentz process (2) converges to the stochastic process (23) as $\rho \rightarrow 0$.

Theorem 3. *Fix a lattice $\mathcal{L}$ and let Λ be a Borel probability measure on $\mathbb{T}^1(\mathbb{R}^d)$ which is absolutely continuous with respect to Lebesgue measure. Then, for any $t_1, \dots, t_M \in \mathbb{R}_{\geq 0}$, and any subsets $\mathcal{D}_1, \dots, \mathcal{D}_M \subset \mathbb{T}^1(\mathbb{R}^d)$ with boundary of Lebesgue measure zero,*

$$\begin{aligned} & \lim_{\rho \rightarrow 0} \Lambda(\{(\mathbf{Q}_0, \mathbf{V}_0) \in \mathbb{T}^1(\rho^{d-1}\mathcal{K}_\rho) : \\ & \quad (\mathbf{Q}(t_1), \mathbf{V}(t_1)) \in \mathcal{D}_1, \dots, (\mathbf{Q}(t_M), \mathbf{V}(t_M)) \in \mathcal{D}_M\}) \\ & = \mathbb{P}_\Lambda(\Xi(t_1) \in \mathcal{D}_1, \dots, \Xi(t_M) \in \mathcal{D}_M). \end{aligned} \quad (28)$$

The convergence is uniform for $t_1, \dots, t_M$ in compact subsets of $\mathbb{R}_{\geq 0}$.

Theorem 3 follows from Theorem 2; the main ingredient in the proof is an estimate that shows that it is unlikely to have many collisions in any fixed time interval [38, Section 5].

The generalized linear Boltzmann equation (9) can now be interpreted as the Fokker-Planck-Kolmogorov equation of the stochastic process (23), and its validity follows from Theorem 3 by standard arguments from the theory of stochastic processes [38, Section 6].

6. The distribution of free path lengths

To give a detailed account of the proof of Theorem 2 would go beyond the scope of this lecture. I however hope to be able to explain the key idea, which is already clearly visible in the case $n = 1$, i.e., the distribution of free path lengths.

In order to explain the proof, it will be more convenient to return to the original microscopic coordinates $(\mathbf{q}, \mathbf{v})$, where the free path length diverges at the rate $\rho^{-(d-1)}$.

Let us denote by $\mathcal{K}_\rho \subset \mathbb{R}^d$ the complement of the union of all scatterers. The free path length for the initial condition $(\mathbf{q}, \mathbf{v}) \in \mathbb{T}^1(\mathcal{K}_\rho)$ is defined as

$$\tau_1(\mathbf{q}, \mathbf{v}; \rho) = \inf\{t > 0 : \mathbf{q} + t\mathbf{v} \notin \mathcal{K}_\rho\}. \quad (29)$$

That is, $\tau_1(\mathbf{q}, \mathbf{v}; \rho)$ is the first time at which a particle with initial data $(\mathbf{q}, \mathbf{v})$ hits a scatterer. We also include the (somewhat artificial) case when $\mathbf{q} \in \mathcal{L}$, i.e., when the particle starts at the center of a scatterer; in this case we think of $\mathcal{K}_\rho \subset \mathbb{R}^d$ as the domain obtained by removing all scatterers *except* the one centered at $\mathbf{q}$. The following is the main result of [37].

Theorem 4. *Fix a lattice $\mathcal{L}$ of covolume one, let $\mathbf{q} \in \mathbb{R}^d$, and let λ be a Borel probability measure on S_1^{d-1} absolutely continuous with respect to Lebesgue measure. Then there exists a C^1 function $F_{\mathcal{L}, \mathbf{q}}$ on $\mathbb{R}_{\geq 0}$ such that, for every $\xi > 0$,*

$$\lim_{\rho \rightarrow 0} \lambda(\{\mathbf{v} \in S_1^{d-1} : \rho^{d-1} \tau_1(\mathbf{q}, \mathbf{v}; \rho) \leq \xi\}) = F_{\mathcal{L}, \mathbf{q}}(\xi). \quad (30)$$

The distribution of the free path lengths in the Lorentz gas was already investigated by Polya, who rephrased the problem in terms of the visibility in a (random and periodic) forest [43]. The problem of the limiting distribution in dimension $d = 2$ was recently solved by Boca and Zaharescu [5] in the case when $\mathbf{q}$ is either random or located at a lattice point; see also their earlier work with Gologan [4], and the paper by Calglioti and Golse [10]. Previous work in higher dimension $d \geq 3$ includes the papers by Bourgain, Golse and Wennberg [7], [24] who provide upper and lower bounds on the tail of the distribution of free path lengths. More details on the existing literature can be found in the survey [25].

7. The space of lattices

The new idea in the joint work with Strömbergsson [37] is to translate the problem of the free path length into a question about the dynamics on the space of lattices. The advantage of this approach over previous attempts is that the technique extends naturally to arbitrary dimension, and that the limiting distributions have a canonical interpretation as the distribution function of random variables on a beautiful geometric object.

A Euclidean lattice $\mathcal{L} \subset \mathbb{R}^d$ of covolume one can be written as $\mathcal{L} = \mathbb{Z}^d M$ for some $M \in \mathrm{SL}(d, \mathbb{R})$, where $\mathbb{Z}^d$ is the standard cubic lattice. Since $\mathrm{SL}(d, \mathbb{Z})$ leaves $\mathbb{Z}^d$ invariant, the homogeneous space $X_1 = \mathrm{SL}(d, \mathbb{Z}) \backslash \mathrm{SL}(d, \mathbb{R})$ parametrizes the

space of lattices of covolume one. Similarly, let $\text{ASL}(d, \mathbb{R}) = \text{SL}(d, \mathbb{R}) \ltimes \mathbb{R}^d$ be the semidirect product group with multiplication law

$$(M, \xi)(M', \xi') = (MM', \xi M' + \xi'). \quad (31)$$

An action of $\text{ASL}(d, \mathbb{R})$ on $\mathbb{R}^d$ can be defined as

$$\mathbf{y} \mapsto \mathbf{y}(M, \xi) := \mathbf{y}M + \xi. \quad (32)$$

Each *affine* lattice (i.e. translate of a lattice) of covolume one in $\mathbb{R}^d$ can then be expressed as $\mathbb{Z}^d g$ for some $g \in \text{ASL}(d, \mathbb{R})$, and the space of affine lattices is then represented by $X = \text{ASL}(d, \mathbb{Z}) \backslash \text{ASL}(d, \mathbb{R})$ where $\text{ASL}(d, \mathbb{Z}) = \text{SL}(d, \mathbb{Z}) \ltimes \mathbb{Z}^d$. We denote by μ_1 and μ the Haar measure on $\text{SL}(d, \mathbb{R})$ and $\text{ASL}(d, \mathbb{R})$, respectively, normalized in such a way that they represent probability measures on X_1 and X .

We are interested in the λ -measure of velocities with free path lengths at most $\rho^{-(d-1)}\xi$,

$$\lambda(\{\mathbf{v} \in S_1^{d-1} : \rho^{d-1}\tau_1(\mathbf{q}, \mathbf{v}; \rho) \leq \xi\}). \quad (33)$$

This is approximately the same as the λ -measure of directions such that a cylinder $\mathbf{q} + \mathcal{Z}(\mathbf{v}, \rho^{-(d-1)}\xi, \rho)$ of length $\rho^{-(d-1)}\xi$ and radius ρ , pointing in direction $\mathbf{v}$ contains at least one lattice point (cf. Fig. 4):

$$\approx \lambda(\{\mathbf{v} \in S_1^{d-1} : \mathbb{Z}^d M \cap \mathbf{q} + \mathcal{Z}(\mathbf{v}, \rho^{-(d-1)}\xi, \rho) \neq \emptyset\}). \quad (34)$$

The approximation comes from the fact that our cylinder should have spherical caps of radius ρ on each end; it is easy to show however that the λ -measure of $\mathbf{v}$ that have a lattice point in these caps is vanishingly small, as $\rho \rightarrow 0$ (see [37, Section 4.1] for details). The next step is to shift lattice and cylinder by $-\mathbf{q}$, and then rotate by $K(\mathbf{v}) \in \text{SO}(d)$ such that $K(\mathbf{v})\mathbf{v} = \mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0, \dots, 0)$:

$$= \lambda(\{\mathbf{v} \in S_1^{d-1} : (\mathbb{Z}^d M - \mathbf{q})K(\mathbf{v}) \cap \mathcal{Z}(\mathbf{e}_1, \rho^{-(d-1)}\xi, \rho) \neq \emptyset\}). \quad (35)$$

The cylinder now lies in the $\mathbf{e}_1$ -direction. We apply the linear transformation $D(\rho) = \text{diag}(\rho^{d-1}, \rho^{-1}, \dots, \rho^{-1})$ which transforms the long and thin cylinder into better proportions:

$$= \lambda(\{\mathbf{v} \in S_1^{d-1} : (\mathbb{Z}^d M - \mathbf{q})K(\mathbf{v})D(\rho) \cap \mathcal{Z}(\xi) \neq \emptyset\}), \quad (36)$$

where

$$\mathcal{Z}(\xi) := \mathcal{Z}(\mathbf{e}_1, \xi, 1) = \{(x_1, \dots, x_d) \in \mathbb{R}^d : 0 < x_1 < \xi, \|(x_2, \dots, x_d)\| < 1\}. \quad (37)$$

Although the above linear transformations seem trivial, we have achieved a different perspective on the problem: Rather than counting lattice points in a long thin cylinder (which looks hard) we now count in a well proportioned object. The price we have paid is that our original lattice $\mathcal{L} = \mathbb{Z}^d M$ has changed to $(\mathbb{Z}^d M - \mathbf{q})K(\mathbf{v})D(\rho)$. Hence we are moving through the space of (affine) lattices, as $\rho \rightarrow 0$, and may now employ ergodic theoretic methods to understand the averages over $\mathbf{v}$ with respect to λ .

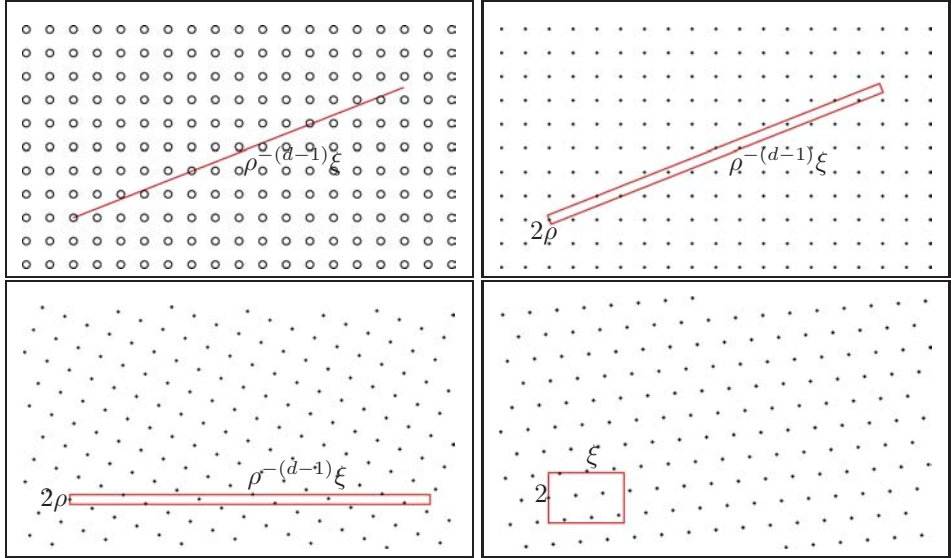


Fig. 4. Renormalization of the periodic Lorentz gas. *Step 1*: Replace the question on the number of intersections of a line with scatterers by a lattice point counting problem in a long stretched cylinder. *Step 2*: Rotate cylinder and lattice so the cylinder lies horizontally. *Step 3*: Apply a diagonal linear transformation that maps the cylinder to the ρ -independent reference cylinder $\mathcal{Z}(\xi)$.

8. Equidistribution in the space of lattices

We begin with the simplest case, $\mathbf{q} \in \mathcal{L}$, i.e., without loss of generality $\mathbf{q} = \mathbf{0}$. The right translation

$$X_1 \rightarrow X_1, \quad \mathrm{SL}(d, \mathbb{Z})M \mapsto \mathrm{SL}(d, \mathbb{Z})M\Phi^t \quad (38)$$

by the element

$$\Phi^t = \begin{pmatrix} e^{-(d-1)t} & \mathbf{0} \\ \mathbf{0} & e^t \mathbf{1}_{d-1} \end{pmatrix} \quad (39)$$

defines a flow on the homogeneous space $X_1 = \mathrm{SL}(d, \mathbb{Z}) \backslash \mathrm{SL}(d, \mathbb{R})$, where ${}^t\mathbf{x}$ denotes the transpose of the row vector $\mathbf{x}$. This flow has many good chaotic features: it is ergodic, mixing and partially hyperbolic.

The horospherical subgroups generated by

$$n_+(\mathbf{x}) = \begin{pmatrix} 1_{d-1} & {}^t\mathbf{0} \\ \mathbf{x} & 1 \end{pmatrix}, \quad n_-(\mathbf{x}) = \begin{pmatrix} 1_{d-1} & {}^t\mathbf{x} \\ \mathbf{0} & 1 \end{pmatrix}, \quad (40)$$

respectively, generate the stable and unstable horospherical subgroups of the flow. Using the mixing property, a standard argument (see e.g. [21], [33]) shows that averages over the unstable horosphere become asymptotically ($t \rightarrow \infty$) equidistributed in X_1 with respect to μ_1 .

Theorem 5. *Let λ be a Borel probability measure on $\mathbb{R}^{d-1}$ absolutely continuous with respect to Lebesgue measure. Let $f : X_1 \rightarrow \mathbb{R}$ be bounded continuous, $M \in \mathrm{SL}(d, \mathbb{R})$. Then*

$$\lim_{t \rightarrow \infty} \int_{\mathbb{R}^{d-1}} f(Mn_-(\mathbf{x})\Phi^t) d\lambda(\mathbf{v}) = \int_{X_1} f(M') d\mu_1(M'). \quad (41)$$

Alternative proofs of this theorem can be obtained by using harmonic analysis (which is particularly feasible when $M = 1$, i.e., the horosphere is closed) or Ratner's theory, which we will revisit below.

In view of (36) we are interested in the distribution of the orbit

$$\mathrm{SL}(d, \mathbb{Z}) \backslash \{ \mathrm{SL}(d, \mathbb{Z})MK(\mathbf{v})\Phi^t : \mathbf{v} \in S_1^{d-1} \} \quad (42)$$

in X_1 , as $t \rightarrow \infty$ (set $t = \log 1/\rho$). By using the fact that this orbit is close to an unstable horosphere, we can exploit Theorem 5 to deduce equidistribution also in this case; we refer the reader to [37, Section 5] for details.

Theorem 6. *Let λ be a Borel probability measure on S_1^{d-1} absolutely continuous with respect to Lebesgue measure. Let $f : X_1 \rightarrow \mathbb{R}$ be bounded continuous, $M \in \mathrm{SL}(d, \mathbb{R})$. Then*

$$\lim_{t \rightarrow \infty} \int_{S_1^{d-1}} f(MK(\mathbf{v})\Phi^t) d\lambda(\mathbf{v}) = \int_{X_1} f(M') d\mu(M'). \quad (43)$$

This theorem thus states, that in the limit $t \rightarrow \infty$ we can replace the λ -average over $\mathbf{v}$ by an average of the entire space of lattices. This yields (modulo some technicalities) the proof of the limit law for the free path length, Theorem 4, in the case $\mathbf{q} \in \mathcal{L}$, plus a formula for the limit distribution:

$$F_{\mathcal{L}, \mathbf{o}}(\xi) = \mu_1(\{M \in X_1 : \mathbb{Z}^d M \cap \mathcal{Z}(\xi) \neq \emptyset\}). \quad (44)$$

That is, *the limit distribution of the free path length for a particle emerging from the center of a scatterer is given by the probability that a random lattice intersects the cylinder $\mathcal{Z}(\xi)$ in at least one point*. Note that the limit distribution $F_{\mathbf{o}}(\xi) := F_{\mathcal{L}, \mathbf{o}}(\xi)$ is independent of the lattice $\mathcal{L}$ and of λ .

Instead of particles emerging from a lattice point we can also consider initial conditions on the boundary of a scatterer; this leads to different limit distributions and is one of the crucial steps in the proof of the existence of the limiting random flight process described in Theorems 2 and 3.

Let us now turn to the case $\mathbf{q} \notin \mathbb{Q}\mathcal{L}$. The right translation

$$X \rightarrow X, \quad \mathrm{ASL}(d, \mathbb{Z})g \mapsto \mathrm{ASL}(d, \mathbb{Z})g\Phi^t \quad (45)$$

by the element

$$\Phi^t = \left(\begin{pmatrix} e^{-(d-1)t} & \mathbf{0} \\ \mathbf{0} & e^t \mathbf{1}_{d-1} \end{pmatrix}, \mathbf{0} \right) \quad (46)$$

now defines a flow on the homogeneous space $X = \mathrm{ASL}(d, \mathbb{Z}) \backslash \mathrm{ASL}(d, \mathbb{R})$. In analogy with the above, we set

$$n_-(\mathbf{x}) = \left(\begin{pmatrix} 1_{d-1} & \mathbf{x} \\ \mathbf{0} & 1 \end{pmatrix}, \mathbf{0} \right). \quad (47)$$

The crucial difference is now that $n_-(\mathbf{x})$ no longer generates the full unstable horosphere for the flow Φ^t , and hence the mixing argument is no longer sufficient. Instead, we need to employ Ratner's classification of measures that are invariant under unipotent actions, which in the present case is given by the right action of $n_-(\mathbf{x})$. We can in particular exploit a very useful theorem of Shah [46] to show the following. (See [37, Section 5] for details, and [36] for a general introduction to applications of Ratner's theory to problems of this kind.)

Theorem 7. *Let λ be a Borel probability measure on $\mathbb{R}^{d-1}$ absolutely continuous with respect to Lebesgue measure. Let $f : X \rightarrow \mathbb{R}$ be bounded continuous, $\alpha \in \mathbb{R}^d \setminus \mathbb{Q}^d$, $M \in \mathrm{SL}(d, \mathbb{R})$. Then*

$$\lim_{t \rightarrow \infty} \int_{\mathbb{R}^{d-1}} f((1, \alpha)(M, \mathbf{0})n_-(\mathbf{x})\Phi^t) d\lambda(\mathbf{x}) = \int_X f(g) d\mu(g). \quad (48)$$

Using the same approximation argument that led to Theorem 6 we deduce:

Theorem 8. *Let λ be a Borel probability measure on S_1^{d-1} absolutely continuous with respect to Lebesgue measure. Let $f : X \rightarrow \mathbb{R}$ be bounded continuous, $\alpha \in \mathbb{R}^d \setminus \mathbb{Q}^d$, $M \in \mathrm{SL}(d, \mathbb{R})$. Then*

$$\lim_{t \rightarrow \infty} \int_{S_1^{d-1}} f((1, \alpha)(M, \mathbf{0})(K(\mathbf{v}), \mathbf{0})\Phi^t) d\lambda(\mathbf{v}) = \int_X f(g) d\mu(g). \quad (49)$$

This proves Theorem 4 in the case $\mathbf{q} \notin \mathbb{Q}\mathcal{L}$, and yields the formula

$$F_{\mathcal{L}, \mathbf{q}}(\xi) = \mu(\{(M, \mathbf{x}) \in X : (\mathbb{Z}^d M + \mathbf{x}) \cap \mathcal{Z}(\xi) \neq \emptyset\}). \quad (50)$$

Hence, the limit distribution of the free path length for a particle starting at a generic position is given by the probability that a random affine lattice intersects the cylinder $\mathcal{Z}(\xi)$ in at least one point. Here, $F(\xi) := F_{\mathcal{L}, \mathbf{q}}(\xi)$ is evidently independent of $\mathcal{L}$, $\mathbf{q}$ and λ .

I omit the discussion of the remaining $\mathbf{q} \in \mathbb{Q}\mathcal{L}$ —the arguments are analogous to the above [37].

Only in dimension $d = 2$ have we been able to turn the formulas (44) and (50) for $F_0(\xi)$ and $F(\xi)$ into explicit functions of ξ . The easiest approach is to integrate our formula for the transition probability (11) (see [39] for details),

$$F_0(\xi) = \int_0^\xi \int_{-1}^1 \Phi_0(\xi', 0, z) dz d\xi', \quad (51)$$

$$F(\xi) = \int_0^\xi \int_{\xi'}^\infty \int_{-1}^1 \int_{-1}^1 \Phi_0(\xi'', w, z) dw dz d\xi'' d\xi'. \quad (52)$$

The resulting explicit expressions coincide with those obtained, using different methods, by Boca, Gologan and Zaharescu [4], and by Boca and Zaharescu [5], respectively.

9. Asymptotics

The geometry of the space of lattices is significantly more complicated in dimension $d \geq 3$, and it seems extremely hard to obtain any explicit formulas. It is possible, however, to describe the asymptotic tails of our limit distributions, by observing that when $\xi > 0$ is very large (or small), then the lattices that contribute to $F_0(\xi)$ and $F(\xi)$ must have at least one very short basis vector. That is, all integration is restricted to the cusps of the spaces X_1 and X , respectively, whose geometry is simpler. Using tools from the geometry of numbers we can show that [40]

$$F_0(\xi) = 1 \quad \text{for } \xi \text{ sufficiently large,} \quad (53)$$

and

$$F_0(\xi) = \frac{\nu_d}{\zeta(d)} \xi + O(\xi^2), \quad \xi \rightarrow 0, \quad (54)$$

where $\nu_d = \frac{\pi^{(d-1)/2}}{\Gamma((d+1)/2)}$ is the volume of the $(d-1)$ -dimensional unit ball. Similarly,

$$F(\xi) = 1 - \frac{\pi^{\frac{d-1}{2}}}{2^d d \Gamma(\frac{d+3}{2}) \zeta(d)} \xi^{-1} + O(\xi^{-1-\frac{2}{d}}), \quad \xi \rightarrow \infty, \quad (55)$$

and

$$F(\xi) = \nu_d \xi + O(\xi^2), \quad \xi \rightarrow 0. \quad (56)$$

We also obtain asymptotic formulas for the collision kernel $\Phi_0(\xi, w, z)$ in the limits $\xi \rightarrow 0$ and $\xi \rightarrow \infty$, see [40] for details.

10. Outlook

The techniques outlined above do not necessarily require that the scatterers are rigid spheres. It is sufficient to assume that the scattering map is dispersive; a Muffin-Tin Coulomb potential would be a good example [38]. A key hypothesis of our approach is however is that the interaction region of the scattering process is finite, so that the test particle moves along straight lines for most of the time. This assumption is no longer valid in the case of long-range potentials. Provided the renormalization approach can be modified accordingly, it seems feasible to generalize our studies to crystals with long-range potential, at least for sufficiently fast power-like decay. A result by Desvillettes and Pulvirenti [15] achieves this objective (with some additional technical assumptions) in the case of random scatterer configurations; cf. also the work by Poupaud and Vasseur [44].

A related problem is to consider different scaling limits for compact potentials, where the strength of the potential is reduced, and at the same time the scatterer

density rescaled suitably to achieve a non-trivial limit. In this case grazing collisions become important, and one expects a different kinetic equation for the macroscopic dynamics; cf. Desvillettes and Ricci [16] for the corresponding result for a random scatterer configuration—here the limiting kinetic equation is the classical Fokker-Planck equation.

The renormalization approach we have developed for the Lorentz gas assumes that the scattering process is the same for each scatterer. It is possible to remove this assumption, as long as the resulting (global) potential of the crystal is still periodic, or in the case of quasicrystals. A natural question leading on from this is whether one can describe the kinetic equations in the case of a (quasi-)crystal with random defects, i.e., a scatterer is removed from a lattice site with probability p , with $0 < p < 1$. Such a set-up will lead to an interesting variation of the renormalization method, since the modular invariance of the crystal will be replaced by a modular invariance in distribution (we assume p is the same for each scatterer). It is no surprise that the limiting case $p \rightarrow 1$ leads back to the linear Boltzmann equation [45].

A further important case that has been extensively studied for stochastic Lorentz gases is the dynamics in the presence of electro-magnetic fields [17]. In the case of constant magnetic or electric fields, the above problem has some connection with beautiful, basic number-theoretic questions on the distribution of lattice points near circles or parabolas, respectively.

Finally, an obvious challenge is to adapt our renormalization approach to the quantum mechanical problem, and derive a corresponding quantum kinetic equation, in suitable weak-coupling or low-density limits discussed for random scatterer configurations by Erdős and Yau [20], and Eng and Erdős [19]. It should be stressed that of course the quantum theory of electrons in *fixed* periodic potentials is well understood, and that the lattice symmetry allows for a wealth of techniques in the spectral analysis (Floquet-Bloch decomposition). In particular the KKR method is an extremely useful tool, and has been applied to the semiclassical analysis of the symmetry-reduced periodic Lorentz gas (the Sinai billiard) in connection with quantum chaos, see Berry [2], and (for the case of small scatterers) Dahlqvist and Vattay [14].

The central idea of the work presented here is to exploit the dynamics of flows on the space of lattices. This technique has proved to be extremely powerful also in other applications in mathematical physics, including KAM theory [27], [28], arithmetic quantum unique ergodicity [30] and the Berry-Tabor conjecture [22], [32], [34], [35]. Flows on the space of lattices (and more general homogeneous spaces) should be viewed as *the* higher-dimensional generalization of the classical continued fraction algorithms—I expect more striking applications in the future.

Acknowledgements

All of the results described above have been obtained in collaboration with Andreas Strömbergsson. It has been a great pleasure to work with him on this project. I would also like to thank EPSRC, the Leverhulme Trust, the Royal Society and the Wolfson Foundation for their generous support.

References

- [1] P. Balint and I.P. Toth, Exponential decay of correlations in multi-dimensional dispersing billiards. *Ann. Henri Poincaré* 9 (2008) 1309–1369.
- [2] M.V. Berry, Quantizing a classically ergodic system: Sinai's billiard and the KKR method. *Ann. Physics* 131 (1981) 163–216.
- [3] P.M. Bleher, Statistical properties of two-dimensional periodic Lorentz gas with infinite horizon. *J. Statist. Phys.* 66 (1992), 315–373.
- [4] F.P. Boca, R.N. Gologan and A. Zaharescu, The statistics of the trajectory of a certain billiard in a flat two-torus. *Comm. Math. Phys.* 240 (2003), 53–73.
- [5] F.P. Boca and A. Zaharescu, The distribution of the free path lengths in the periodic two-dimensional Lorentz gas in the small-scatterer limit, *Commun. Math. Phys.* 269 (2007), 425–471.
- [6] C. Boldrighini, L.A. Bunimovich and Y.G. Sinai, On the Boltzmann equation for the Lorentz gas. *J. Statist. Phys.* 32 (1983), 477–501.
- [7] J. Bourgain, F. Golse and B. Wennberg, On the distribution of free path lengths for the periodic Lorentz gas. *Comm. Math. Phys.* 190 (1998), 491–508.
- [8] L.A. Bunimovich and Y.G. Sinai, Statistical properties of Lorentz gas with periodic configuration of scatterers. *Comm. Math. Phys.* 78 (1980/81), 479–497.
- [9] V.A. Bykovskii and A.V. Ustinov, Trajectory statistics in inhomogeneous Sinai problem for 2-dimensional lattice, *Izv. Ran. Ser. Mat.*, 73:4 (2009), 17–36
- [10] E. Caglioti and F. Golse, On the distribution of free path lengths for the periodic Lorentz gas. III. *Comm. Math. Phys.* 236 (2003), 199–221.
- [11] E. Caglioti and F. Golse, The Boltzmann-Grad limit of the periodic Lorentz gas in two space dimensions, *C. R. Math. Acad. Sci. Paris* 346 (2008) 477–482.
- [12] N.I. Chernov, Statistical properties of the periodic Lorentz gas. Multidimensional case. *J. Statist. Phys.* 74 (1994), 11–53.
- [13] P. Dahlqvist, The Lyapunov exponent in the Sinai billiard in the small scatterer limit. *Nonlinearity* 10 (1997), 159–173.
- [14] P. Dahlqvist and G. Vattay, Periodic orbit quantization of the Sinai billiard in the small scatterer limit. *J. Phys. A* 31 (1998) 6333–6345.
- [15] L. Desvillettes and M. Pulvirenti, The linear Boltzmann equation for long-range forces: a derivation from particle systems. *Math. Models Methods Appl. Sci.* 9 (1999) 1123–1145.
- [16] L. Desvillettes and V. Ricci, A rigorous derivation of a linear kinetic equation of Fokker-Planck type in the limit of grazing collisions. *J. Statist. Phys.* 104 (2001) 1173–1189.
- [17] L. Desvillettes and V. Ricci, Non-Markovianity of the Boltzmann-Grad limit of a system of random obstacles in a given force field. *Bull. Sci. Math.* 128 (2004) 39–46.
- [18] D. Dolgopyat, D. Szasz and T. Varju, Recurrence properties of the planar Lorentz process, *Duke Math. J.* 142 (2008) 241–281.
- [19] D. Eng and L. Erdős, The linear Boltzmann equation as the low density limit of a random Schrödinger equation. *Rev. Math. Phys.* 17 (2005) 669–743.

- [20] L. Erdős and H.-T. Yau, Linear Boltzmann equation as the weak coupling limit of a random Schrödinger equation. *Comm. Pure Appl. Math.* 53 (2000) 667–735.
- [21] A. Eskin and C. McMullen, Mixing, counting, and equidistribution in Lie groups, *Duke Math. J.* 71 (1993), 181–209.
- [22] A. Eskin, G. Margulis and S. Mozes, Quadratic forms of signature $(2, 2)$ and eigenvalue spacings on rectangular 2-tori. *Ann. of Math.* (2) 161 (2005) 679–725.
- [23] G. Gallavotti, Divergences and approach to equilibrium in the Lorentz and the Wind-tree-models, *Physical Review* 185 (1969), 308–322.
- [24] F. Golse and B. Wennberg, On the distribution of free path lengths for the periodic Lorentz gas. II. *M2AN Math. Model. Numer. Anal.* 34 (2000) 1151–1163.
- [25] F. Golse, The periodic Lorentz gas in the Boltzmann-Grad limit. *Proc. ICM 2006*, Madrid, Spain, 183–201.
- [26] F. Golse, On the periodic Lorentz gas and the Lorentz kinetic equation, *Ann. Fac. Sci. Toulouse Math.* (6) 17 (2008) 735–749.
- [27] K. Khanin, J. Lopes Dias and J. Marklof, Renormalization of multidimensional Hamiltonian flows. *Nonlinearity* 19 (2006) 2727–2753.
- [28] K. Khanin, J. Lopes Dias and J. Marklof, Multidimensional continued fractions, dynamical renormalization and KAM theory. *Comm. Math. Phys.* 270 (2007) 197–231.
- [29] O.E. Lanford III, Time evolution of large classical systems. *Dynamical systems, theory and applications* (Recontres, Battelle Res. Inst., Seattle, Wash., 1974), pp. 1–111. *Lecture Notes in Phys.*, Vol. 38, Springer, Berlin, 1975.
- [30] E. Lindenstrauss, Invariant measures and arithmetic quantum unique ergodicity. *Ann. of Math.* (2) 163 (2006) 165–219.
- [31] H. Lorentz, Le mouvement des électrons dans les métaux, *Arch. Néerl.* 10 (1905), 336–371.
- [32] J. Marklof, Spectral form factors of rectangle billiards. *Comm. Math. Phys.* 199 (1998) 169–202.
- [33] J. Marklof, The n -point correlations between values of a linear form, *Ergod. Th. & Dynam. Sys.* 20 (2000), 1127–1172.
- [34] J. Marklof, Pair correlation densities of inhomogeneous quadratic forms, *Ann. of Math.* 158 (2003) 419–471.
- [35] J. Marklof, Pair correlation densities of inhomogeneous quadratic forms II, *Duke Math. J.* 115 (2002) 409–434, correction *ibid.* 120 (2003) 227–228.
- [36] J. Marklof, Distribution modulo one and Ratner’s theorem. *Equidistribution in number theory, an introduction*, 217–244, *NATO Sci. Ser. II Math. Phys. Chem.*, 237, Springer, Dordrecht, 2007.
- [37] J. Marklof and A. Strömbergsson, The distribution of free path lengths in the periodic Lorentz gas and related lattice point problems, arXiv:0706.4395; to appear in the *Annals of Mathematics*.
- [38] J. Marklof and A. Strömbergsson, The Boltzmann-Grad limit of the periodic Lorentz gas, arXiv:0801.0612
- [39] J. Marklof and A. Strömbergsson, Kinetic transport in the two-dimensional periodic Lorentz gas. *Nonlinearity* 21 (2008) 1413–1422.
- [40] J. Marklof and A. Strömbergsson, The periodic Lorentz gas in the Boltzmann-Grad limit: Asymptotic estimates. In preparation.
- [41] I. Melbourne and M. Nicol, Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys.* 260 (2005), 131–146.
- [42] I. Melbourne and M. Nicol, A vector-valued almost sure invariance principle for hyperbolic dynamical systems, *Ann. Probab.* 37 (2009), 478–505.
- [43] G. Polya, Zahlentheoretisches und Wahrscheinlichkeitstheoretisches über die

Sichtweite im Walde, Arch. Math. Phys. 27 (1918), 135–142.

- [44] F. Poupaud and A. Vasseur, Classical and quantum transport in random media. J. Math. Pures Appl. (9) 82 (2003) 711–748.
- [45] V. Ricci and B. Wennberg, On the derivation of a linear Boltzmann equation from a periodic lattice gas. Stochastic Process. Appl. 111 (2004) 281–315.
- [46] N.A. Shah, Limit distributions of expanding translates of certain orbits on homogeneous spaces. Proc. Indian Acad. Sci. Math. Sci. 106 (1996) 105–125.
- [47] H. Spohn, The Lorentz process converges to a random flight process, Comm. Math. Phys. 60 (1978), 277–290.
- [48] D. Szasz and T. Varju, Limit laws and recurrence for the planar Lorentz process with infinite horizon, J. Stat. Phys. 129 (2007) 59–80.

ISING MODELS, UNIVERSALITY AND THE NON RENORMALIZATION OF THE QUANTUM ANOMALIES

VIERI MASTROPIETRO

Università' di Roma Tor Vergata, Roma, Italy

E-mail: mastropi@mat.uniroma2.it

A number of universal relations (proposed by Kadanoff, Luther, Peschel and Haldane) are believed to be true in a wide class of systems with continuously varying indices, among which are interacting planar Ising models, vertex or Ashkin-Teller models, quantum spin chains and 1D Fermi systems; by such relations one can predict several quantities in terms of a few measurable parameters without relying on the specific microscopic details. The validity of such relations can be checked in special solvable models but, despite several attempts, the proof of their general validity was up to now an open problem. A rigorous derivation of several of such relations (for solvable and not solvable models and without any use of exact solutions) has been recently obtained in [8] and [11] through Renormalization Group methods. The proof is based on the representation in terms of Grassmann integrals and the validity of the Adler-Bardeen property of the non renormalization of the quantum anomalies in the asymptotic *Ward identities*. Gauge invariance is exact only in the scaling limit but the lattice corrections can be rigorously taken into account.

Keywords: Universality, critical exponents, Renormalization Group

1. Introduction

The principle of *Universality* in statistical physics is central in the modern understanding of critical phenomena. Universality means that the critical properties close to phase transitions are *insensitive* to the microscopic details inside a certain *universality class* of systems. For models in the class of universality of the Ising model, whose exponents are pure numbers, this principle says that the exponents are *exactly the same*. Experimentally this happens in systems like carbon dioxide [22]; even if they are described by very complex Hamiltonians, depending on a number of microscopic details, their critical exponents *coincide* with the 3D Ising model ones. Despite widely accepted, rigorous results on universality in the Ising classes have only been proven in dimension 4 and above.

There are however systems in which the indices are not pure numbers but *depend on all the microscopic details*; this happens in a wide class of models, including planar Ising-like models with quartic interactions, vertex or Ashkin-Teller models, quantum spin chains and 1D fermionic systems. It has long been conjectured, mainly by Kadanoff [17–19], Luther and Peschel [29] and Haldane [21], that a number of

universal relations between critical exponents and other observables are valid in this class of models. Such relations express how the *universality principle* works in such systems: the critical exponents are model dependent (non-universal) but satisfy model independent formulas, so allowing, for instance, *to express all the exponents in terms of a single one*. The universal relations have been verified only in certain special exactly solvable models, but the conjecture is that they are generally valid in a larger class of models, for which an exact solution is not available.

The interest in these universal relations has been renewed by recent experiments on materials described by models in this class, like quantum spin chain models (KCuF3) [28], carbon nanotubes [3] or even 1D Bose systems [26]. In such systems the critical exponents depend on the extraordinary complex and largely unknown microscopic details of the compounds, but the universal relations allow concrete and testable predictions for them in terms of a few measurable parameters.

A mathematical proof of such universal relations has shown to be a rather challenging problem. Several attempts in the last thirty years have been devoted to their proof [36], [35], [41], [38], using a variety of methods ranging from operator product expansions, perturbation theory, Renormalization Group, bosonization and several others. It is common to all such approaches to start from a formal continuum limit in which extra Lorentz and Gauge symmetries are verified. However strictly speaking the formal continuum limit is singular, as it is plagued by ultraviolet divergences which were absent in the original lattice model. Moreover lattice effects destroy such symmetries and change the exponents, and it is not clear at all while the relations between exponents should be true also when such symmetries are violated.

On the other hand, not all the relations which are valid in the special solvable models are generically true; a counterexample is found for the exponents involved in the dynamic correlations [26] and another one will be shown below. It is therefore important to know rigorously, and therefore unambiguously, under which conditions and which one of the relations valid in the solvable models are generically true.

The proof of several of such universal relations has been achieved in [8, 11], and in this paper we will outline such results. Up to recent times, the exponents were known only for a very small number of models where an exact solution was known. New methods have been introduced in [39] and [31] to study 2D statistical mechanics models, which can be considered as a perturbation of the Ising model. These methods take advantage of the fact that such systems can be mapped in systems of interacting fermions in $d = 1 + 1$ dimensions. This mapping was known since a long time [23, 40], but only in recent years a great progress has been achieved in the evaluation of the Grassmann integrals involved in the analysis of the interacting models (see *e.g.* [33] for a review). The outcome is that the exponents and other physical observables can be written as *convergent series*; this means that they can be computed with arbitrary precision with an explicit computation of lowest orders and with a rigorous bound on the rest. The complexity of the expansions make however essentially impossible to prove the universal relations directly from them. It is

then shown that such models have the same exponents to the ones of an *effective continuum fermionic model*, provided that the bare parameters of such model are chosen properly. The effective model is expressed in terms of Grassmann integrals which are identical to the ones appearing in certain Quantum Field Theory models; we take advantage of the Gauge symmetry and of a property called *anomaly non-renormalization* to get exact expressions of the critical indices and other quantities in terms of the bare parameters; by them, several of the universal relations can be proved.

2. Ising, Vertex and Ashkin-Teller models

The paradigmatic model for statistical mechanics is the 2D Ising model with Hamiltonian

$$H = -J \sum_{j=0,1} \sum_{\mathbf{x} \in \Lambda} \sigma_{\mathbf{x}} \sigma_{\mathbf{x}+\mathbf{e}_j} \quad (1)$$

$\sigma_{\mathbf{x}} = \pm 1$, Λ is a square subset of $\mathbb{Z}^2$, $\mathbf{x} = (x_0, x_1) \in \Lambda$, $\mathbf{e}_0 = (0, 1)$, $\mathbf{e}_1 = (1, 0)$. By the exact solution, due to Onsager, the critical exponents can be computed and they appear to be J -independent. The universality hypothesis says that a small next to nearest neighbor or a quartic short ranged perturbation *do not change* the exponents, and for a set of exponents this has been recently proved in [39].

The simplest model displaying indices which are non trivial functions of the parameters is obtained considering *two* Ising models coupled by a quartic interaction; the Hamiltonian is

$$H(\sigma, \sigma') = H_J(\sigma) + H_{J'}(\sigma') - \lambda V(\sigma, \sigma') \quad (2)$$

with V is a short ranged, quartic interaction in the spin and invariant in the spin exchange, like

$$V = \sum_{j=0,1} \sum_{\mathbf{x}, \mathbf{y} \in \Lambda} v(\mathbf{x} - \mathbf{y}) \sigma_{\mathbf{x}} \sigma_{\mathbf{x}+\mathbf{e}_j} \sigma'_{\mathbf{y}} \sigma'_{\mathbf{y}+\mathbf{e}_j} \quad (3)$$

with $v(\mathbf{x})$ a short range potential.

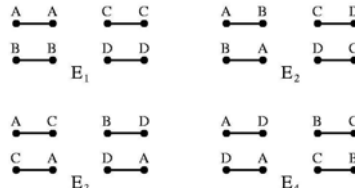


Fig. 1. Interactions between neighbor spins in the Ashkin-Teller models

The interest in the above model is increased by the fact that several systems in statistical mechanics, like the *Ashkin-Teller* and the *Eight Vertex* models, can be rewritten as coupled Ising models.

The Ashkin-Teller [2] model is a generalization of the Ising model in which the spin has four values A, B, C, D , and to two neighbor spins is associated an energy E_1 for AA, BB, CC, DD , E_2 for AB, CD , E_3 for AC, BD , E_4 for AD, BC .

The Eight vertex model [5] is a generalization of the Ice model for the hydrogen bounding in which at each point is associated one among eight vertices with four different energies.



Fig. 2. The vertices in the Eight Vertex model

Both models can be rewritten in the form of coupled Ising models with Hamiltonian (2), see [5]. In the case of the Ashkin-Teller model this is straightforward as it is sufficient to associate to each lattice point $\mathbf{x}$ a couple of spins $(\sigma_{\mathbf{x}}, \sigma'_{\mathbf{x}})$ and to associate the four couples $(\pm 1, \pm 1)$ with the four states A, B, C, D . One can then immediately verify that the Ashkin-Teller hamiltonian is equivalent to (2), up to an additive constant $-J_0$, provided that we choose

$$\begin{aligned} -J &= (E_1 + E_2 - E_3 - E_4)/4 & -J' &= (E_1 + E_3 - E_4 - E_2)/4 \\ -\lambda &= (E_1 + E_4 - E_2 - E_3)/4 & -J_0 &= (E_1 + E_2 + E_3 + E_4)/4 \end{aligned} \quad (4)$$

and

$$V = \sum_{j=0,1} \sum_{\mathbf{x} \in \Lambda} \sigma_{\mathbf{x}} \sigma_{\mathbf{x}+\mathbf{e}_j} \sigma'_{\mathbf{x}} \sigma'_{\mathbf{x}+\mathbf{e}_j} \quad (5)$$

For a choice of parameters such that $J = J'$ the Ashkin-Teller model is called *isotropic*, while for $J \neq J'$ is called *anisotropic*. When $\lambda = 0$ the model is exactly solvable as its hamiltonian is the sum of two independent Ising models, and two *critical temperatures* are present if $J \neq J'$ which reduce to one in the $J = J'$ case.

Analogously also the Eight vertex model can be mapped in (2) with a suitable identification of the parameters; in such a case $J = J'$ and

$$V = \sum_{j=0,1} \sum_{\mathbf{x} \in \Lambda} \sigma_{\mathbf{x}+j(\mathbf{e}_0+\mathbf{e}_1)} \sigma_{\mathbf{x}+\mathbf{e}_0} \sigma'_{\mathbf{x}+j(\mathbf{e}_0+\mathbf{e}_1)} \sigma'_{\mathbf{x}+\mathbf{e}_1} \quad (6)$$

Despite the similarity between the Eight Vertex and the Ashkin-Teller model, quite apparent from (5),(6), an exact solution is known *only* in the case of the Eight Vertex model, due to Baxter ([4]), and, even in that case, only a few indices can be computed. This shows how much the solvability is a delicate property depending on all details of the Hamiltonian; even if the two interactions appear equivalent, as

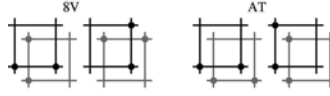


Fig. 3. Graphical representation of the interactions (5) and (6).

far as the long distance properties are considered, only for a one of them an exact solution is known. By the exact solution of the Eight vertex model ([4]) one sees that the exponents are non trivial functions of λ ; it is *not* in the Ising universality class.

3. Quantum spin chain and 1D Fermi systems

A model very related to the previous ones is the Heisenberg spin chain, physically realized in several compounds like KCuF_3 . Its Hamiltonian is given by

$$H = - \sum_{x=1}^{L-1} [J_1 S_x^1 S_{x+1}^1 + J_2 S_x^2 S_{x+1}^2 - h S_x^3] + \lambda \sum_{1 \leq x, y \leq L} v(x-y) S_x^3 S_y^3 \quad (7)$$

where $S_x^\alpha = \sigma_x^\alpha/2$ for $i = 1, 2, \dots, L$ and $\alpha = 1, 2, 3$, σ_x^α being the Pauli matrices and $|v(x-y)| \leq C e^{-\kappa|x-y|}$. In the special case ($\lambda = -J_3$)

$$v(x-y) = \delta_{|x-y|,1}/2, \quad h = 0 \quad (8)$$

the model is known as *XYZ chain* and it is exactly solvable. Despite it looks very different, it is related to the previous models: the Hamiltonian of the *XYZ* model *commutes* with the transfer matrix of the 8V model ([5]). For a generic short range interaction, the solvability is lost.

The spin chain (7) can be exactly mapped in a system of interacting fermions through the *Jordan-Wigner* transformation. It is indeed well known that the operators

$$a_x^\pm = \prod_{y=1}^{x-1} (-\sigma_y^3) \sigma_x^\pm \quad (9)$$

are a set of anticommuting fermionic operators and that, if $\sigma_x^\pm = (\sigma_x^1 \pm i\sigma_x^2)/2$, we can write

$$\sigma_x^- = e^{-i\pi \sum_{y=1}^{x-1} a_y^+ a_y^-} a_x^+ a_x^-, \quad \sigma_x^+ = a_x^+ e^{i\pi \sum_{y=1}^{x-1} a_y^+ a_y^-}, \quad \sigma_x^3 = 2a_x^+ a_x^- - 1. \quad (10)$$

Hence, if we fix the units so that $J_1 + J_2 = 2$ we get

$$\begin{aligned} H = & -\frac{1}{2} \sum_{x=1}^{L-1} [a_x^+ a_{x+1}^- + a_{x+1}^+ a_x^-] - u \sum_{x=1}^{L-1} [a_x^+ a_{x+1}^+ + a_{x+1}^- a_x^-] \\ & + h \sum_{x=1}^L (a_x^+ a_x^- - \frac{1}{2}) + \lambda \sum_{1 \leq x, y \leq L} v(x-y) (a_x^+ a_x^- - \frac{1}{2}) (a_y^+ a_y^- - \frac{1}{2}) \end{aligned} \quad (11)$$

where $\rho_x = a_x^+ a_x^-$, $u = (J_1 - J_2)/2$. In this form, the model describes interacting non relativistic 1D fermions on a lattice with a short range interaction and a BCS-like term (in the anisotropic case $J_1 \neq J_2$), and it can be used to understand the properties of the conduction electrons of one-dimensional metals.

It is convenient to introduce the density and the current operators:

$$\begin{aligned}\rho_x &= S_x^3 + \frac{1}{2} = a_x^+ a_x^- , \\ J_x &= S_x^1 S_{x+1}^2 - S_x^2 S_{x+1}^1 = \frac{1}{2i} [a_{x+1}^+ a_x^- - a_x^+ a_{x+1}^-] .\end{aligned}\quad (12)$$

If $\rho_{\mathbf{x}} = e^{x_0 H} \rho_x e^{-x_0 H}$, $a_{\mathbf{x}}^\pm = e^{x_0 H} a_x^\pm e^{-x_0 H}$ the above definition of the current is justified by the (imaginary time) conservation equation

$$\frac{\partial \rho_{\mathbf{x}}}{\partial x_0} = e^{H x_0} [H, \rho_x] e^{-H x_0} = -i \partial_x^{(1)} J_{\mathbf{x}} = -i [J_{x, x_0} - J_{x-1, x_0}] \quad (13)$$

where we have used that ρ_x commutes with the quartic part of H . Note that J_x does not verify a simple equation like (13), as J_x does not commute with the quartic part of H . Note also that

$$[H_0, \hat{J}_p] = \frac{1}{L} \sum_k \sin k (\cos(k+p) - \cos k) \hat{a}_{k+p}^+ \hat{a}_k \quad (14)$$

a relation which will be useful in the following.

4. Universal relations

It has been conjectured that the above models are in the same *universality class*; this does not mean that the exponents are the same (on the contrary, the indices depend on all the details of the Hamiltonian), rather it means that (for instance) there are *universal relations* between them, such that *all the indices of a single model can be expressed in terms of any one of them*. As an example of such relations, in the coupled Ising model the following relations have been conjectured

$$X_- = \frac{1}{X_+} \quad \nu = \frac{1}{2 - X_+} \quad \alpha = \frac{2 - 2X_+}{2 - X_+} = 2 - 2\nu \quad (15)$$

where $X_\pm$ are the exponents of the energy or crossover correlations, ν is the exponents of the correlation length, α the exponent of the specific heat (see below for their exact definition).

The first of above relations was proposed by Kadanoff [17], the second by Kadanoff and Wegner [19] and the third is the hyperscaling relation. In the case of the Eight vertex model, the index ν and α can be computed from the exact solution, and one can check the validity of the last of (15); the indices $X_\pm$ cannot be computed from the solution even in the Eight vertex case. In the spin chains or 1D fermions, Luther and Peschel [29] proposed similar relations, with a different identification of the exponents. Except from a partial verification in the Eight or XYZ models, there were no proof (until now) of the validity of (15) in non solvable models.

In general even the knowledge of a single exponent can be lacking; in the case of spin chains or 1D fermions, Haldane [21] conjectured other relations *allowing the determination of the exponents in terms of thermodynamic quantities (Luttinger liquid conjecture)*. In particular if v_s is the Fermi velocity and κ is the susceptibility, calling $v_N = (\pi\kappa)^{-1}$ the following universal relation was conjectured in [21]

$$\frac{v_s}{v_N} = X_+ \quad (16)$$

The validity of this relation can be checked in the XYZ case; the correlation length exponent index $\bar{\nu}$ is, if $\cos \bar{\mu} = -J_3/J_1$

$$\bar{\nu} = \frac{\pi}{2\bar{\mu}} \quad (17)$$

from Baxter solution [4]. Moreover $X_- = 2(1 - \frac{\bar{\mu}}{\pi})$ from the relation $\bar{\nu} = \frac{1}{2-X_-}$ (conjectured in [29]) and from Bethe ansatz

$$v_s = \frac{\pi}{\bar{\mu}} \sin \bar{\mu} \quad \kappa = \frac{1}{[2\pi(\pi/\bar{\mu} - 1) \sin \bar{\mu}]} \quad (18)$$

from which (16) follows. *The conjecture is that this relation is true generically*, for instance in the model (7) with $h \neq 0$ and any short range interaction $v(x-y)$.

5. Renormalization Group analysis for coupled Ising models

It happens sometimes that problems which appear rather hard in certain variables appear simpler in others. This is the case of the planar Ising model: it is a system of *interacting* spins which can be mapped in a system of *non interacting* fermions. Such a mapping is particularly useful because models which are perturbation of the Ising model, like the Eight Vertex or Ashkin-Teller models, can be mapped in models of interacting fermions, which can be analyzed via the methods developed in Quantum Field Theory.

The starting point is the following representation of the Ising model partition function:

$$Z_I = (\cosh \beta J)^B 2^S \sum_{\gamma} (\tanh \beta J)^{|\gamma|} . \quad (19)$$

where B is the total number of bonds, S is the total number of sites and the sum is over all the multipolygon γ with length $|\gamma|$. If open boundary conditions are assumed, only multipolygons *not* winding up the lattice are allowed. In the case of periodic boundary conditions the representation is the same, but the polygons are allowed to wind up the lattice. It was proved in [20], [37], [23] that Z_I can be written as the sum of four *Pfaffians* which can be written as four *Grassmann integrals* with different boundary conditions

$$Z_I = -Z_{+,+} + Z_{+,-} + Z_{-,+} + Z_{-,-} \quad (20)$$

$$Z_{\epsilon,\epsilon'} = (\cosh \beta J)^B 2^S \frac{1}{2} \int \prod_{\mathbf{x} \in \Lambda} dH_{\mathbf{x}} d\bar{H}_{\mathbf{x}} dV_{\mathbf{x}} d\bar{V}_{\mathbf{x}} e^{S_{\epsilon,\epsilon'}} \quad (21)$$

where

$$S_{\epsilon, \epsilon'} = \sum_{\mathbf{x} \in \Lambda} \tanh \beta J [\bar{H}_{x, x_0} H_{x+1, x_0} + \bar{V}_{x, x_0} V_{x, x_0+1}] + \sum_{\mathbf{x} \in \Lambda} [\bar{H}_{x, x_0} H_{x, x_0} + \bar{V}_{x, x_0} V_{x, x_0} + \bar{V}_{x, x_0} \bar{H}_{x, x_0} + V_{x, x_0} \bar{H}_{x, x_0} + H_{x, x_0} \bar{V}_{x, x_0} + V_{x, x_0} H_{x, x_0}] \quad (22)$$

and $H_{\mathbf{x}}, \bar{H}_{\mathbf{x}}, V_{\mathbf{x}}, \bar{V}_{\mathbf{x}}$ are anticommuting variables such that

$$\begin{aligned} \bar{H}_{x, x_0+L} &= \epsilon \bar{H}_{x, x_0} & \bar{H}_{x+L, x_0} &= \epsilon' \bar{H}_{x, x_0} \\ H_{x, x_0+L} &= \epsilon H_{x, x_0} & H_{x+L, x_0} &= \epsilon' H_{x, x_0} \end{aligned} \quad (23)$$

and identical relations hold for the variables $V, \bar{V}$. Note that periodic or antiperiodic boundary conditions are imposed in the x_0, x variables depending on $\epsilon, \epsilon' = \pm$. The Grassmann integral is a linear operation such that $\int d\eta_{\mathbf{x}} = 0$, $\int d\eta_{\mathbf{x}} \eta_{\mathbf{x}} = 1$ if $\eta_{\mathbf{x}}$ is any of $(H_{\mathbf{x}}, \bar{H}_{\mathbf{x}}, V_{\mathbf{x}}, \bar{V}_{\mathbf{x}})$.

Let us consider now the coupled Ising model (2); we will be interested in particular in the specific heat C_v and the energy $\epsilon = +$ and cross-over ($\epsilon = -$) correlations, defined as

$$G_{\beta}^{\epsilon}(\mathbf{x} - \mathbf{y}) = \lim_{\Lambda \rightarrow \infty} \langle O_{\mathbf{x}}^{\epsilon} O_{\mathbf{y}}^{\epsilon} \rangle_{\Lambda} - \langle O_{\mathbf{x}}^{\epsilon} \rangle_{\Lambda} \langle O_{\mathbf{y}}^{\epsilon} \rangle_{\Lambda} \quad \epsilon = \pm \quad (24)$$

where

$$O_{\mathbf{x}}^{\epsilon} = \sum_{j=0,1} \sigma_{\mathbf{x}} \sigma_{\mathbf{x}+\mathbf{e}_j} + \epsilon \sum_{j=0,1} \sigma'_{\mathbf{x}} \sigma'_{\mathbf{x}+\mathbf{e}_j} \quad (25)$$

where $\langle \cdot \rangle_{\Lambda}$ is the average over all configurations of the spins with statistical weight $e^{-\beta H}$, H given by (2). Starting from (21) and performing suitable change of variables, such correlations can be written as sums of functional derivatives (with respect to A^{ϵ} , $\epsilon = +$ for the energy and $\epsilon = -$ for the crossover) of Grassmann integrals with different boundary conditions; in the thermodynamic limit it is sufficient to consider only one of them, for instance

$$Z(A) = \int P_{Z_1, \mu_1}(d\psi) e^{L^2 \mathcal{N} + \mathcal{V}^{(1)}(\sqrt{Z_1} \psi) + \mathcal{B}^{(1)}(\sqrt{Z_1} \psi, A)}, \quad (26)$$

where $\mathcal{N}$ is a constant, $\psi_{\mathbf{k}, \omega}^{\pm}$ is a finite set of Grassman variables, $\mathbf{k} = (k_0, k)$, $k_0 = \frac{2\pi}{L}(n_0 + \frac{1}{2})$, $k = \frac{2\pi}{L}(n + \frac{1}{2})$, $n_0, n_1 = -L/2, \dots, L/2 - 1$, $P_{Z_1, \mu_1}(d\psi)$ is a gaussian Grassman integration with propagator, in the case $J = J'$

$$g(\mathbf{k}) = \frac{1}{Z_1} \begin{pmatrix} -i \sin k_0 + \sin k + \mu_{++}(\mathbf{k}) & -\mu_1 - \mu_{-+}(\mathbf{k}) \\ -\mu_1 - \mu_{+-}(\mathbf{k}) & -i \sin k_0 - \sin k_1 + \mu_{++}(\mathbf{k}) \end{pmatrix}^{-1} \quad (27)$$

with $\mu_1 = O(|t - t_c|)$, $t = \tanh \beta J$, $t_c = \tanh \beta_c J = \sqrt{2} - 1 - \nu$, $Z_1 = O(1)$; finally $\mu_{\epsilon, \epsilon} = O(\mathbf{k}^3)$ and $\mu_{\epsilon, -\epsilon} = O(\mathbf{k}^2)$ are such that the determinant vanishes only when $\mu_1 = 0$ and $\mathbf{k} = (0, 0)$. Moreover (27) is given by

$$\mathcal{V}^{(1)}(\psi) = \nu \sum_{\mathbf{x}, \omega=\pm} \psi_{\mathbf{x}, \omega}^+ \psi_{\mathbf{x}, \omega}^- + \lambda_1 \sum_{\mathbf{x}} \psi_{\mathbf{x}, +}^+ \psi_{\mathbf{x}, +}^- \psi_{\mathbf{x}, -}^+ \psi_{\mathbf{x}, -}^- + R_1(\psi) \quad (28)$$

$$\mathcal{B}^{(1)}(\psi, A) = \sum A_{\epsilon, \mathbf{x}} O_{\mathbf{x}}^{\epsilon} + R_2(A, \psi) \quad (29)$$

with $\nu_1 = O(\nu)$, $\lambda_1 = O(\lambda)$; R_1 is a sum of monomials in ψ more than quartic in ψ or quartic with at least a derivatives and R_2 is a sum of monomials in A, ψ more than quadratic in ψ or quadratic with at least a derivative; finally

$$O_{\mathbf{x}}^{+} = \psi_{\mathbf{x},+}^{+} \psi_{\mathbf{x},-}^{-} + \psi_{\mathbf{x},-}^{-} \psi_{\mathbf{x},+}^{+} \quad O_{\mathbf{x}}^{-} = i[\psi_{\mathbf{x},+}^{+} \psi_{\mathbf{x},-}^{+} + \psi_{\mathbf{x},+}^{-} \psi_{\mathbf{x},-}^{-}] \quad (30)$$

Note that ν is a *counterterm* to be chosen so that β_c is the critical temperature.

Grassmann integrals appear in the analysis of the fermionic sector of Quantum Field Theory models; in particular, (26) strongly resembles the generating functions of the current correlations of a interacting 2D *Dirac fermions* with mass μ_1 and wave function renormalization Z_1 on a lattice. Zero mass $\mu_1 = 0$ corresponds to criticality. Note also that the lattice regularization of massless Dirac fermions suffer by a well known problem called *fermion doubling*, which is usually solved by the introduction of the so-called *Wilson terms*; in (26) such Wilson terms appear naturally (they are the $\mu_{\epsilon, \epsilon'}(\mathbf{k})$ terms in (27)). In the case $J \neq J'$, the correlations are expressed in terms of Dirac fermions with two different kinds of mass terms.

In the case of the Ising model, the Grassmann integrals are quadratic, and this corresponds to *free* Dirac particles on a lattice; on the contrary, coupled Ising models correspond to *interacting* Dirac particles. The fact that such models can be mapped in interacting fermions was known since a long time, but this was not really used in rigorous analysis for the lack of methods to control the perturbative expansions appearing in interacting quantum field theories. However in recent years a great progress has been achieved in the analysis of non quadratic Grassmann integrals; the key observation of Caianiello [12], saying that fermionic perturbative expansions have much better convergence properties due to the relative signs, has been substantiated in the rigorous construction of 2D fermionic quantum field theories, either renormalizable and asymptotically free [14, 15] and superrenormalizable [30]; the key tool of such analysis are the Gram bound for determinants and the Battle-Brydges-Federbush formula for truncated expectations. The Grassmann integral in (26) is much more difficult to analyze, as the theory is renormalizable but *not* asymptotically free; it belongs to a class of models with *vanishing* beta function (see below for details). Such kinds of models can be constructed only exploiting non trivial cancellations in the expansions. The first example of rigorous construction of a model in this class was in [7], relying on the *exact solution* of the Luttinger model [34]. Later on, models with vanishing Beta function were constructed using a technique capable of combining Ward Identities based on local symmetries with Renormalization Group methods which was developed in [10] (and without any use of exact solutions); the main problem to face is that the momentum cut-off breaks local symmetries producing additional terms in the Ward Identities which can be however rigorously taken into account. By using such methods, the following Theorem can be proved.

Theorem 5.1. (*Mastropietro [31]*) *The coupled Ising model (2) with $J = J'$ and*

λ small enough is critical at $\tanh \beta_c J = \sqrt{2} - 1 + O(\lambda)$ and the specific heat is

$$C_v \sim -\frac{1}{\alpha}[1 - |\beta - \beta_c|^{-\alpha}] \quad (31)$$

with $\alpha = O(\lambda)$. If $\beta \neq \beta_c$ the density and crossover correlations $G_\beta^\epsilon(\mathbf{x} - \mathbf{y})$, $\epsilon = \pm$ (24),(25) decay faster than any power of $\xi^{-1}|\mathbf{x} - \mathbf{y}|$, with $\xi^{-1} \sim C|\beta - \beta_c|^\nu$ with $\nu = 1 + O(\lambda)$ and C is a constant. Moreover

$$G_{\beta_c}^\epsilon(\mathbf{x} - \mathbf{y}) \sim \frac{C_\epsilon}{|\mathbf{x} - \mathbf{y}|^{2X_\epsilon}}, \text{ as } |\mathbf{x} - \mathbf{y}| \rightarrow \infty, \quad (32)$$

with $X_\pm = 1 + O(\lambda)$ and $C_\pm$ constants.

From the above theorem we see that the interaction has two main effects. The first one is to change the value of the critical temperature. The second and more dramatic one is to modify qualitatively the critical properties. The logarithmic singularity in the specific heat of the Ising model is changed in a power law singularity when $\lambda > 0$; on the contrary for $\lambda < 0$ the specific heat is a continuous function. Moreover, the exponents $X_\pm, \nu$, which were equal to 1, are now continuous non trivial functions of λ .

The above result gives the first proof of the fact that the critical exponents are non trivial functions of the interaction in coupled Ising models with a generic quartic interaction (2); previously this was known only in the Eight Vertex case thanks to the exact solution. The series for X_+, X_-, ν, X_T are *convergent* for small λ , and the indices can be computed with arbitrary precision by an explicit computation of the first orders. In the case of a single perturbed Ising model, it was proved by Pinson and Spencer [39] that the indices $\nu = 1, X_\pm = 1$, that is they are the same as the Ising ones.

A similar Renormalization Group analysis can be performed in the anisotropic Ashkin-Teller model when $J \neq J'$.

Theorem 5.2. (Giuliani, Mastropietro [16]) *In the case of the anisotropic Ashkin-Teller model (2),(5) ($J \neq J'$) there are two critical temperatures, T_c^+ and T_c^- such that*

$$|T_c^- - T_c^+| \sim |J - J'|^{X_T} \quad (33)$$

with $X_T = 1 + O(\lambda)$ and

$$C_v \sim -\Delta^\alpha \log \frac{|T - T_c^-| \cdot |T - T_c^+|}{\Delta^2} \quad (34)$$

where $2\Delta^2 = (T - T_c^-)^2 + (T - T_c^+)^2$.

In this case the specific heat has the same logarithmic singularity as in the Ising model; however, even if we are in the universality class of the Ising model, the difference between the two critical temperatures rescale with an anomalous exponent in the isotropic limit $|T_{1,c} - T_{2,c}| \sim |J - J'|^{X_T}$; the existence of such a *transition index* was overlooked in the physical literature.

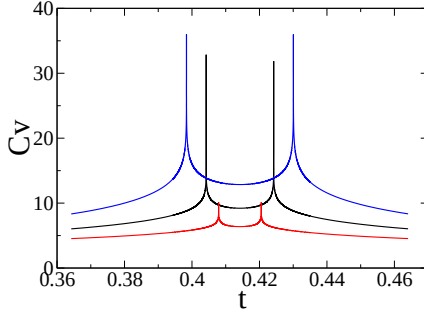


Fig. 4. The specific heat (34), in the case $\lambda < 0$, $\lambda = 0$ and $\lambda > 0$.

Let us sketch the proof of Theorem 5.1 for the case $J = J'$. The Grassmann variables are written as $\psi_{\mathbf{k}} = \sum_{h=-\infty}^0 \psi_{\mathbf{k}}^{(h)}$, and after the integration of the fields $\psi^{(0)}, \dots, \psi^{(h+1)}$, the partition function can be written as

$$Z(A) = e^{S^{(h)}(A)} \int P_{\bar{Z}_{h-1}, \mu_{h-1}}(d\psi^{(\leq h)}) e^{\mathcal{V}^{(h)}(\sqrt{\bar{Z}_{h-1}}\psi^{(\leq h)}) + \mathcal{B}^{(h)}(\sqrt{\bar{Z}_{h-1}}\psi^{(\leq h)}, A)}, \quad (35)$$

where $\psi^{(\leq h)} = \sum_{j=-\infty}^h \psi^{(j)}$ and $P_{Z_h, \mu_h}(d\psi^{(\leq h)})$ is the *Gaussian Grassmann integration* with propagator

$$g^{(\leq h)}(\mathbf{k}) = \frac{\chi_h(\mathbf{k})}{Z_h} \begin{pmatrix} -i \sin k_0 + \sin k + \mu_{++} & -\mu_h - \mu_{-+} \\ -\mu_h - \mu_{+-} & -i \sin k_0 - \sin k_1 + \mu_{--} \end{pmatrix}^{-1} \quad (36)$$

with $\chi_h(\mathbf{k})$ a smooth compact support function non vanishing for $|\mathbf{k}| \leq 2^h$. The *effective interaction* $\mathcal{V}^{(h)}(\psi)$ is a sum over monomials in the Grassmann variables

$$\mathcal{V}^{(h)}(\psi) = \gamma^h \nu_h F_\nu^{(h)} + \lambda_h F_\lambda^{(h)} + R_h, \quad (37)$$

where

$$F_\nu^{(h)} = \frac{1}{L^2} \sum_{\omega=\pm} \sum_{\mathbf{k}} \hat{\psi}_{\mathbf{k}, \omega}^{(\leq h)+} \hat{\psi}_{\mathbf{k}, -\omega}^{(\leq h)-}, \quad (38)$$

$$F_\lambda^{(\leq h)} = \frac{1}{L^8} \sum_{\mathbf{k}_1, \dots, \mathbf{k}_4} \hat{\psi}_{\mathbf{k}_1, +}^{(\leq h)+} \hat{\psi}_{\mathbf{k}_3, -}^{(\leq h)+} \hat{\psi}_{\mathbf{k}_2, +}^{(\leq h)-} \hat{\psi}_{\mathbf{k}_4, -}^{(\leq h)-} \delta(\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3 - \mathbf{k}_4).$$

and R^h contains sum of monomials with more than four fields, or quartic with at least a derivative, or bilinear with at least two derivatives. In the same way

$$\mathcal{B}^{(h)}(\sqrt{\bar{Z}_{h-1}}\psi^{(\leq h)}, A) = \sum_{\epsilon=\pm, \mathbf{x}} Z_{h-1}^{(\epsilon)} A_{\mathbf{x}}^\epsilon O_{\mathbf{x}}^{(\leq h)\epsilon} + \bar{R}_h, \quad (39)$$

where $O^\pm$ is given by (30) and $\bar{R}_h$ contains terms more than quadratic, or quadratic with a derivative. It is natural the interpretation of λ_h in (37) as the *effective coupling* of the model at momentum scales $O(2^h)$.

Notice that the propagator of the field $\psi^{(h)}$ can be written, for $h \leq 0$, as

$$g^{(h)}(\mathbf{x}, \mathbf{y}) = g_T^{(h)}(\mathbf{x}, \mathbf{y}) + r^{(\leq h)}(\mathbf{x}, \mathbf{y}) , \quad (40)$$

where

$$g_T^{(h)}(\mathbf{x}, \mathbf{y}) = \frac{1}{L^2} \sum_{\mathbf{k}} e^{-i\mathbf{k}(\mathbf{x}-\mathbf{y})} \frac{1}{Z_h} T_h^{-1}(\mathbf{k}) , \quad (41)$$

$$T_h(\mathbf{k}) = f_h(\mathbf{k}) \begin{pmatrix} -ik_0 + k & -\mu_h \\ \mu_h & -ik_0 - k \end{pmatrix} \quad (42)$$

with $f_h(\mathbf{k})$ a smooth compact support function non vanishing for $2^{h-1} \leq |\mathbf{k}| \leq 2^{h+1}$ and, for any positive integer M ,

$$|r^{(h)}(\mathbf{x}, \mathbf{y})| \leq C_M \frac{2^{2h}}{1 + (2^h |\mathbf{x} - \mathbf{y}|^M)} . \quad (43)$$

On the other hand, $g_T^{(h)}(\mathbf{x}, \mathbf{y})$ verifies a bound similar to (43) with 2^h replacing 2^{2h} .

The running couplings λ_j (which, by construction, are the same in the massless $\mu_1 = 0$ or in the massive $\mu_1 \neq 0$) satisfy a recursive equation of the form

$$\lambda_{j-1} = \lambda_j + \beta_\lambda^{(j)}(\lambda_j, \dots, \lambda_0) + \bar{\beta}_\lambda^{(j)}(\lambda_j, \nu_j; \dots; \lambda_0, \nu_0) , \quad (44)$$

where $\beta_\lambda^{(j)}$, $\bar{\beta}_\lambda^{(j)}$ are μ_1 -independent and expressed by a *convergent* expansion in $\lambda_j, \nu_j, \dots, \lambda_0, \nu_0$ (convergence follows from Gram bounds); moreover $\bar{\beta}_\lambda^{(j)}$ vanishes if at least one of the ν_k is zero. The running coupling λ_j stays close to λ_1 for any j as a consequence of the following property, called *vanishing of the Beta function*, which was proved in Theorem 2 of [9]; for suitable positive constants C and $\theta < 1$:

$$|\beta_\lambda^{(j)}(\lambda_j, \dots, \lambda_j)| \leq C |\lambda_j|^2 2^{\theta j} . \quad (45)$$

It is possible to prove that, for a suitable choice of $\nu_1 = O(\lambda)$, $\nu_j = O(2^{\theta j} \lambda)$, and this implies $\bar{\beta}_\lambda^{(j)} = O(2^{\theta j} \lambda^2)$ so that the sequence λ_j converges, as $j \rightarrow -\infty$, to a smooth function $\lambda_{-\infty}(\lambda) = \lambda_1 + O(\lambda^2)$, such that

$$|\lambda_j - \lambda_{-\infty}| \leq C \lambda^2 2^{\theta j} . \quad (46)$$

Moreover

$$\frac{Z_{j-1}}{Z_j} = 1 + \beta_z^{(j)}(\lambda_j, \dots, \lambda_0) + \bar{\beta}_z^{(j)}(\lambda_j, \nu_j; \dots, \lambda_0, \nu_0) , \quad (47)$$

with $\bar{\beta}_z^{(j)}$ vanishing if at least one of the ν_k is zero so that $\bar{\beta}_z^{(j)} = O(\lambda 2^{\theta j})$. Finally

$$\beta_z(\lambda_j, \dots, \lambda_0) = \beta_z(\lambda_{-\infty}, \dots, \lambda_{-\infty}) + O(\lambda 2^{\theta h}) , \quad (48)$$

where the last identity follows from (46) and the function $\beta_z(\lambda_{-\infty}, \dots, \lambda_{-\infty})$ is by definition sum of terms in which only the propagators $g_T^{(h)}$ (41) appear (the terms containing $r^{(j)}$ are included in the second term in (48)). Similar equations hold for $Z_h^{(\pm)}$, μ_h , with

$$\beta_\pm(\lambda_j, \dots, \lambda_0) = \beta_\pm(\lambda_{-\infty}, \dots, \lambda_{-\infty}) + O(\lambda 2^{\theta h}) . \quad (49)$$

By an explicit computation and (48), (49), there exist $\eta_+(\lambda_{-\infty}) = c_1\lambda_{-\infty} + O(\lambda_{-\infty}^2)$, $\eta_-(\lambda_{-\infty}) = -c_1\lambda_{-\infty} + O(\lambda_{-\infty}^2)$, $\eta_\mu(\lambda_{-\infty}) = c_1\lambda_{-\infty} + O(\lambda_{-\infty}^2)$ and $\eta_z(\lambda_{-\infty}) = c_2\lambda_{-\infty}^2 + O(\lambda_{-\infty}^3)$, with c_1 and c_2 strictly positive, such that, for any $j \leq 0$,

$$\begin{aligned} |\log_2(Z_{j-1}/Z_j) - \eta_z(\lambda_{-\infty})| &\leq C\lambda^2 2^{\theta j}, \\ |\log_2(\mu_{j-1}/\mu_j) - \eta_\mu(\lambda_{-\infty})| &\leq C|\lambda| 2^{\theta j}, \\ |\log_2(Z_{j-1}^{(\pm)}/Z_j^{(\pm)}) - \eta_\pm(\lambda_{-\infty})| &\leq C\lambda^2 2^{\theta j}. \end{aligned} \quad (50)$$

The critical indices in Theorem 5.1 are functions of $\lambda_{-\infty}$ only, as it is clear from (48); moreover the exponents appearing there are such that

$$X_\pm = 1 - \eta_\pm + \eta_z \quad \mu = \eta_+ - \eta_z = 1 - x_+ . \quad (51)$$

If $\mu_1 \neq 0$ (that is, if the temperature is not the critical one), the correlations decay faster than any power with rate proportional to μ_{h^*} , where, if $[x]$ denotes the largest integer $\leq x$, h^* is given by $h^* = \left\lceil \frac{\log_2 |\mu_1|}{1 + \eta_\mu} \right\rceil$.

In conclusion, the exponents are written as convergent series so that they can be computed with arbitrary precision; the complexity of the expansions makes however essentially impossible to prove the universal relations directly from the expansions and new ideas are necessary.

6. Proof of universality and anomaly non renormalization

Some of the universal relations for coupled Ising models (2) with a generic quartic interaction have been recently proved.

Theorem 6.1. (*Benfatto, Falco, Mastropietro [6]*). *Given the coupled Ising model with quartic interaction (2), with the same definitions as in Theorems 5.1 and 5.2 and λ small enough the following relations are true*

$$\begin{aligned} X_- &= \frac{1}{X_+} & \alpha &= \frac{2 - 2X_+}{2 - X_+}, \\ \nu &= \frac{1}{2 - X_+} & X_T &= \frac{1 - X_+}{1 - X_+^{-1}} \end{aligned} \quad (52)$$

Despite the exponents depend from all the details of the model, they verify universal, model independent relations allowing to express all the exponents (which we can compute) in terms of any one of them. The first two of the above relations are the ones conjectured by Kadanoff [17] and the second by Kadanoff and Wegner [19] while the last one is completely new.

In order to prove such theorem we introduce a continuum fermionic theory defined as the formal scaling limit of the original one plus an ultraviolet regularization; more exactly, we prove that the critical indices X_+ , X_- , ν and X_T of the spin model (2) are *equal* to the indices of a fermionic theory provided that the bare coupling λ_∞ of the new theory is properly chosen as a suitable function of the parameter λ in (2).

The correlations of this continuum fermionic theory are the functional derivatives of the following Grassmann integral

$$\int P_Z(d\psi^{(\leq N)}) e^{V^{(N)}(\sqrt{Z}\psi^{(\leq N)}) + \sum_{\omega=\pm} \int d\mathbf{x} [\psi_{\mathbf{x},\omega}^+ \phi_{\mathbf{x},\omega}^- + \psi_{\mathbf{x},\omega}^- \phi_{\mathbf{x},\omega}^+]} \quad (53)$$

with $\psi, \bar{\psi}$ are Euclidean $d = 1 + 1$ spinors ($\psi^\pm = (\psi_+^\pm, \psi_-^\pm)$), $P_Z(d\psi^{(\leq N)})$ is the fermionic gaussian integration with propagator $g^{(\leq N)} = \frac{\chi_N(\mathbf{k})}{\mathbf{k}}$, $\mathbf{k} = \gamma_0 k_0 + c\gamma_1 k_1$ (which in components appear to be equal to $g_T^{(\leq N)}(\mathbf{x})$ (41)), and

$$V^{(N)}(\psi^{(\leq N)}) = \lambda_\infty \int d\mathbf{x} d\mathbf{y} v(\mathbf{x} - \mathbf{y}) j_\mu(\mathbf{x}) j_\mu(\mathbf{y}) \quad (54)$$

with $j_\mu(\mathbf{x}) = \bar{\psi}_\mathbf{x} \gamma_\mu \psi_\mathbf{x}$ and $v(\mathbf{x} - \mathbf{y})$ a short range symmetric interaction; moreover

$$\gamma_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \gamma_1 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad (55)$$

A multiscale integration is now necessary also in the ultraviolet region to perform the limit $N \rightarrow \infty$; no ultraviolet divergences are present due to the non locality of the interaction in (54). The multiscale integration for the infrared scales can be done exactly as described in the previous section, with the only difference that, by the oddness of the free propagator, $\nu_j = 0$ and

$$\lambda_{j-1} = \lambda_j + \hat{\beta}_\lambda^{(j)}(\lambda_j, \dots, \lambda_0), \quad (56)$$

where

$$\hat{\beta}_\lambda^{(j)}(\lambda_j, \dots, \lambda_0) = \beta_\lambda^{(j)}(\lambda_j, \dots, \lambda_0) + O(\lambda_\infty^2 2^{\theta j}), \quad (57)$$

$\beta_\lambda^{(j)}(\lambda_j, \dots, \lambda_j)$ being the function appearing in (45), so that we can prove that $\lambda_{-\infty} = \lambda_0 + O(\lambda_0^2)$; since $\lambda_0 = \lambda_\infty + O(\lambda_\infty^2)$, we have

$$\lambda_{-\infty} = h(\lambda_\infty) = \lambda_\infty + O(\lambda_\infty^2), \quad (58)$$

for some analytic function $h(\lambda_\infty)$, invertible for λ_∞ small enough. Moreover

$$\frac{Z_{j-1}^\pm}{Z_j^\pm} = 1 + \hat{\beta}_\pm^{(j)}(\lambda_j, \dots, \lambda_0), \quad (59)$$

with

$$\hat{\beta}_\pm^{(j)}(\lambda_j, \dots, \lambda_0) = \beta_\pm^{(j)}(\lambda_j, \dots, \lambda_0) + O(\lambda_\infty^2 2^{\theta j}), \quad (60)$$

$\beta_\pm^{(j)}$ being the functions appearing in (49) (as consequence of (40)). This implies that

$$\eta_\pm = \log_2[1 + \beta_\pm^{(-\infty)}(\lambda_{-\infty}, \dots, \lambda_{-\infty})], \quad (61)$$

that is *the critical indices in the AT or 8V or in the model (53) are the same as functions of $\lambda_{-\infty}$.*

If we call $\lambda'_j(\lambda)$ the effective couplings of the lattice model (2) appearing in the previous section, the invertibility of $h(\lambda_\infty)$ implies that we can choose λ_∞ so that (chosen $c = 1$)

$$h(\lambda_\infty) = \lambda'_{-\infty}(\lambda) . \quad (62)$$

With the above choice of λ_∞ the exponents in the models (2) and (53) are the same.

What have we gained by this? The point is that the continuum fermionic theory (53) has correlations expressed by Grassmann integrals which are identical to the ones appearing in certain Quantum Field Theory models; in particular it verifies extra *Gauge symmetries* with respect to the original spin Hamiltonian. If $\langle \dots \rangle_{th}$ are the correlations with respect to $P(d\psi^{(\leq N)})e^{V^{(N)}}$ in (54), by the transformation $\psi \rightarrow e^{i\alpha \times} \psi_{\mathbf{x}}$ one finds

$$-i\mathbf{p}_\mu \langle j_{\mu,\mathbf{p}} \psi_{\mathbf{k}} \bar{\psi}_{\mathbf{k}+\mathbf{p}} \rangle_{th} = \langle \psi_{\mathbf{k}} \bar{\psi}_{\mathbf{k}} \rangle_{th} - \langle \psi_{\mathbf{k}+\mathbf{p}} \bar{\psi}_{\mathbf{k}+\mathbf{p}} \rangle_{th} + \Delta_N(\mathbf{k}, \mathbf{p}) \quad (63)$$

where

$$\Delta_N = \langle \delta j_{\mathbf{p}} \psi_{\mathbf{k}} \bar{\psi}_{\mathbf{k}+\mathbf{p}} \rangle_{th} \quad (64)$$

with

$$\delta j_{\mathbf{p}} = \int d\mathbf{k} [(\chi_N^{-1}(\mathbf{k} + \mathbf{p}) - 1)(\mathbf{k} + \mathbf{p}) - (\chi_N^{-1}(\mathbf{k}) - 1)\mathbf{k}] \bar{\psi}_{\mathbf{k}} \psi_{\mathbf{k}+\mathbf{p}} \quad (65)$$

An analogous expression is obtained for the axial current $\bar{\psi} \gamma_\mu \gamma_5 \psi$. The term Δ_N is due to the momentum regularization which is necessary to have a well defined Grassmann integral but which breaks the continuum phase symmetries. By a multiscale analysis it is found, in the limit of removed cut-off [32]

$$\lim_{N \rightarrow \infty} \Delta_N(\mathbf{k}, \mathbf{p}) = -i\tau \hat{v}(\mathbf{p}) \mathbf{p}_\mu \langle j_{\mu,\mathbf{p}} \psi_{\mathbf{k},\omega} \bar{\psi}_{\mathbf{k}+\mathbf{p},\omega} \rangle_{th} \quad \tau = \frac{\lambda_\infty}{4\pi c} \quad (66)$$

A similar expression holds for the axial Ward Identity, with j_μ replaced by j_μ^5 and τ replaced by $-\tau$. The fact that $\Delta_N(\mathbf{k}, \mathbf{p})$ is non vanishing removing the ultraviolet cut-off $N \rightarrow \infty$ is a manifestation of a *quantum anomaly*. The anomaly coefficients τ is *linear* in λ_∞ : this is the non-perturbative analogue of the *anomaly non renormalization* in 4D Quantum Electrodynamics [1]. In the proof of the validity of such property a crucial role is played by the non locality of the interaction (53); higher orders in λ_∞ are present in τ with other choices of $V^{(N)}$, for instance replacing $v(\mathbf{x} - \mathbf{y})$ with a delta function, as shown in [6].

The fact that τ has no contributions from higher orders implies that also the exponents can be exactly computed in terms of λ_∞ . By combining the Ward Identities (63), (66) with the Schwinger-Dyson equation one gets equations for the correlations from which the indices can be computed as functions of λ_∞ : it is found

$$X_+ = 1 - \frac{1}{1 + \tau} \frac{\lambda_\infty}{2\pi c} \quad X_- = 1 + \frac{1}{1 - \tau} \frac{\lambda_\infty}{2\pi c} \quad (67)$$

from which the relation $X_+ X_- = 1$ follows (the other relations between indices follow by similar arguments).

The indices have simple expressions in λ_∞ , as consequence of the linearity of τ ; all the model dependence is included in the function λ_∞ , which is given by a convergent non trivial series with coefficients depending from all the details of the spin model (2). The simple and universal expressions of the exponents in terms of λ_∞ allow to check the validity of the universal relations.

7. Renormalization Group for Quantum spin chain and 1D Fermi systems

We now come to the quantum spin chain model or 1D interacting Fermi system (11). If O_x is a local monomial in the S_x^α or $a_x^\pm$ operators, we call $O_{\mathbf{x}} = e^{Hx_0} O_x e^{-Hx_0}$ where $\mathbf{x} = (x, x_0)$; moreover, if $A = O_{\mathbf{x}_1} \cdots O_{\mathbf{x}_n}$,

$$\langle A \rangle_{L,\beta} = \frac{\text{Tr}[e^{-\beta H} \mathbf{T}(A)]}{\text{Tr}[e^{-\beta H}]} \quad (68)$$

$\mathbf{T}$ being the time order product, denotes its expectation in the grand canonical ensemble, while $\langle A \rangle_{T,L,\beta}$ denotes the corresponding truncated expectation. We will use also the notation $\langle A \rangle_T = \lim_{L,\beta \rightarrow \infty} \langle A \rangle_{T,L,\beta}$.

For λ small enough [9] if $\mathbf{x} = (x, x_0)$, x_0 imaginary time, when $J_1 = J_2$, one can prove that, for small λ , $J_1 = J_2 = 1$ and large $\mathbf{x}$,

$$\langle a_{\mathbf{x}}^- a_{\mathbf{0}}^+ \rangle_T \sim g_0(\mathbf{x}) \frac{1 + \lambda f(\lambda)}{(x_0^2 + v_s^2 x^2)^{(\eta/2)}}, \quad (69)$$

where $f(\lambda)$ is a bounded function, $\eta = a_0 \lambda^2 + O(\lambda^3)$, with $a_0 > 0$, and

$$g_0(\mathbf{x}) = \sum_{\omega=\pm} \frac{e^{i\omega p_F x}}{-ix_0 + \omega v_s x}, \quad (70)$$

$$v_s = v_F + O(\lambda), \quad p_F = \cos^{-1}(h + \lambda) + O(\lambda), \quad v_F = \sin p_F. \quad (71)$$

From (69) we see that the interaction has two main effects. The first one is to change the value of the Fermi momentum from $\cos^{-1}(h)$ to p_F and the Fermi velocity from v_F in the non interacting case to v_s . The second effect is that the power law decay is changed; the 2-point function is asymptotically given by the product of the non-interacting one (with a different velocity) times an extra power law decay factor with non-universal index η .

It was also proved in [9] that the spin-spin correlation in the direction of the 3-axis (or, equivalently, the fermionic density-density correlation) is given, for large $\mathbf{x}$, by

$$\langle S_{\mathbf{x}}^{(3)} S_{\mathbf{0}}^{(3)} \rangle_T \sim \cos(2p_F x) \Omega^{3,a}(\mathbf{x}) + \Omega^{3,b}(\mathbf{x}), \quad (72)$$

$$\Omega^{3,a}(\mathbf{x}) = \frac{1 + A_1(\mathbf{x})}{2\pi^2 [x^2 + (v_s x_0)^2]^{X_+}}, \quad (73)$$

$$\Omega^{3,b}(\mathbf{x}) = \frac{1}{2\pi^2 [x^2 + (v_s x_0)^2]} \left\{ \frac{x_0^2 - (x/v_s)^2}{x^2 + (v_s x_0)^2} + A_2(\mathbf{x}) \right\}, \quad (74)$$

with $|A_1(\mathbf{x})|, |A_2(\mathbf{x})| \leq C|\lambda|$ and $X_+ = 1 - a_1\lambda + O(\lambda^2)$ with

$$a_1 = [\hat{v}(0) - \hat{v}(2p_F)]/(\pi \sin p_F) \quad (75)$$

Finally the Cooper pair density correlation, that is the correlation of the operator $\rho_{\mathbf{x}}^c = a_{\mathbf{x}}^+ a_{\mathbf{x}'}^+ + a_{\mathbf{x}}^- a_{\mathbf{x}'}^-$, $\mathbf{x}' = (x+1, x_0)$, behaves as

$$\langle \rho_{\mathbf{x}}^c \rho_{\mathbf{0}}^c \rangle_T \sim \frac{1 + A_3(\mathbf{x})}{2\pi^2(x^2 + v_s^2 x_0^2)^{X_-}}, \quad (76)$$

with $X_- = 1 + a_1\lambda + O(\lambda^2)$, a_1 being the same constant appearing in the first order of X_+ . In the case $J_1 \neq J_2$ the correlations decay faster than any power with rate ξ such that $\xi \sim C|J_1 - J_2|^{\bar{\nu}}$ with $\bar{\nu} = 1 + a_1\lambda + O(\lambda^2)$, a_1 given by (75).

If $J_{\mathbf{x}} = v_F j_{\mathbf{x}}$ ($J_{\mathbf{x}}$ was defined in (12)), in the $\lambda = 0$ the commutation relations (13), (14) imply the following Ward Identities, for $\mathbf{k}, \mathbf{k} + \mathbf{p}$ close to $(0, \omega p_F)$, $\omega = \pm$

$$\begin{aligned} -ip_0 \langle \hat{\rho}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle + \omega p v_F \langle \hat{j}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle &\sim [\langle \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}}^- \rangle - \langle \hat{a}_{\mathbf{k}+\mathbf{p}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle] \\ -ip_0 \langle \hat{j}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle + \omega p v_F \langle \hat{\rho}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle &\sim [\langle \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}}^- \rangle - \langle \hat{a}_{\mathbf{k}+\mathbf{p}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle] \end{aligned} \quad (77)$$

In the presence of the interaction $\lambda \neq 0$, the Ward Identities have the form [11]

$$\begin{aligned} -ip_0 \langle \hat{\rho}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle + \omega p \tilde{v}_J \langle \hat{j}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle &\sim B[\langle \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}}^- \rangle - \langle \hat{a}_{\mathbf{k}+\mathbf{p}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle] \\ -ip_0 \langle \hat{j}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle + \omega p \tilde{v}_N \langle \hat{\rho}_{\mathbf{p}} \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle &\sim \bar{B}[\langle \hat{a}_{\mathbf{k}}^+ \hat{a}_{\mathbf{k}}^- \rangle - \langle \hat{a}_{\mathbf{k}+\mathbf{p}}^+ \hat{a}_{\mathbf{k}+\mathbf{p}}^- \rangle] \end{aligned} \quad (78)$$

with $B = 1$, $\bar{B} = 1 + O(\lambda)$ and $\tilde{v}_J, \tilde{v}_N = v_s(1 + O(\lambda))$; in particular $\frac{\tilde{v}_N}{\tilde{v}_J} = 1 + 2a_1\lambda + O(\lambda^2)$ with a_1 the constant defined above in (75). The interaction has the effect that the normalization $\bar{B}$ is not 1 (the current does not commute with the quartic part of the Hamiltonian) and two different velocities, the charge $\tilde{v}_J$ and the current velocity $\tilde{v}_N$, appear. The presence of the lattice, breaking the Lorentz symmetry valid in the continuum limit, causes the presence of three distinct velocities, $\tilde{v}_N, \tilde{v}_J, v_s$.

8. Universality for quantum spin chains and 1D Fermi systems

The same strategy followed for proving the universal relations in the coupled Ising models (2) allows to derive the same relations between the indices appearing in the correlations of the spin chain; again all the indices can be expressed in terms of a single one. There is in this case also an extra relation connecting the indices with the Fermi velocity v_s and the susceptibility, defined as

$$\kappa = \lim_{p \rightarrow 0} \hat{\Omega}(0, p) \quad (79)$$

where $\hat{\Omega}(0, p)$ is the bidimensional Fourier transform of $\langle S_{\mathbf{x}}^{(3)} S_{\mathbf{0}}^{(3)} \rangle_T$. In the fermionic interpretation, $\kappa \rho^{-2}$ is the compressibility (ρ is the fermionic density).

Theorem 8.1. (*Benfatto, Mastropietro [11]*) *In the model (7) for λ small enough the exponents in (69), (72), (76) verify*

$$X_+ X_- = 1 \quad \bar{\nu} = \frac{1}{2 - X_+^{-1}} \quad 2\eta = X_+ + X_+^{-1} - 2 \quad (80)$$

Moreover, the velocities in the Ward Identity (78) are such that

$$\tilde{v}_N \tilde{v}_J = v_s^2 \quad \tilde{v}_J = v_F \quad (81)$$

and the susceptibility κ verifies

$$\kappa = \frac{1}{\pi} \frac{X_+}{v_s} \quad (82)$$

The relation (82) has been proposed in [21] as a part of the *Luttinger liquid conjecture* and checked previously only in the case of the XYZ chain using the explicit exact formulas, see (17),(18).

Note that, in the notation of [21], $v_N \equiv (\pi\kappa)^{-1}$ should not be confused with $\tilde{v}_N$ appearing in the WI (77); they are coinciding only in the special case of the Luttinger model. Therefore $\tilde{v}_N = v_N$ is an example of relation true in the (exactly solvable) Luttinger model but not in the generic 1D Fermi system model (7).

The Ward Identities of the relativistic model (53) with $c = v_s$ and a suitable choice of λ_∞ and the following relations, valid for $\mathbf{k}$ $\mathbf{k}$, $\mathbf{k} + \mathbf{p}$ small, if $\mathbf{p}_F = (0, \omega p_F)$

$$\begin{aligned} < \hat{\rho}_{\mathbf{p}} \hat{a}_{\mathbf{k}+\mathbf{p}_F}^+ \hat{a}_{\mathbf{k}+\mathbf{p}+\mathbf{p}_F}^- > \sim Z^{(3)} < j_{\mathbf{p}}^0 \hat{\psi}_{\mathbf{k},\omega}^+ \hat{\psi}_{\mathbf{k}+\mathbf{p},\omega}^- >_{th} \\ < \hat{j}_{\mathbf{p}} \hat{a}_{\mathbf{k}+\mathbf{p}_F}^+ \hat{a}_{\mathbf{k}+\mathbf{p}+\mathbf{p}_F}^- > \sim \tilde{Z}^{(3)} < j_{\mathbf{p}}^1 \hat{\psi}_{\mathbf{k},\omega}^+ \hat{\psi}_{\mathbf{k}+\mathbf{p},\omega}^- >_{th} \end{aligned} \quad (83)$$

imply the Ward Identities (78) with, if τ is given by (66)

$$B = \frac{Z^{(3)}}{Z} (1 - \tau)^{-1}, \quad \bar{B} = \frac{\tilde{Z}^{(3)}}{Z} (1 + \tau)^{-1} \quad (84)$$

and

$$\tilde{v}_N = v_s \frac{Z^{(3)}}{\tilde{Z}^{(3)}} \quad \tilde{v}_J = v_s \frac{\tilde{Z}^{(3)}}{Z^{(3)}} \quad (85)$$

On the other hand, by comparing the first of (78) with the Ward Identity obtained from (13) we get

$$\frac{Z^{(3)}}{(1 - \tau)Z} = 1 \quad \tilde{v}_J = v_F \quad (86)$$

In the same way we can derive from the Ward Identity for the densities the following expressions

$$< \rho_{\mathbf{p}} \rho_{\mathbf{p}} > = \frac{1}{4\pi v_s Z^2} \frac{(Z^{(3)})^2}{1 - (\lambda_\infty/4\pi v_s)^2} \left[2 - \frac{D_-(\mathbf{p})}{D_+(\mathbf{p})} - \frac{D_+(\mathbf{p})}{D_-(\mathbf{p})} \right] + O(\mathbf{p}) \quad (87)$$

where $D_\omega(\mathbf{p}) = -ip_0 + \omega v_s p$, so that from (86) and (79)

$$\kappa = \frac{1}{\pi v_s} \frac{1}{Z^2} \frac{(Z^{(3)})^2}{1 - (\lambda_\infty/4\pi v_s)^2} = \frac{1}{\pi v_s} \frac{1 - (\lambda_\infty/4\pi v_s)}{1 + (\lambda_\infty/4\pi v_s)} = \frac{X_+}{\pi v_s} \quad (88)$$

so that (82) is proved.

9. Conclusions

We have established for the first time the validity of a number of universal relations between exponents and other quantities in a wide class models, including solvable and *not solvable* models. Several of such relations were conjectured by Kadanoff [17–19], Luther and Peschel [29] and Haldane [21]. These universal relations are used for the analysis of experiments in carbon nanotubes or spin chains, but their interest goes much beyond this, as they provide one of the very cases in which the *universality* hypothesis can be *rigorously* verified. Extensions of our methods will allow hopefully to prove universal relations in an even wider class of models and to prove other relations between exponents; for instance the ones between the exponents appearing in the spin-spin correlations in coupled Ising models (2) or the dynamic exponents in 1D Fermi systems (7).

References

- [1] Adler S.L. WBardeen Phys. Rev. **182**, 1517 - 1536, (1969).
- [2] Ashkin J., Teller E.: Phys. Rev. **64**, 178 - 184, (1943).
- [3] Auslaender O.M. et al. Phys. Rev. Lett. **84**, 1764 (2000)
- [4] Baxter R.J.: Phys. Rev. Lett. **26**, 832–833, (1971).
- [5] Baxter R.J.: *Exactly solved models in statistical mechanics*. Academic Press, Inc. London, (1989).
- [6] Benfatto G., Falco P., Mastropietro V.: Comm. Math. Phys. **273**, 67–118, (2007).
- [7] Benfatto G., Gallavotti G., Procacci A., Scoppola B.: Comm. Math. Phys. **160**, 93 (1994)
- [8] Benfatto G., Falco P., Mastropietro V.: Comm. Math. Phys., Comm. Math. Phys. **292**, 2, 569–605 (2009)
- [9] Benfatto G., Mastropietro V.: Rev. Math. Phys. **13**, 1323–1435, (2001).
- [10] Benfatto G., Mastropietro V.: Comm. Math. Phys. **258**, 609–655, (2005).
- [11] Benfatto G., Mastropietro V.: arXiv:0907.2837
- [12] E. Caianiello. Nuovo Cimento **12**, 561 (1954)
- [13] Bockrath, M et al. Nature **397**, 598–601
- [14] J.Feldman, J. Magnen, V. Rivasseau, R.Seneor Comm. Math. Phys. **103**, 1 67–103 (1986)
- [15] K. Gawedzki A. Kupiainen Comm.Math.Phys. **99**, 197–252 (1985)
- [16] Giuliani A., Mastropietro V.: Phys. Rev. Lett. **93**, 190603–07, (2004); Comm. Math. Phys. **256**, 681–725, (2005).
- [17] Kadanoff L.P.: Phys. Rev. Lett. **39**, 903–905, (1977).
- [18] Kadanoff L.P., Brown A.C.: *Ann. Phys.* **121**, 318–345, (1979).
- [19] Kadanoff L.P., Wegner: Phys. Rev. B **4**, 11, 3989–3993, (1971).
- [20] Kasteleyn P.W.: J. Math. Phys. **4**, 287, (1963).
- [21] F.D.M. Haldane. *Phys.Rev.Lett.* **45**, 1358–1362 (1980); J. Phys. C. **14**, 2575–2609 (1981).
- [22] Hocken R, Moldover MR Phys Rev Lett **37**, 29–32 (1976)
- [23] Hurst C.A., Green H.S. J. Chem. Phys. **33**, 1059, (1960).
- [24] Itzykson C., Drouffe J. Statistical field theory. Cam. Un. Press (1989)
- [25] Ishii H et al. Nature **426**, 540 2005
- [26] Imambekov A, Glazman L.: Science **323**, 228 (2009); Phys.Rev.Lett. **102**, 126405 (2009)

- [27] Yao Z et al, Nature 401, 273 (1999)
- [28] Lake B et al. Nature materials 4, 329, 2005
- [29] Luther A., Peschel I.: Phys. Rev.B **12**, 3908–3917, (1975).
- [30] A. Lesniewski: Comm. Math. Phys. **108**, 437–467, (1987).
- [31] Mastropietro V.: Comm. Math. Phys. **244** 595–64 (2004).
- [32] Mastropietro V.: J. Math. Phys **48**, 022302, (2007).
- [33] Mastropietro V.: *Non-perturbative Renormalization*. World Scientific, (2008).
- [34] D. Mattis, E. Lieb. *J. Math. Phys.* **6**, 304–312 (1965).
- [35] den Nijs M.P.M.: Phys. Rev. B **23**, 6111–6125, (1981).
- [36] Pruisken A.M.M. Brown A.C.: Phys. Rev. B **23**, 1459–1468, (1981); Pruisken A.M.M., Kadanoff L.P.: Phys. Rev. B **22**,1 5154–5170, (1980).
- [37] Samuel S. J.Math. Phys. **21**, 2806, (1980).
- [38] H. Spohn, Phys. Rev. E **60**, 6411 (1999).
- [39] Spencer T. Physica A **279**, 250–259, (2000); Pinson H., Spencer T.: *Unpublished*
- [40] Schultz T.D., Mattis D.C., Lieb E.H.: Rev. Mod. Phys. **36**, 856 - 871 (1964)
- [41] Zamolodchikov A.B., Zamolodchikov Al. B.: Soviet Scientific Reviews A **10**, 269, (1989).

THE INFRARED PROBLEM IN NONRELATIVISTIC QED

ALESSANDRO PIZZO

Department of Mathematics, UC Davis,

Davis, CA 95616, USA

E-mail: pizzo@math.ucdavis.edu

www.math.ucdavis.edu

The fact that photons are massless particles introduces substantial difficulties into the mathematical analysis of the interactions between nonrelativistic quantum matter and the quantized radiation field. These difficulties are known as the infrared problem in (nonrelativistic) QED. This issue is of particular interest in atomic physics. After a review of the different aspects of the infrared problem, we report on recent progress in tackling some of the open mathematical questions. We describe some novel spectral and scattering techniques based on multi-scale analysis.

1. Introduction

In the last ten to fifteen years, we have seen a renewed interest in nonrelativistic Quantum Electrodynamics (QED), that is the mathematical theory describing the interaction of charged quantum matter and the quantized radiation field at low energy scales, where pair-creation and other relativistic phenomena are negligible. The renewed interest in this subject is based on: (a) The rigorous treatment of the *infrared divergences* that plague the calculation of the radiative corrections to some important physical quantities; this topic is the main subject of these notes and will be analyzed in some detail; (b) Questions concerning the stability of matter, and the interaction with strong electromagnetic fields; (c) The development of quantum optics applications that show phenomena whose theoretical explanation requires nonperturbative arguments and, in general, a more refined control of the underlying mathematical theory.

In textbooks, the *infrared problem* in QED is usually presented as a pathology of the scattering states, linked to the infamous *infrared catastrophe*. It became clear very early in the development of QED that, at the level of perturbation theory, the transition amplitudes (e.g., for Compton scattering) between formal scattering states with charges and a *finite* number of photons are *ill-defined*, because, typically, Feynman amplitudes containing vertex or electron self-energy corrections exhibit logarithmic infrared divergences; [35]. A pragmatic approach proposed by Jauch and Rohrlich, [34] [42], and extended by Yennie, Frautschi, and Suura, [47],

circumvents this difficulty by considering *inclusive cross sections*. The inclusive cross sections are defined by summing over all possible final states that include photons whose total energy lies below an arbitrary threshold energy $\epsilon > 0$. Then the infrared divergences due to soft virtual photons are *formally* canceled by those corresponding to the emission of *soft* photons of total energy below ϵ , order by order in perturbation theory in powers of the fine structure constant α . A drawback of this approach becomes apparent when one tries to formulate a scattering theory that is ϵ -independent, because the transition probability, P^ϵ , is estimated to be $\mathcal{O}(\epsilon^{\text{const.}\alpha})$. Consequently, the threshold energy ϵ cannot be allowed to approach zero, unless “Bremsstrahlung” processes (emission of photons) are properly incorporated in the calculation. A more satisfactory approach to solve the infrared problem must go beyond inclusive cross sections and define α -dependent scattering states containing an infinite number of photons, the commonly referred *soft-photon clouds*, which are expected to yield finite and, in general, nonzero transition amplitudes, order by order in perturbation theory. This approach is guided by an *ansatz* identified in the analysis of certain solvable models introduced in early work by Bloch and Nordsieck, [9], and by Pauli and Fierz, [39], in the late 1930’s. In their seminal paper, [9], Bloch and Nordsieck have shown that, under certain approximations that render their model solvable, formal scattering states with a finite number of soft (asymptotic) photons do not belong to the physical Hilbert space of a system of asymptotically freely moving electrons interacting with the quantized radiation field. The works of Chung [16], Kibble [36], and Faddeev and Kulish [20], between 1965 and 1970, represent promising, albeit incomplete progress in this direction. In fact, the perturbative recipes for the construction of scattering states did not remove some of the major conceptual problems. New puzzles appeared that have stimulated a more fundamental analysis of the infrared features of QED in the framework of *general quantum field theory*; see [46], [10], [11], [26], and [37].

The previously described situation explains the interest in mathematically well defined models that are believed to retain the infrared features of QED, while describing a nontrivial interaction. After the fundamental works by J. Fröhlich [21], [22] from early 1970’s (where important constructive quantum field theory techniques have been put in place), the infrared problem has been investigated in the *standard model* of nonrelativistic QED or *Pauli Fierz model*, starting from the late 1990’s. This investigation is related to the study of atomic physics as a mathematically consistent theory. In spite of a host of qualitative and perturbative results which make atomic physics one of the most confirmed physical theories, very few mathematical proofs of basic facts of spectroscopy were available. Some of these mathematical difficulties are simply a manifestation of the infrared problem. In this respect, the extension of some analytical tools introduced in the *theory of Schrödinger operators* has permitted remarkable advance, as far as rigorous results are concerned. Conversely, the study of nonrelativistic QED models has stimulated new insights in areas like multi-scale methods, spectral analysis, scattering and expansion techniques.

Regarding our presentation, the reader is advised that we do not intend to provide a review of all the many and quite technical results produced in this area. This goes beyond the scope of a short review article. Rather we focus on a restricted number of problems, trying to explain the mathematical and physical contents with an emphasis on the links between them. In doing so, we will make use of some simplified models to explain more general concepts and results, which the reader can find in research papers. Therefore, in general, the mathematical details are limited to what is strictly necessary to introduce and explain the results. Two technical appendices complete these notes and provide more details about two novel techniques. The purpose of the appendices is to convey some basic ideas and the flavor of the techniques without providing complete proofs.

The structure of these notes is as follows. In Section 2, we describe the general set-up of nonrelativistic QED (at zero temperature). In Section 2.1, we analyze the spectral and dynamical features of a nonrelativistic model of an atom for energies below the ionization threshold. In Section 2.2, we discuss the *infrared catastrophe* in nonrelativistic QED. Appendix A is devoted to the construction of the atomic groundstate. Appendix B deals with the construction of the *infraparticle scattering states* for a nonrelativistic electron.

2. Spectroscopy in nonrelativistic QED

QED describes the interactions between electrically charged quantum-mechanical matter and the quantized electromagnetic field. The standard model of nonrelativistic QED or Pauli Fierz model arises as the low-energy limit of relativistic QED in which matter is treated nonrelativistically. It enables us to describe with considerable mathematical precision atoms, ions and molecules, i.e., electrons bound to nuclei, interacting with electromagnetic radiation. More precisely, in this model, matter consists of low-energy electrons, treated as nonrelativistic quantum-mechanical point particles with charge $-e$ and spin $1/2$, whose total number is conserved, and nuclei, treated as static or dynamical sources of an electrostatic potential corresponding to a localized charge Ze , where $Z = 1, 2, 3, \dots$ is the number of protons in the nucleus. The photons, which are the modes of the quantized electromagnetic field, are massless, neutral particles of helicity ± 1 . Configurations of photons can be described in terms of a quantized electromagnetic vector potential, $\vec{A}$. Interactions between electrons and photons are described by *minimal substitution*: The gauge-invariant velocity operator of an electron is given by

$$\vec{v} = \frac{\hbar}{im} \vec{\nabla}_{\vec{x}} + \frac{e}{mc} \vec{A}(\vec{x}),$$

where $\hbar$ is Planck's constant, m the mass of an electron, c the speed of light, and $\vec{A}(\vec{x})$ is the vector potential evaluated at the position, $\vec{x}$, of the electron. We choose the Coulomb gauge, $\vec{\nabla} \cdot \vec{A} = 0$. The finestructure constant $\alpha = \frac{e^2}{\hbar c}$ measures the strength of the interaction between electrons and photons. In the following, we use

units such that $\hbar = c = 1$.

In this framework, the Hamiltonian

$$H = \sum_{j=1}^N \frac{[\vec{\sigma}_j \cdot (-i\vec{\nabla}_{\vec{x}_j} + \alpha^{1/2}\vec{A}(\vec{x}_j))]^2}{2m} + \sum_{j=1}^N \sum_{m=1}^M \frac{-\alpha Z_m}{|\vec{x}_j - \vec{R}_m|} + \sum_{1 \leq i < j \leq N} \frac{\alpha}{|\vec{x}_i - \vec{x}_j|} + H^f$$

provides an effective description of a molecule with N electrons, where $\vec{\sigma}_j = (\sigma_j^1, \sigma_j^2, \sigma_j^3)$ denotes the three Pauli matrices associated with the j^{th} electron, $\vec{x}_j$ is its position, $\vec{R}_1, \dots, \vec{R}_M \in \mathbb{R}^3$ are the positions of M static nuclei with atomic numbers $Z_1, \dots, Z_M$, and H^f is the Hamiltonian of the free electromagnetic field. (Rigorous definitions are deferred to the next section.) The Hilbert space of the system is the tensor product, $\mathcal{H}_{\text{el}} \otimes \mathcal{F}$, of the Hilbert space, $\mathcal{H}_{\text{el}}$, of N electrons and of the photon Fock space $\mathcal{F}$ where:

- $\mathcal{H}_{\text{el}} = \mathcal{A}_N L^2[(\mathbb{R}^3 \times \mathbb{Z}_2)^N]$, with $\mathcal{A}_N$ the projection onto completely antisymmetric functions in N arguments;
- $\mathcal{F} = \oplus_{n \geq 0} \mathcal{F}_n$, where $\mathcal{F}_0 = \mathbb{C}\Omega$ is the non-degenerate subspace corresponding to the vacuum vector Ω , and $\mathcal{F}_n = \mathcal{S}_n L^2(\mathbb{R}^3 \times \mathbb{Z}_2)^{\otimes n}$, with $\mathcal{S}_n$ the projection onto completely symmetric functions in n arguments.

Despite simplifications, two prototypes of *Pauli Fierz models* are usually studied, and most of the results can be generalized to heavier atoms and to molecules. They are:

- i) An atom, for simplicity consisting of a single, static proton of charge e surrounded by one electron, interacting with the quantized radiation field (i.e., the Hydrogen atom);
- ii) One electron only interacting with the radiation field.

For most of the qualitative spectral and scattering features of the models, the spin of the electron turns out to be an inessential complication, and, for simplicity, the Zeeman coupling of the magnetic moment of the electron to the quantized magnetic field can be neglected.

In the formal expression of the Hamiltonian H , interactions between electrons and photons of very large momentum and energy become so strong that the model is ill-defined, mathematically. This is known as the *ultraviolet problem*, and it is avoided by introducing a cutoff in $\vec{A}(\vec{x})$ that turns off the interactions between electrons and photons with energy above a certain threshold Λ . The ultraviolet cutoff Λ can be chosen to be arbitrarily large, but, for physical reasons, is typically chosen to be of the order of the rest energy of an electron, i.e., $\Lambda \sim m$.

When the ultraviolet cutoff is introduced, the dynamics is generated by a self-adjoint Hamilton operator acting on the Hilbert space of pure state vectors of the system, and the resulting model of QED is mathematically well defined. Nevertheless, the fact that photons are massless particles introduces substantial difficulties into the mathematical analysis of the Pauli Fierz models known collectively as the *infrared problem* of (nonrelativistic) QED.

2.1. Atoms and molecules

To summarize some of the main results achieved so far in the spectral analysis of atomic systems and in the study of scattering of photons at an atom, our reference model will be the Hydrogen atom in a minimal description where we also neglect the electron spin. In this approximation, the Hilbert space of pure states of the system is given by $\mathcal{H} = \mathcal{H}_{\text{el}} \otimes \mathcal{F}$, where $\mathcal{H}_{\text{el}} = L^2(\mathbb{R}^3)$ is the Hilbert space of the spinless electron.

The Hamiltonian generating the dynamics of the system is

$$H_{\text{at}} := \frac{(-i\vec{\nabla}_{\vec{x}} + \alpha^{\frac{1}{2}}\vec{A}_{\Lambda}(\vec{x}))^2}{2m} - \frac{\alpha}{|\vec{x}|} + H^f, \quad (1)$$

where $\vec{A}_{\Lambda}(\vec{x})$ is the vector potential, $\vec{A}(\vec{y})$, cutoff at energies larger than Λ and evaluated at $\vec{y} \equiv \vec{x}$. It is convenient to describe the vector potential $\vec{A}_{\Lambda}(\vec{y})$ and the Hamiltonian H^f in terms of the usual photon creation- and annihilation operator-valued distributions, $\{a_{\vec{k},\lambda}^*, a_{\vec{k},\lambda} : \vec{k} \in \mathbb{R}^3, \lambda \in \mathbb{Z}_2\}$, acting on the Fock space and satisfying the canonical commutation relations

$$[a_{\vec{k},\lambda}, a_{\vec{k}',\lambda'}^*] = \delta_{\lambda\lambda'} \delta(\vec{k} - \vec{k}'), \quad [a_{\vec{k},\lambda}^{\#}, a_{\vec{k}',\lambda'}^{\#}] = 0, \quad (2)$$

with $a^{\#} = a$ or a^* . The vacuum vector Ω obeys the condition $a_{\vec{k},\lambda} \Omega = 0$ for all $\vec{k} \in \mathbb{R}^3$ and $\lambda \in \mathbb{Z}_2 \equiv \{+, -\}$. Thereafter the quantized electromagnetic vector potential cutoff at Λ is given by

$$\vec{A}_{\Lambda}(\vec{y}) := \frac{1}{(2\pi)^{3/2}} \sum_{\lambda=\pm} \int_{\mathcal{B}_{\Lambda}} \frac{d^3k}{\sqrt{2|\vec{k}|}} \{ \vec{\varepsilon}_{\vec{k},\lambda} e^{-i\vec{k}\cdot\vec{y}} a_{\vec{k},\lambda}^* + \vec{\varepsilon}_{\vec{k},\lambda}^* e^{i\vec{k}\cdot\vec{y}} a_{\vec{k},\lambda} \}, \quad (3)$$

where $\vec{\varepsilon}_{\vec{k},-}$, $\vec{\varepsilon}_{\vec{k},+}$ are photon polarization vectors (i.e., two unit vectors in $\mathbb{R}^3 \otimes \mathbb{C}$ satisfying $\vec{\varepsilon}_{\vec{k},\lambda}^* \cdot \vec{\varepsilon}_{\vec{k},\mu} = \delta_{\lambda\mu}$, $\vec{k} \cdot \vec{\varepsilon}_{\vec{k},\lambda} = 0$, for $\lambda, \mu = \pm$), and $\mathcal{B}_{\Lambda}$ is a ball of radius Λ centered at the origin in momentum space. In order to derive some of the results described in the later sections, for technical reasons, the step function, in the $|\vec{k}|$ -variable, corresponding to the sharp ultraviolet cutoff Λ must be replaced by a smooth function $\Lambda(\vec{k})$, with properties dictated by the technical tool that is employed. The reader can find the precise requirements in the quoted research papers. We implicitly assume these modifications in the definition of the model whenever it is necessary. The Hamiltonian H^f is given by

$$H^f := \sum_{\lambda=\pm} \int d^3k |\vec{k}| a_{\vec{k},\lambda}^* a_{\vec{k},\lambda}. \quad (4)$$

In this review of rigorous results in atomic spectroscopy, we are interested in configurations of the system where the electron is bound to the nucleus. The energy of the corresponding states is strictly below the ionization threshold Σ and above the ground state energy E_{gs} . Within this energy range, the dynamics of the system can be described in terms of atomic transitions between metastable states that originate from the perturbation of the excited eigenstates of the Coulomb system. Therefore,

it is convenient to re-scale the physical variables with respect to typical lengths and energies involved in these atomic transitions, and to express the Hamiltonian in terms of the re-scaled variables. A conventional change of variables goes as follows

$$\vec{x} \rightarrow \frac{r_{\text{Bohr}}}{2} \vec{x} \quad , \quad \vec{k} \rightarrow \frac{2\alpha}{r_{\text{Bohr}}} \vec{k}, \quad (5)$$

where $r_{\text{Bohr}} = \frac{\hbar^2}{me^2} = \frac{1}{m\alpha}$ is the Bohr radius. The unit of energy is chosen such that $2m\alpha^2 = 4Rydberg = 1$. The Hamiltonian H_{at} expressed in the new variables reads

$$H_{\text{at}} := (-i\vec{\nabla}_{\vec{x}} + \alpha^{\frac{3}{2}} \vec{A}_{\Lambda}(\alpha\vec{x}))^2 - \frac{1}{|\vec{x}|} + H^f, \quad \Lambda \sim 1/(2\alpha^2). \quad (6)$$

We note that, in Expression (6), the Coulomb interaction is α -independent.

2.1.1. Spectral analysis

The spectrum of the Hamiltonian H_{at} results by perturbation of the spectrum (see Figure 1) of the unperturbed Hamiltonian

$$H_{\text{at}}^{(0)} := -\Delta_{\vec{x}} - \frac{1}{|\vec{x}|} + H^f, \quad (7)$$

by means of the interaction

$$W(\alpha^{\frac{1}{2}}) := H_{\text{at}} - H_{\text{at}}^{(0)} = 2\alpha^{\frac{3}{2}} \vec{A}_{\Lambda}(\alpha\vec{x}) \cdot (-i\vec{\nabla}_{\vec{x}}) + \alpha^3 \vec{A}_{\Lambda}(\alpha\vec{x})^2. \quad (8)$$

The spectrum of $H_{\text{Coul}} := -\Delta_{\vec{x}} - \frac{1}{|\vec{x}|}$ is absolutely continuous above an ionization threshold Σ_0 . Below Σ_0 , the spectrum is discrete, starting from the non-degenerate groundstate energy E_0 . H^f has a non-degenerate eigenvalue $E_{\Omega} \equiv 0$ at the bottom of a branch of absolutely continuous spectrum of infinite multiplicity. The analytic

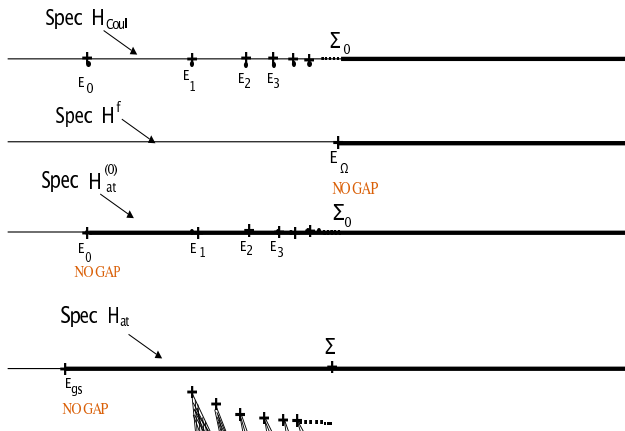


Fig. 1. The spectra of the Hamiltonians H_{Coul} , H^f , $H_{\text{at}}^{(0)}$, and H_{at} .

continuation of the resolvent of H_{at} onto the second Riemann sheet of the complex-energy plane displays branch-type singularities (resonances) in the lower half plane (see Figure 1), located near the eigenvalues of $H_{\text{at}}^{(0)}$; for the details of the singularities see [43]. This explains basic facts of spectroscopy: *spectral lines, lifetimes, relaxation to the groundstate*. Although the physical arguments that explain the features of the spectrum of H_{at} are cornerstones of quantum mechanics, they have been translated into rigorous mathematics only in recent times. This is mostly because methods of conventional analytic perturbation theory applied to the unperturbed Hamiltonian $H_{\text{at}}^{(0)}$ and to the interaction $W(\alpha^{\frac{1}{2}})$ are inapplicable, and methods of *multi-scale analysis* have been developed in order to rigorously study problems like the construction of the groundstate vector $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$, and to analyze the fate of the excited eigenstates of $H_{\text{at}}^{(0)}$ (see Figure 1).

Listed below are some of the main results from this study.

- i-1)** H_{at} has an eigenvalue, $E_{\text{gs}} \equiv E_{\text{gs}}(\alpha)$, at the bottom of its spectrum corresponding to a normalizable atomic groundstate. This result holds for arbitrarily large values of α and Λ ; see [31]. The spectrum of H_{at} is a half-infinite interval $[E_{\text{gs}}, \infty)$ of the real line. In particular, the continuous spectrum of H_{at} is given by $[E_{\text{gs}}, \infty)$. Thus, E_{gs} is an eigenvalue of H_{at} that is *not* isolated; rather, it is at the bottom of a branch of continuous spectrum of infinite multiplicity. This is because photons are *massless* particles.

The construction of the groundstate $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$ and the α -dependence of $E_{\text{gs}}(\alpha)$ and of $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$ have been open problems for a long time. Indeed, if one attempts to formally apply perturbation theory in $\alpha^{\frac{1}{2}}$ to derive formulae to $E_{\text{gs}}(\alpha)$ and $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$, one finds contributions to the Taylor coefficients of sufficiently high order that are divergent. The fact that photons are massless particles manifests itself in an accumulation of nearly vanishing energy denominators in the formal expressions of Rayleigh-Schrödinger perturbation theory yielding infrared-divergent integrals.

- i-2)** In the small coupling regime, i.e., for small values of α , the groundstate vector, $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$, has been constructed by using multi-scale methods. First (see [2] and references therein), the *operator-theoretic renormalization group analysis* based on the Feshbach map has been applied to a regularized model where the field-electron interaction is turned off at large distances from the nucleus. Later, the groundstate vector of the Hamiltonian H_{at} has been constructed by applying the *iterative analytic perturbation theory*, a new multi-scale technique by which one can handle the singularities of Rayleigh-Schrödinger perturbation theory without any infrared regularization of the interaction between the electromagnetic field and the electron (see Appendix A). Recently, the construction of $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$ has been also carried out in [43], by a refinement of the technique in [2].
- i-3)** For the treatment of the infrared singularities appearing in a formal expansion of $E_{\text{gs}}(\alpha)$ in α , an asymptotic expansion up to some finite order has been derived

in [32] by using minimizers. Later, an algorithm based on multi-scale expansions in the photon momenta has been introduced in [5] to calculate $E_{\text{gs}}(\alpha)$ and $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$ to arbitrary precision, namely up to rigorously controlled remainder terms of $o(\alpha^N)$, for arbitrary $N < \infty$. The expansion for the energy is

$$E_{\text{gs}}(\alpha) = E_0 + \sum_{l=3}^N \varepsilon_l(\alpha) \alpha^l + o(\alpha^N) \quad (9)$$

with $\lim_{\alpha \rightarrow 0} \alpha^\delta |\varepsilon_l(\alpha)| = 0 \ \forall \delta > 0$, where $\varepsilon_l(\alpha)$ are computable in terms of finitely many convergent integrals, for any $3 \leq l < \infty$. (We note, however, that for some slighter simple models with less severe infrared divergences it has been shown that quantities analogous to $E_{\text{gs}}(\alpha)$ and $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$ are analytic in the coupling constant in a small disk around the origin; see [30].)

We remark that the control on the α -dependence of $E_{\text{gs}}(\alpha)$ and of $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$ is one of the ingredients in a mathematically precise analysis of the scattering of low energy photons at atoms, i.e., *Rayleigh scattering*; see point ii-4), Section 2.1.2.

Regarding the fate of the excited state energy levels of $H_{\text{at}}^{(0)}$ as the interaction is turned on, the *operator-theoretic renormalization group analysis* has been combined with the method of complex spectral deformation to show that higher eigenvalues of $H_{\text{at}}^{(0)}$ migrate to the lower complex half-plane when the perturbation $W(\alpha^{\frac{1}{2}})$ is added to $H_{\text{at}}^{(0)}$. The proof requires a condition on the electric-dipole matrix elements ensuring that the state may decay by emission of just one photon.

i-4) For small values of α , and in presence of one-photon transitions, it has been proven that the excited state energy levels of $H_{\text{at}}^{(0)}$ dissolve in the continuum and turn to resonances, i.e., branch-point singularities in the lower complex half plane. For references, see [7], [8], [18], [24], [29], [27].

Mourre's method of positive commutators has also been employed to prove that, under the same condition on the electric-dipole matrix elements, the spectrum of H_{at} is absolutely continuous above E_{gs} up to $\Sigma_0 - \delta_\alpha$ where δ_α is a small positive number, $\delta_\alpha \rightarrow 0$ as $\alpha \rightarrow 0$; see [8], [24]. More generally, a multi-scale virial argument can be used (see [27]) to exclude eigenvalues in certain intervals, also under the condition that the decay is due to a two-photon transition. (In fact, this happens for the 2S energy level of the Hydrogen atom.)

2.1.2. Asymptotic completeness, metastable states, and Bohr's frequency condition.

In the analysis of atomic transitions between states of energy below the ionization threshold, various questions have been addressed in order to add mathematical rigor to basic facts of spectroscopy. Among them we focus on four mathematical goals:

(a) The construction of the asymptotic states describing the atom in the ground-state plus asymptotic photons; (b) The proof of unitarity of the scattering matrix (asymptotic completeness), which means that the spectral subspace corresponding to energies below the ionization threshold Σ is spanned by the asymptotic state vectors; (c) A nonperturbative derivation of *Fermi's Golden Rule* for the decay of the excited states of the Hamiltonian $H_{\text{at}}^{(0)}$; (d) A mathematical justification of *Bohr's frequency condition*.

The results obtained so far can be summarized as follows.

ii-1) We recall that the asymptotic creation- and annihilation operators describing a free asymptotic electromagnetic field are defined by the (strong) limits

$$a_{\text{out/in}}^{\#}(f) \quad (10)$$

$$:= s - \lim_{t \rightarrow \pm\infty} \sum_{\lambda=\pm} e^{iH_{\text{at}}t} e^{-iH_{\text{at}}^{(0)}t} \int a_{\lambda}^{\#}(\vec{k}) f_{\lambda}^{\#}(k) d^3k e^{iH_{\text{at}}^{(0)}t} e^{-iH_{\text{at}}t}$$

with $(a^{\#}, f^{\#}) = (a, \bar{f})$ or (a^*, f) , and where $f_{\lambda}(k)$, $\lambda = \pm$, are photon test functions.

The vectors of the form

$$a_{\text{out/in}}^*(f^{(m)}) \dots a_{\text{out/in}}^*(f^{(1)}) \phi_{\text{gs}}(\alpha^{\frac{1}{2}}) \quad (11)$$

have been rigorously constructed and represent states where the system consists of the atom in the groundstate $\phi_{\text{gs}}(\alpha^{\frac{1}{2}})$ plus a finite (arbitrary) number of out/in asymptotic photons with wave functions $\{f_{\lambda}^{(j)} | 1 \leq j \leq m\}$. The S -matrix elements are defined by

$$\langle a_{\text{in}}^*(g^{(m')}) \dots a_{\text{in}}(g^{(1)}) \phi_{\text{gs}}(\alpha^{\frac{1}{2}}), S a_{\text{in}}^*(f^{(m)}) \dots a_{\text{in}}^*(f^{(1)}) \phi_{\text{gs}}(\alpha^{\frac{1}{2}}) \rangle \quad (12)$$

$$:= \langle a_{\text{out}}^*(g^{(m')}) \dots a_{\text{out}}^*(g^{(1)}) \phi_{\text{gs}}(\alpha^{\frac{1}{2}}), a_{\text{in}}^*(f^{(m)}) \dots a_{\text{in}}^*(f^{(1)}) \phi_{\text{gs}}(\alpha^{\frac{1}{2}}) \rangle$$

where $\langle, \rangle$ is the scalar product in $\mathcal{H}$.

ii-2) Asymptotic completeness has been proven in the subspace corresponding to energies below the ionization threshold Σ (see Figure (1)), but under some conditions, namely either by assuming that the photons have a mass or/and by introducing an arbitrarily small infrared cutoff $\sigma > 0$ in the interaction $W(\alpha^{\frac{1}{2}})$ (i.e., in the definition of $\vec{A}_{\Lambda}$, in Eq. (3), one restricts the integral to the region $\mathcal{B}_{\Lambda} \setminus \mathcal{B}_{\sigma}$). The proofs rely on techniques from N -body potential scattering that have been adapted to quantum systems with infinitely many degrees of freedom; see [19], [23]. A crucial role is played by an *a priori* estimate on the number of the emitted photons in terms of the energy of the state. Such an easy bound is available only because of the unphysical photon mass or because of the infrared cut-off in the interaction between the atom and the quantized radiation field.

We remark that the scattering matrix restricted to the subspace corresponding to energies below the ionization threshold (see Figure 1) is expected to be unitary irrespective of any infrared regularization of the type described above. Currently,

the best result in this direction is *relaxation to the groundstate*, thanks to the proof of LAP (limiting absorption principle) for a spectral interval at the infimum of the energy spectrum; for a precise formulation of this result, see [24], [25]. In order to establish asymptotic completeness we still miss a key mathematical idea to substantiate the physical intuition that no cloud of infinitely many soft photons should be emitted in *Rayleigh scattering*. (*Rayleigh scattering* refers to the scattering of low energy photons at atoms such that the total energy of the system does not exceed the ionization threshold Σ).

- ii-3) An alternative approach to shed light on the dynamical features of the *metastable* atomic states is based on the analysis of the eigenvectors of a complex-dilated Hamiltonian, H_{at}^θ , obtained from H_{at} by complex dilation, where θ is the complex parameter associated to the dilation (for details see, e.g., [7]). In fact, one starts from a state that is identical or perturbatively close to the unperturbed eigenstate corresponding to an embedded eigenvalue of the Hamiltonian $H_{\text{at}}^{(0)}$ that is turned to a resonance (see Figure 1) when the interaction is turned on. This vector is then multiplied by an operator that is a bounded function of H_{at} with support in an interval around the unperturbed eigenvalue. One then studies the time-evolution under $\exp(-itH_{\text{at}})$ of the state so defined, and, using the dilatation analyticity of the unperturbed eigenvector and of H_{at} , estimates the survival probability. This argument shows the relaxation of the metastable states, and relates their lifetime to the inverse of the imaginary part of the corresponding branch-point in the complex-energy plane; see also [1].

Bohr's frequency condition in atomic spectroscopy has been rigorously justified at least in the following weaker version. Consider the transition amplitude (12) for just one incoming and one outgoing photon, with the same wave function f . If the frequencies $\omega(\vec{k}) = |\vec{k}|$, with $\vec{k}$ any point in the support of f , do not fulfill the resonant condition, more precisely if $\omega(\vec{k}) \neq E_i - E_0 \quad \forall i$, where E_i and E_0 are the energies of any excited state and of the groundstate of $H_{\text{at}}^{(0)}$, respectively, then, to leading order in α , the imaginary part of the corresponding T -matrix element is identically zero, where $T := i(S - 1)$. (Only the imaginary part of T matters in the (formal) calculation of the total cross section of the incoming photon state.)

- ii-4) Concerning *Rayleigh scattering*, in a recent paper, [6], it has been understood how to calculate the transition amplitudes in Eq. (12) up to any order in the coupling constant α . This expansion provides a mathematical proof of the validity of Bohr's frequency condition previously described.

However, we remark that the result in [6] is not completely satisfactory yet, because it is restricted to scattering amplitudes (see Eq. (12)) where the wave functions $\{f^{(j)}, g^{(j')}\}$ are α -independent. Therefore the widths, in the $|\vec{k}|$ -variable, of the wave functions are comparable to the energy shifts $\Delta_i := E_i - E_0 = \mathcal{O}(1)$. More ambitiously, one would like to translate a stronger physical statement, based

on *formal* expansions, into rigorous mathematics: To leading order in α , the total cross section for an incoming photon whose wave function does not contain any resonant frequency (i.e., $\omega(\vec{k}) \neq E_i - E_0, \forall i$) is identically zero. This type of result requires control on scattering amplitudes where the wave functions of the incoming and outgoing photons may have arbitrarily small widths in the $|\vec{k}|$ -variable, in particular for widths comparable to the inverse of the life-time of the metastable states, which is $\mathcal{O}(\alpha^3)$ in the rescaled variables.

2.2. Infraparticles and Čerenkov radiation

In the framework of nonrelativistic QED, the system consisting of one (spinless) electron only interacting with the quantized radiation field is described by the Hamiltonian

$$H_{\text{el}} := \frac{(-i\vec{\nabla}_{\vec{x}} + \alpha^{\frac{1}{2}}\vec{A}_{\Lambda}(\vec{x}))^2}{2m} + H^f, \quad (13)$$

acting on $L^2(\mathbb{R}^3) \otimes \mathcal{F}$, and where Λ is comparable to m (no rescaling of units is implemented here). For “small” (i.e., < 1) electron velocities, this simple model system represents an important tool to analyze the *infrared catastrophe* in QED. Furthermore, for “large” (i.e., > 1) electron velocities, we observe the appearance of the Čerenkov radiation, which is an example of quantum friction, a mathematically poorly understood phenomenon.

2.2.1. Dressed electron states

The spectrum of the Hamiltonian H_{el} has been analyzed with regard to the existence of the single-particle or dressed electron states, $\{\psi\}$, which form the mass shell subspace of the electron, i.e., the subspace spanned by the vectors in the Hilbert space that satisfy the equation

$$H_{\text{el}}\psi = E(\vec{P})\psi, \quad (14)$$

where the function $E(\cdot)$ is the renormalized electron dispersion, and $\vec{P} = -i\vec{\nabla}_{\vec{x}} + \vec{P}^f$ is the total momentum operator of the system, with $\vec{P}^f := \sum_{\lambda=\pm} \int d^3k \vec{k} a_{\vec{k},\lambda}^* a_{\vec{k},\lambda}$ the momentum operator of the photon field. By translation invariance, H_{el} commutes with $\vec{P}$. Using the joint spectral decomposition of $\vec{P}$ and H_{el} , the Hilbert space can be decomposed as the direct integral $\mathcal{H} = \int^{\oplus} d^3P \mathcal{H}_{\vec{P}}$, where the fiber spaces $\mathcal{H}_{\vec{P}}$ are invariant under the action of $\vec{P}$ and of H_{el} , i.e., $H_{\text{el}} = \int^{\oplus} d^3P (H_{\text{el}})_{\vec{P}}$ and $(H_{\text{el}})_{\vec{P}} : \mathcal{H}_{\vec{P}} \rightarrow \mathcal{H}_{\vec{P}}$. If the dressed electron states exist, they are of the form

$$\psi \equiv \psi(h) = \int^{\oplus} d^3P h(\vec{P}) \psi_{\vec{P}}, \quad (15)$$

where the functions $\{h\}$ are square-integrable and the fiber vectors $\psi_{\vec{P}}$ are the groundstates of the fiber Hamiltonians $(H_{\text{el}})_{\vec{P}}$.

- i-1)** For small values of α , as long as an arbitrarily small infrared cutoff, $\sigma > 0$, in the photon energy is imposed on $\vec{A}_\Lambda(\vec{x})$, the dressed electron states, $\{\psi^\sigma\}$, of the corresponding Hamiltonian (that we denote by H_{el}^σ) have been constructed in the spectral region $|\vec{P}| < m - \delta_\alpha$, with $(0 <) \delta_\alpha \rightarrow 0$ as $\alpha \rightarrow 0$. As $\sigma \rightarrow 0$, the mass-shell subspace dissolves, i.e., $w - \lim_{\sigma \rightarrow 0} \psi^\sigma = 0$. Particles sharing this property are called *infraparticles*; see [44]. The asymptotic properties, as $\sigma \rightarrow 0$, of the states $\{\psi^\sigma\}$ provide crucial information concerning the scattering features of the system; see [40], [13], [14], [3], and [28].
- i-2)** In classical electrodynamics, the Čerenkov radiation is the electromagnetic radiation emitted by a charged particle that travels through a dielectric medium with a speed greater than that at which light propagates in the same medium. In the quantum model described by the Hamiltonian H_{el} , for values of the total momentum, $\vec{P}$, such that $|\vec{P}| > m$, the group velocity of the electron is expected to exceed $1(= c)$. Therefore, the electron mass shell subspace is expected not to exist in that region even imposing an infrared cutoff on $\vec{A}_\Lambda(\vec{x})$ that turns off the interaction with low energy photons. The only result (see [17]) concerning this spectral region is concerned with the instability of the unperturbed mass shell when the interaction is switched on. The result is formulated as follows: For $|\vec{P}| > m + \delta_\alpha$, no putative mass shell, *regular* in the variable $\vec{P}$, i.e. $\|\frac{\partial \psi_{\vec{P}}}{\partial \vec{P}}\| < \infty$, can be perturbatively close (in α) to the unperturbed mass shell.

2.2.2. Infraparticle scattering states

An asymptotic electron is always surrounded by a cloud of infinitely many low energy asymptotic photons with wave function depending on the electron asymptotic velocity according to the Bloch-Nordsieck paradigm [9] (see also Eq. (49) in Appendix B).

- ii-1)** For the nonrelativistic QED model described by the Hamiltonian in Eq. (13), the infraparticle scattering states have been constructed in [15], by extension of a scattering technique introduced, in [41], for the Nelson model. The key ingredients for this result are the spectral analysis of the infrared singularities of the electron mass shell, point i-1) in Section 2.2.1, and a suitable treatment of the long range interactions.

While referring the reader to Appendix B for details about this construction, we now recall a useful point of view based on *classical* electrodynamics to understand why free radiation parametrized by the asymptotic velocities of the charged particles must be expected to be present in *all* the scattering states.

We consider a single, classical charged point-particle, e.g., an electron, moving along a world line $(t, \vec{x}(t))$ in Minkowski space, with $\vec{x}(0) = \vec{0}$. For simplicity, we may suppose that, for $t \leq 0$, it moves at a constant velocity $\vec{v}_{in}$, and, for $t > \bar{t} > 0$,

at a constant velocity $\vec{v}_{out} \neq \vec{v}_{in}$, $|\vec{v}_{out}|, |\vec{v}_{in}| < 1 (= c)$. Thus,

$$\vec{x}(t) = \vec{v}_{in} \cdot t, \quad \text{for } t \leq 0, \quad (16)$$

$$\vec{x}(t) = \vec{x}_* + \vec{v}_{out} \cdot t, \quad \text{for } t \geq \bar{t}, \quad (17)$$

for some $\vec{x}_*$. For times $t \in [0, \bar{t}]$, the particle performs an accelerated motion. We propose to analyze the properties of the free electromagnetic radiation at very early times ($t \rightarrow -\infty$, “in”) and very late times ($t \rightarrow +\infty$, “out”). For this purpose, we must solve Maxwell’s equations for the electromagnetic field tensor, $F^{\mu\nu}(t, \vec{y})$, given the 4-current density corresponding to the trajectory of the particle; (back reaction of the electromagnetic field on the motion of the charged particle is neglected):

$$\partial_\mu F^{\mu\nu}(t, \vec{y}) = J^\nu(t, \vec{y}) \quad (18)$$

with

$$J^\nu(t, \vec{y}) := -\alpha^{1/2} (\delta^{(3)}(\vec{y} - \vec{x}(t)), \dot{\vec{x}}(t) \delta^{(3)}(\vec{y} - \vec{x}(t))) \quad |\dot{\vec{x}}(t)| < 1. \quad (19)$$

We solve equation (18) with prescribed spatial asymptotics ($|\vec{y}| \rightarrow \infty$): Let $F_{[\vec{v}_{L.W.}]}^{\mu\nu}(t, \vec{y})$ be a solution of (18) that, to leading order in $|\vec{y}|^{-1}$ ($|\vec{y}| \rightarrow \infty$), approaches the Liénard-Wiechert field tensor for a point-particle with charge $-\alpha^{1/2}$ and a *constant* velocity $\vec{v}_{L.W.}$ at *all* times. Let us denote the Liénard-Wiechert field tensor of a point-particle with charge $-\alpha^{1/2}$ moving along a trajectory $(t, \vec{x}(t))$ in Minkowski space with $\vec{x}(0) =: \vec{x}$ and $\dot{\vec{x}}(t) \equiv \vec{v}$, for *all* t , by $F_{\vec{x}, \vec{v}}^{\mu\nu}(t, \vec{y})$. Apparently, we are looking for solutions, $F_{[\vec{v}_{L.W.}]}^{\mu\nu}(t, \vec{y})$, of (18) with the property that, for all times t ,

$$|F_{[\vec{v}_{L.W.}]}^{\mu\nu}(t, \vec{y}) - F_{\vec{x}, \vec{v}_{L.W.}}^{\mu\nu}(t, \vec{y})| = o(|\vec{y}|^{-2}), \quad (20)$$

as $|\vec{y}| \rightarrow \infty$, for any $\vec{x}$. This class of solutions of (18) is denoted by $\mathcal{C}_{\vec{v}_{L.W.}}$. It is important to observe that, by causality, the class $\mathcal{C}_{\vec{v}_{L.W.}}$ is *non-empty*, for any $\vec{v}_{L.W.}$, with $|\vec{v}_{L.W.}| < 1 (= c)$. (This can be seen by choosing Cauchy data for the solution of (18) satisfying (20) at some time t_0 , e.g., $t_0 = 0$.)

Let us now consider a specific solution, $F_{[\vec{v}_{L.W.}]}^{\mu\nu}(t, \vec{y})$, of Eq. (18) in the class $\mathcal{C}_{\vec{v}_{L.W.}}$. We are interested in the behavior of this solution at very early times ($t \ll 0$) and at very late times ($t \gg 0$). We note that, by (16), $F_{0, \vec{v}_{in}}^{\mu\nu}(t, \vec{y})$ solves Eq. (18), for times $t < 0$. Thus,

$$\phi_{in}^{\mu\nu}(t, \vec{y}) := F_{[\vec{v}_{L.W.}]}^{\mu\nu}(t, \vec{y}) - F_{0, \vec{v}_{in}}^{\mu\nu}(t, \vec{y}) \quad t < 0 \quad (21)$$

solves the *homogenous* Maxwell equation, i.e., Eq. (18) with $J^\nu \equiv 0$. Analogously, by (17), $F_{\vec{x}_*, \vec{v}_{out}}^{\mu\nu}(t, \vec{y})$ solves Eq. (18), for times $t > \bar{t}$. Thus,

$$\phi_{out}^{\mu\nu}(t, \vec{y}) := F_{[\vec{v}_{L.W.}]}^{\mu\nu}(t, \vec{y}) - F_{\vec{x}_*, \vec{v}_{out}}^{\mu\nu}(t, \vec{y}) \quad t > \bar{t} \quad (22)$$

also solves the homogenous Maxwell equation.

Next, we recall that $\phi_{as}^{\mu\nu}(t, \vec{y})$, with $as = out/in$, can be derived from an electromagnetic vector potential, A_{as}^μ , by

$$\phi_{as}^{\mu\nu}(t, \vec{y}) = \partial^\mu A_{as}^\nu(t, \vec{y}) - \partial^\nu A_{as}^\mu(t, \vec{y}). \quad (23)$$

We can impose the *Coulomb gauge condition* on A_{as}^μ : $A_{as}^\mu = (0, \vec{A}_{as}(t, \vec{y}))$, with $\vec{\nabla} \cdot \vec{A}_{as}(t, \vec{y}) \equiv 0$. It turns out that, to leading order in $|\vec{y}|^{-1}$ ($|\vec{y}| \rightarrow \infty$), $\vec{A}_{as}(t, \vec{y})$ is given by

$$\begin{aligned} \vec{A}_{as}(t, \vec{y}) \\ [|\vec{y}| \rightarrow \infty] = & \alpha^{\frac{1}{2}} \sum_{\lambda} \int \frac{d^3 k}{(2\pi)^{3/2} \sqrt{2|\vec{k}|}} \left\{ \frac{\vec{v}_{as} \cdot \vec{\varepsilon}_{\vec{k}, \lambda}^*}{|\vec{k}|^{\frac{3}{2}} (1 - \vec{v}_{as} \cdot \hat{\vec{k}})} \vec{\varepsilon}_{\vec{k}, \lambda} e^{-i\vec{k} \cdot \vec{y} + i|\vec{k}|t} + c.c. \right\} \\ & - \alpha^{\frac{1}{2}} \sum_{\lambda} \int \frac{d^3 k}{(2\pi)^{3/2} \sqrt{2|\vec{k}|}} \left\{ \frac{\vec{v}_{L.W.} \cdot \vec{\varepsilon}_{\vec{k}, \lambda}^*}{|\vec{k}|^{\frac{3}{2}} (1 - \vec{v}_{L.W.} \cdot \hat{\vec{k}})} \vec{\varepsilon}_{\vec{k}, \lambda} e^{-i\vec{k} \cdot \vec{y} + i|\vec{k}|t} + c.c. \right\}. \end{aligned} \quad (24)$$

Analogous statements hold in the model of quantum electrodynamics with non-relativistic matter described by the Hamiltonian H_{el} , where the motion of the quantum particle depends on the back reaction of the field, and the asymptotic *in*- and *out*-velocities of this particle are not attained at finite times.

In fact, the set of classes $\mathcal{C}_{\vec{v}_{L.W.}}$, associated with different currents but at fixed $\vec{v}_{L.W.}$, corresponds to one of the superselection sectors of the quantized theory; see e.g. [3]. In particular, the Fock representation, which is the usual (but not the only possible) choice for the representation of the algebra of photon creation- and annihilation operators, corresponds to $\vec{v}_{L.W.} = 0$. This implies that, in the Fock representation of the interpolating photon creation- and annihilation operators, an infrared-singular asymptotic electromagnetic-field configuration must be present for all values of the asymptotic velocity of the electron different from zero. In particular, after replacing the classical velocities with the spectral values of the *quantum* operators $\vec{v}_{\text{out/in}}$ – which are the asymptotic velocities out/in of the electron in the model described by the Hamiltonian H_{el} – the background field in Eq. (24) with $\vec{v}_{L.W.} = 0$ corresponds to the background radiation described by the coherent (non-Fock) representations of the asymptotic photon algebra labeled by $\vec{v}_{\text{out/in}}$; see Appendix B.

Appendix A

The *iterative analytic perturbation method* enables us to construct the ground-state of the Hamiltonian H_{at} by subsequent perturbations of the groundstate of the Hamiltonian

$$H_{\text{at}}^{(0)} := (-i\vec{\nabla}_{\vec{x}})^2 - \frac{1}{|\vec{x}|} + H^f. \quad (25)$$

The method is based on the following preparatory definitions:

i) The regularized Hamiltonians

$$H_{\text{at}}^{\sigma_j} := (-i\vec{\nabla}_{\vec{x}} + \alpha^{3/2} \vec{A}_{\Lambda}^{\sigma_j}(\alpha \vec{x}))^2 - \frac{1}{|\vec{x}|} + H^f, \quad (26)$$

with

$$\begin{aligned} & \vec{A}_\Lambda^{\sigma_j}(\alpha\vec{x}) \\ &:= \frac{1}{(2\pi)^{\frac{3}{2}}} \sum_\lambda \int_{\mathcal{B}_\Lambda \setminus \mathcal{B}_{\sigma_j}} \frac{d^3k}{\sqrt{2|\vec{k}|}} \{ \vec{\varepsilon}_{\vec{k},\lambda} a_{\vec{k},\lambda}^* e^{-i\vec{k} \cdot \alpha\vec{x}} + \vec{\varepsilon}_{\vec{k},\lambda}^* a_{\vec{k},\lambda} e^{i\vec{k} \cdot \alpha\vec{x}} \}, \end{aligned} \quad (27)$$

where $\mathcal{B}_{\sigma_j}$ is a ball of radius σ_j centered at the origin, and $\{\sigma_j := \Lambda\epsilon^j\}$ is a sequence of infrared cut-offs with $0 < \epsilon < 1$ and $j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$;

ii) The Fock spaces

$$\mathcal{F}_{\sigma_j} := \mathcal{F}(L^2((\mathbb{R}^3 \setminus \mathcal{B}_{\sigma_j}) \times \mathbb{Z}_2)) \quad , \quad \mathcal{F}_{\sigma_{j+1}}^{\sigma_j} := \mathcal{F}(L^2((\mathcal{B}_{\sigma_j} \setminus \mathcal{B}_{\sigma_{j+1}}) \times \mathbb{Z}_2)), \quad (28)$$

where Ω_f denotes the vacuum state in anyone of these Fock spaces.

Note that

$$\mathcal{F}_{\sigma_{j+1}} = \mathcal{F}_{\sigma_j} \otimes \mathcal{F}_{\sigma_{j+1}}^{\sigma_j}. \quad (29)$$

Starting from the Fock space $\mathcal{F}_{\sigma_j}$, one defines the scale-dependent Hilbert space

$$\mathcal{H}_j := \mathcal{H}_{\text{el}} \otimes \mathcal{F}_{\sigma_j}. \quad (30)$$

In next formulae, we use that any vector η in $\mathcal{F}_{\sigma_j}$ can be identified with the corresponding vector, $\eta \otimes \Omega_f$, in $\mathcal{F}_{\sigma_{j'}}$, $j' > j$. In the same way, we identify vectors $\phi \otimes \eta$ and $\phi \otimes (\eta \otimes \Omega_f)$ in $\mathcal{H}_j$ and in $\mathcal{H}_{j'}$, respectively;

iii) Momentum-slice interaction Hamiltonians defined by

$$\begin{aligned} \Delta H_{\text{at}}|_{\sigma_{j+1}}^{\sigma_j} &:= H_{\text{at}}^{\sigma_j} - H_{\text{at}}^{\sigma_{j+1}} \\ &= 2\alpha^{\frac{3}{2}} (-i\vec{\nabla}_{\vec{x}} + \alpha^{3/2} \vec{A}_\Lambda^{\sigma_j}(\alpha\vec{x})) \cdot \vec{A}|_{\sigma_{j+1}}^{\sigma_j}(\alpha\vec{x}) + \alpha^3 (\vec{A}|_{\sigma_{j+1}}^{\sigma_j}(\alpha\vec{x}))^2, \end{aligned} \quad (31)$$

where

$$\begin{aligned} & \vec{A}|_{\sigma_{j+1}}^{\sigma_j}(\alpha\vec{x}) \\ &:= \frac{1}{(2\pi)^{3/2}} \sum_\lambda \int_{\mathcal{B}_{\sigma_j} \setminus \mathcal{B}_{\sigma_{j+1}}} \frac{d^3k}{\sqrt{2|\vec{k}|}} \{ \vec{\varepsilon}_{\vec{k},\lambda} a_{\vec{k},\lambda}^* e^{-i\vec{k} \cdot \alpha\vec{x}} + \vec{\varepsilon}_{\vec{k},\lambda}^* a_{\vec{k},\lambda} e^{i\vec{k} \cdot \alpha\vec{x}} \}; \end{aligned} \quad (32)$$

iv) Three real parameters, ϵ , ρ^- , and μ which have the properties

$$0 < \rho^- < \mu < 1, \quad 0 < \epsilon < \rho^- < \frac{E_1 - E_0}{\Lambda}. \quad (33)$$

By analytic perturbation theory (see [4]), one starts from the groundstate $\Psi^{\sigma_0} \equiv \phi_C \otimes \Omega_f$ of the Hamiltonian $H_{\text{at}}^{\sigma_0} \equiv H_{\text{at}}^{(0)}$, where ϕ_C is the groundstate of the Coulomb Hamiltonian ($H_{\text{Coul}} := -\Delta_{\vec{x}} - \frac{1}{|\vec{x}|}$), and derives the following results, valid for sufficiently small α , depending on the choice of $\Lambda, \epsilon, \rho^-$ and μ (see also Figure 2 below):

(A1) E^{σ_j} is an isolated simple eigenvalue of $H_{\text{at}}^{\sigma_j}|_{\mathcal{H}_j}$ (i.e., $H_{\text{at}}^{\sigma_j}$ restricted to $\mathcal{H}_j$) with spectral gap larger or equal to $\rho^- \sigma_j$. Furthermore, E^{σ_j} is also the groundstate energy of $H_{\text{at}}^{\sigma_j}|_{\mathcal{H}_{j+1}}$, and it is an isolated simple eigenvalue of $H_{\text{at}}^{\sigma_j}|_{\mathcal{H}_{j+1}}$ with spectral gap larger or equal to σ_{j+1} ;

- (A2) The groundstate energies E^{σ_j} and $E^{\sigma_{j+1}}$ of the Hamiltonians $H_{\text{at}}^{\sigma_j}$ and $H_{\text{at}}^{\sigma_{j+1}}$, respectively, (acting on the same space $\mathcal{H}_{j+1}$) satisfy the inequalities

$$E^{\sigma_{j+1}} \leq E^{\sigma_j} + c \alpha^3 \sigma_j^2, \quad (34)$$

where c is a j - and α -independent positive constant;

- (A3) The (unnormalized) groundstate vectors, $\Psi^{\sigma_{j+1}}$, of $H_{\text{at}}^{\sigma_{j+1}}$ can be recursively constructed starting from $\Psi^{\sigma_0} \equiv \phi_C \otimes \Omega_f$ with the help of the spectral projection

$$\frac{1}{2\pi i} \oint_{\gamma_{j+1}} dz \frac{1}{H_{\text{at}}^{\sigma_{j+1}} - z}, \quad (35)$$

where $\gamma_{j+1} := \{z \in \mathbb{C} \mid |z - E^{\sigma_j}| = \mu \sigma_{j+1}\}$, with μ as in (33). More precisely,

$$\Psi^{\sigma_{j+1}} := \frac{1}{2\pi i} \oint_{\gamma_{j+1}} dz \frac{1}{H_{\text{at}}^{\sigma_{j+1}} - z} \Psi^{\sigma_j} \quad (36)$$

$$= \frac{1}{2\pi i} \sum_{n=0}^{\infty} \oint_{\gamma_{j+1}} dz \frac{1}{H_{\text{at}}^{\sigma_j} - z} [-\Delta H_{\text{at}}|_{\sigma_{j+1}}^{\sigma_j} \frac{1}{H_{\text{at}}^{\sigma_j} - z}]^n \Psi^{\sigma_j}, \quad (37)$$

where one uses that Ψ^{σ_j} is the (unnormalized) groundstate vector of $H_{\text{at}}^{\sigma_j}|_{\mathcal{H}_j}$, for any $\infty \geq j' \geq j$ ($\mathcal{H}_{\infty} := \mathcal{H}$), in particular for $j' = j + 1$.

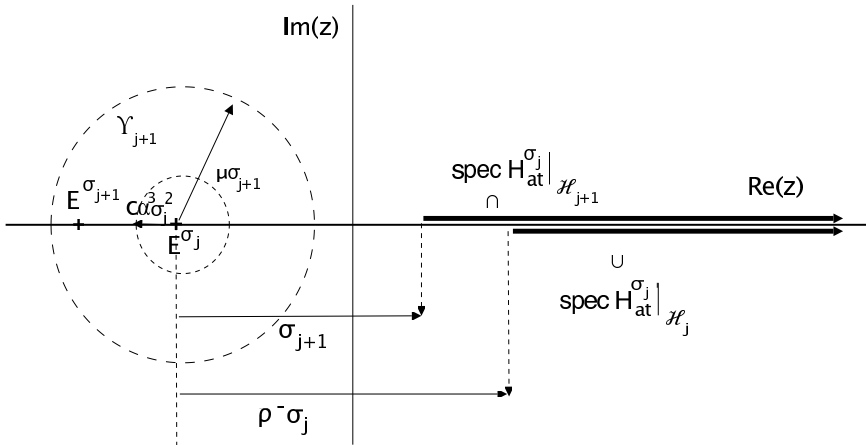


Fig. 2. The contour integral in the energy plane and the gaps.

The construction of the sequence $\{\Psi^{\sigma_j} : j \in \mathbb{N}_0\}$ only requires the estimate

$$\sup_{z \in \gamma_{j+1}} \left\| \frac{1}{H_{\text{at}}^{\sigma_j} - z} [-\Delta H_{\text{at}}|_{\sigma_{j+1}}^{\sigma_j}] \right\|_{\mathcal{H}_{j+1}} \leq \mathcal{O}(\alpha^{\frac{3}{2}}) \quad (38)$$

where we use the notation $\|A\|_{\mathcal{H}'} = \|A|_{\mathcal{H}'}\|$ for the norm of a bounded operator A acting on a Hilbert space $\mathcal{H}'$. For the proof of convergence of the sequence, one has to use:

a) The identity

$$-i\vec{\nabla}_{\vec{x}} + \alpha^{\frac{3}{2}} \bar{A}_{\Lambda}^{\sigma_j}(\alpha\vec{x}) = \frac{i[H_{\text{at}}^{\sigma_j}, \vec{x}]}{2}; \quad (39)$$

b) The localization of the electron near the nucleus that holds in any state of the system with energy below the ionization threshold, in particular in the groundstate; thus

$$\|\vec{x} \Psi^{\sigma_j}\| < \infty \quad \forall j \in \mathbb{N}_0. \quad (40)$$

In fact, these two additional ingredients are enough to conclude that

$$\|\Psi^{\sigma_{j+1}} - \Psi^{\sigma_j}\| \leq \mathcal{O}(\alpha^{3/2} \sigma_j). \quad (41)$$

Consequently, for α small but independent of j , there exists a nonzero vector $\Psi \equiv \lim_{j \rightarrow \infty} \Psi^{\sigma_j}$, and, as shown in [4], $\phi_{\text{gs}} := \frac{\Psi}{\|\Psi\|}$ is the unique groundstate of H_{at} .

Appendix B

The analysis of Bloch and Nordsieck, [9], suggests that the infrared behavior of the state describing the soft photons accompanying an electron should be singular (i.e., not square-integrable at the origin in photon momentum space), and that it should be determined by the momentum of the asymptotic electron. In mathematical terms, this means that the asymptotic electron velocity is expected to determine an asymptotic Weyl operator (creating a cloud of asymptotic photons), which when applied to an *improper* dressed one-electron state $\psi^{\sigma=0}(h)$ yields a well defined vector in the Hilbert space $\mathcal{H}$. This vector is expected to describe an asymptotic electron, with wave function h , surrounded by a cloud of infinitely many asymptotic free photons, in accordance with the observations sketched in (16) – (24).

In [14], the infraparticle scattering states are constructed by using a *time-dependent* approach to scattering theory. Within this approach, one has to construct a time-dependent approximating vector converging, as $t \rightarrow \infty$, to a vector in $\mathcal{H}$ that describes an asymptotic electron with wave function h whose momentum space support is contained in $\mathcal{S} \subset \{\vec{P} : |\vec{P}| < m\}$, and a cloud of asymptotic free photons with an upper photon energy κ , where $\kappa > 0$ is a counter threshold that can be arbitrarily small. This interpretation is justified a posteriori.

The task is to give a mathematically rigorous meaning to the formal expression

$$\Phi_{\kappa}^{\text{out}}(h) := \lim_{t \rightarrow \infty} \lim_{\sigma \rightarrow 0} e^{iH_{\text{el}} t} \mathcal{W}_{\kappa, \sigma}(\vec{v}(t), t) e^{-iH_{\text{el}}^{\sigma} t} \psi^{\sigma}(h), \quad (42)$$

where H_{el}^{σ} is the Hamiltonian introduced in i-1), Section 2.2.1, and

$$\mathcal{W}_{\kappa, \sigma}(\vec{v}(t), t) := \exp \left(\alpha^{\frac{1}{2}} \int_{\mathcal{B}_{\kappa} \setminus \mathcal{B}_{\sigma}} \frac{d^3 k}{\sqrt{2|\vec{k}|}} \frac{\vec{v}(t) \cdot \{\vec{\varepsilon}_{\vec{k}, \lambda} a_{\vec{k}, \lambda}^* e^{-i|\vec{k}|t} - \vec{\varepsilon}_{\vec{k}, \lambda}^* a_{\vec{k}, \lambda} e^{i|\vec{k}|t}\}}{(2\pi)^{3/2} |\vec{k}| (1 - \hat{k} \cdot \vec{v}(t))} \right). \quad (43)$$

The operator $\vec{v}(t)$ is not known a priori; but, in the limit $t \rightarrow \infty$, it must converge to the asymptotic velocity operator of the electron. The latter is determined by the operator $\vec{\nabla} E_{\vec{P}}$, applied to the *improper* vectors $\{\psi^{\sigma=0}(h)\}$; see Eq. (15). This can be seen by first considering the infrared regularized model (i.e, with $\sigma > 0$) which has dressed electron states $\{\psi^\sigma(h)\}$ in $\mathcal{H}$ with energy function $E^\sigma(\vec{P})$ (i.e., $H_{el}^\sigma \psi^\sigma(h) = E^\sigma(\vec{P}) \psi^\sigma(h)$), and by subsequently passing to the limit $\sigma \rightarrow 0$. Formally, for $\sigma \rightarrow 0$, the Weyl operator

$$e^{iH_{el}t} \mathcal{W}_{\sigma,\kappa}(\vec{v}(t), t) e^{-iH_{el}^\sigma t} \quad (44)$$

is an interpolating operator used in the L.S.Z. (Lehmann-Symanzik-Zimmermann) approach to scattering theory for the electromagnetic field. In the definition of this operator the photon test functions are evolved backwards in time with the free evolution, and the photon creation- and annihilation operators are evolved forward in time with the interacting time evolution.

The Weyl operator $\mathcal{W}_{\kappa,\sigma}(\vec{v}(t), t)$ is expressed in terms of the operators $\{a_{\vec{k},\lambda}, a_{\vec{k},\lambda}^*\}$, as it must be when describing real photons in a scattering process, and hence does *not* preserve the fiber spaces $\mathcal{H}_{\vec{P}}$.

The conjecture behind the definition of the vector on the R.H.S. of Eq. (42) is as follows.

- The time-dependent vector converges to an asymptotic vector of the form

$$\lim_{t \rightarrow \pm\infty} \lim_{\sigma \rightarrow 0} e^{iH_{el}t} \mathcal{W}_{\kappa,\sigma}(\vec{v}(t), t) e^{-iH_{el}^\sigma t} \psi^\sigma(h) = \mathcal{W}_{\kappa,\sigma=0}^{out/in}(\vec{v}(\pm\infty)) \psi^{\sigma=0}(h), \quad (45)$$

where

$$\begin{aligned} & \mathcal{W}_{\kappa,\sigma=0}^{out/in}(\vec{v}(\pm\infty)) \\ &:= \exp \left(\alpha^{\frac{1}{2}} \sum_{\lambda} \int_{B_\kappa} \frac{d^3k}{\sqrt{2|\vec{k}|}} \frac{\vec{v}(\pm\infty) \cdot \{\vec{\varepsilon}_{\vec{k},\lambda} a_{\vec{k},\lambda}^{out/in*} - \vec{\varepsilon}_{\vec{k},\lambda}^* a_{\vec{k},\lambda}^{out/in}\}}{(2\pi)^{3/2} |\vec{k}| (1 - \widehat{k} \cdot \vec{v}(\pm\infty))} \right), \end{aligned} \quad (46)$$

and $a_{\vec{k},\lambda}^{out/in*}$, $a_{\vec{k},\lambda}^{out/in}$ are the creation- and annihilation operators of the asymptotic photons.

- The operators $\vec{v}(\pm\infty)$ commute with the algebra of asymptotic creation- and annihilation operators $\{a_{\vec{k},\lambda}^{out/in*}, a_{\vec{k},\lambda}^{out/in}\}$; (this can be expected to be a consequence of asymptotic decoupling of the photon dynamics from the dynamics of the electron).
- The restriction of the asymptotic velocity operators, $\vec{v}(\pm\infty)$, to the improper dressed electron state is given by the operator $\vec{\nabla} E_{\vec{P}}$, i.e.,

$$\vec{v}(\pm\infty) \Psi_{\vec{P}} \equiv \vec{\nabla} E_{\vec{P}} \Psi_{\vec{P}}. \quad (47)$$

Starting from the expected relation (see (47)) between $\vec{v}(t)$ and $\vec{\nabla} E_{\vec{P}}^\sigma$, as $t \rightarrow \infty$ and $\sigma \rightarrow 0$, two key ideas used to make (45) precise are to render the infrared cut-off *time-dependent*, with $\sigma_t \rightarrow 0$ as $t \rightarrow \infty$, and to discretize the region $\mathcal{S}$, with a grid

size decreasing in time t . This discretization also applies to the velocity operator $\vec{v}(t)$ in expression (45).

The existence of infraparticle scattering states in $\mathcal{H}$ is established by proving that the corresponding sequence of time-dependent approximating states, which depend on the cutoff σ_t and on the discretization, defines a norm convergent sequence of vectors in $\mathcal{H}$. This is accomplished by appropriately tuning the convergence rates of σ_t and of the discretization of $\mathcal{S}$. The precise definition of the approximating vectors includes a phase factor with a role similar to the Coulomb phase in Coulomb scattering. For all the details of this construction, we refer the reader to [14]. Starting from the infraparticle scattering states $\{\Phi_\kappa^{out/in}(h)\}$, one can construct scattering subspaces $\mathcal{H}^{out/in}$ by adding *hard* (asymptotic) photons. Here below, we summarize the main properties of the limiting vectors, $\{\Phi_\kappa^{out/in}(h)\}$, that confirm the starting conjecture and the classical picture outlined in Section 2.2.2.

- On the infraparticle scattering states the asymptotic velocity of the electron, $\vec{v}_{out/in}$, is defined starting from

$$\begin{aligned} \lim_{t \rightarrow \pm\infty} e^{iH_{el}t} f\left(\frac{\vec{x}}{t}\right) e^{-iH_{el}t} \Phi_\kappa^{out/in}(h) &= \Phi_\kappa^{out/in}(f(\vec{\nabla} E^{\sigma=0})h) \\ &=: f(\vec{v}_{out/in}) \Phi_\kappa^{out/in}(h) \end{aligned} \quad (48)$$

for any smooth function f .

- The interpolating L.S.Z. electromagnetic fields (defined by expressions analogous to (10)) converge on the states $\{\Phi_\kappa^{out/in}(h)\}$. The representation of the asymptotic fields is coherent non-Fock, namely

$$a_{\vec{k},\lambda}^{out/in} \sim_{\vec{k} \rightarrow 0} a_{\vec{k},\lambda} + \alpha^{\frac{1}{2}} \frac{\vec{v}_{out/in} \cdot \vec{\epsilon}_{\vec{k},\lambda}}{(2\pi)^{3/2} \sqrt{2} |\vec{k}|^{\frac{3}{2}} (1 - \widehat{\vec{k}} \cdot \vec{v}_{out/in})}. \quad (49)$$

This result coincides with the prediction by Block and Nordsieck in the nonrelativistic regime.

- Consider the limits

$$\lim_{t \rightarrow \pm\infty} (\Phi_\kappa^{out/in}(h), \int d^3z e^{iH_{el}t} \vec{E}(\vec{x} + \vec{z}) e^{-iH_{el}t} \rho(\vec{z} - \vec{d}) \Phi_\kappa^{out/in}(h)), \quad (50)$$

$$\lim_{t \rightarrow \pm\infty} (\Phi_\kappa^{out/in}(h), \int d^3z e^{iH_{el}t} \vec{B}(\vec{x} + \vec{z}) e^{-iH_{el}t} \rho(\vec{z} - \vec{d}) \Phi_\kappa^{out/in}(h)), \quad (51)$$

where $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields, respectively, and ρ is a smooth delta function. For $|\vec{d}| \rightarrow \infty$, the limits above coincide with the superposition of the Liénard-Wiechert fields associated with the spectral values of the velocity $\vec{v}_{out/in}$, that are related to $\vec{P}$ in the support of the electron wave function h through the relation in Eq. (47).

Acknowledgments

The author was supported by NSF grant DMS-0905988.

References

- [1] W.K. Abou Salem, J. Faupin, J. Fröhlich, and I.M. Sigal On the theory of resonances in non-relativistic quantum electrodynamics and related models *Advances in Applied Mathematics* **43** (3) 201-230 (2009)
- [2] V. Bach, T. Chen, J. Fröhlich, and I.M. Sigal. Smooth Feshbach map and operator-theoretic renormalization group methods. *J. Funct. Anal.*, **203**(1), 44–92 (2003).
- [3] V. Bach, T. Chen, J. Fröhlich, and I.M. Sigal. The renormalized electron mass in Non-Relativistic Quantum Electrodynamics . *J. Funct. Anal.*, **243**(2), 426–535 (2007).
- [4] V. Bach, J. Fröhlich, and A. Pizzo. Infrared-Finite Algorithms in QED. The Ground-state of an Atom Interacting with the Quantized Radiation Field. *Comm. Math. Phys.*, **264**(1), 145–165 (2006).
- [5] V. Bach, J. Fröhlich, and A. Pizzo. Infrared-Finite Algorithms in QED. The Expansion in the Coupling Constant of the Groundstate of an atom interacting with the Quantized radiation field. *Adv. Math.* **220** (4), 1023-1074 (2009)
- [6] V. Bach, J. Fröhlich, and A. Pizzo. An Infrared-Finite Algorithm for Rayleigh Scattering Amplitudes, and Bohr’s Frequency Condition. *Comm. Math. Phys.*, **274**(2), 457–486 (2007)
- [7] V. Bach, J. Fröhlich, and I. M. Sigal. Spectral analysis for systems of atoms and molecules coupled to the quantized radiation field. *Comm. Math. Phys.*, **207**(2), 249–290 (1999).
- [8] V. Bach, J. Fröhlich, I. M. Sigal, and A. Soffer. Positive commutators and the spectrum of Pauli-Fierz Hamiltonian of atoms and molecules. *Commun. Math. Phys.*, **207**(3), 557–587 (1999).
- [9] F. Bloch and A. Nordsieck. Note on the Radiation Field of the Electron. *Phys. Rev.*, **52**, **54** (1937).
- [10] D. Buchholz. Collision theory for massless bosons. *Comm. Math. Phys.*, **52**, 147–173 (1977).
- [11] D. Buchholz. Gauss’ law and the infraparticle problem. *Phys. Lett. B*, **174**, 331–334 (1986).
- [12] T. Chen. Infrared renormalization in non-relativistic QED and scaling criticality. *J. Funct. Anal.*, **254**(10), 2555–2647 (2008).
- [13] T. Chen and J. Fröhlich. Coherent infrared representations in nonrelativistic QED. *Spectral Theory and Mathematical Physics: A Festschrift in Honor of Barry Simon’s 60th Birthday Proc. Symp. Pure Math.* AMS, 2007.
- [14] T. Chen, J. Fröhlich and A. Pizzo. Infraparticle Scattering States in QED: II. Mass Shell properties. *J. Math. Phys.* **50** 012103 (2009) <http://link.aip.org/link/?JMAPAQ/50/012103/1>
- [15] T. Chen, J. Fröhlich and A. Pizzo. Infraparticle Scattering States in QED: I. The Bloch-Nordsieck Paradigm. *To appear in Comm. Math. Phys.* http://xxx.lanl.gov/PS_cache/arxiv/pdf/0709/0709.2493v2.pdf
- [16] V. Chung. Infrared Divergence in Quantum Electrodynamics *Phys. Rev.*, **140B** 1110–1122 (1965).
- [17] W. De Roeck, J. Fröhlich, and A. Pizzo *in preparation*
- [18] J. Dereziński and V. Jaksic. Spectral theory of Pauli-Fierz operators. *J. Func. Anal.*, **180**(2), 243–327 (2001).
- [19] J. Dereziński and C. Gerard. Asymptotic completeness in quantum field theory. Massive Pauli-Fierz Hamiltonians. *Rev. Math. Phys.*, **11**(4), 383–450 (1999).
- [20] L. Faddeev and P. Kulish. Asymptotic conditions and infrared divergences in quantum electrodynamics. *Theor. Math. Phys.*, **5**, 153–170 (1970).
- [21] J. Fröhlich. On the infrared problem in a model of scalar electrons and massless,

- scalar bosons. *Inst. Henri Poincaré, Section Physique Théorique*, 19(1):1–103, 1973.
- [22] J. Fröhlich. Existence of dressed one electron states in a class of persistent models. *Fortschritte der Physik*, **22**, 159–198 (1974).
 - [23] J. Fröhlich, M. Griesemer, and B. Schlein. Asymptotic completeness for Rayleigh scattering. *Ann. H. Poincaré*, **3**, 107–170 (2002).
 - [24] J. Fröhlich, M. Griesemer, and I.M. Sigal. Spectral Theory for the Standard Model of Non-Relativistic QED. *Comm. Math. Phys.* **283** (3), 613–646 (2008)
 - [25] J. Fröhlich, M. Griesemer, and I.M. Sigal. Spectral Renormalization Group and Local Decay in the Standard Model of the Non-relativistic Quantum Electrodynamics *arXiv:0904.1014*
 - [26] J. Fröhlich, G. Morchio, F. Strocchi. Charged sectors and scattering state in quantum electrodynamics *Annals of Physics*, **119** (2), June 1979
 - [27] J. Fröhlich and A. Pizzo. On the Absence of Excited Eigenstates in QED. *Comm. Math. Phys.* **286**, (3) <http://dx.doi.org/10.1007/s00220-008-0704-1> (2009)
 - [28] J. Fröhlich and A. Pizzo. The renormalized electron mass in non-relativistic QED. *To appear in Comm. Math. Phys.* mp-arc 08-131. http://www.ma.utexas.edu/mp_arc/
 - [29] M. Griesemer. Exponential decay and ionization threshold in non-relativistic quantum electrodynamics. *J. Funct. Anal.*, **210**(3), 321–340 (2004).
 - [30] M. Griesemer, D. Hasler Analytic Perturbation Theory and Renormalization Analysis of Matter Coupled to Quantized Radiation. *Ann. H. Poincaré*, **10** (3) 577–621 (2009)
 - [31] M. Griesemer, E. Lieb, and M. Loss. Ground states in non-relativistic quantum electrodynamics. *Invent. math.*, **145**(3), 557–595 (2001).
 - [32] C. Hainzl, M. Hirokawa, and H. Spohn. Binding energy for hydrogen-like atoms in the Nelson model without cutoffs. *Journal of Funct. Anal.*, 220(2):424–459 (2005)
 - [33] M. Hubner and H. Spohn. Spectral model of the spin-boson Hamiltonian. *Ann. Inst. H. Poincaré Phys. Theor.*, **62**(3), 289–323 (1995).
 - [34] J.M. Jauch, F. Rohrlich. *Helv. Phys. Acta* **27**, 613 (1954).
 - [35] J.M. Jauch, F. Rohrlich. Theory of photons and electrons. *Addison-Wesley*.
 - [36] T. Kibble. Coherent Soft-Photon States and Infrared Divergences. I. Classical Currents *J. Math. Phys.* **9**, 315, (1968).
 - [37] G. Morchio, F. Strocchi. Infrared singularities, vacuum structure and pure phases in local quantum field theory. *Ann. Inst. H. Poincaré Sect. A (N.S.)* **33**, no. 3, 251–282 (1980).
 - [38] E. Nelson. Interaction of nonrelativistic particles with a quantized scalar field. *J. Math. Phys.*, **5** 1190–1197, 1964.
 - [39] M. Fierz and W. Pauli. *Nuovo. Cim.*, **15** 167 (1938).
 - [40] A. Pizzo. One Particle (improper) States in Nelson’s Massless Model. *Ann. H. Poincaré*, **4**(3), 439–486 (2003).
 - [41] A. Pizzo. Scattering of an Infraparticle: The One Particle Sector in Nelson’s Massless Model. *Ann. H. Poincaré*, **6**, 553–606 (2005).
 - [42] F. Rohrlich. Infrared Divergence in Bound State Problems. *Phys. Rev.* **98**, 181–182 (1955).
 - [43] I.M. Sigal. Ground States and Resonances in the Standard Model of the Non-Relativistic QED *arXiv:0806.3297*
 - [44] B. Schroer. Infrateilchen in der Quantenfeldtheorie. (German) *Fortschr. Physik* **11**, 1–31 (1963).
 - [45] M. Reed, B. Simon. Methods of modern mathematical physics. Vol. I – IV *Academic Press*.

- [46] F. Strocchi, A. S. Wightman. Proof of the charge superselection rule in local relativistic quantum field theory. *J. Math. Phys.* **15**, 2198–2224 (1974).
- [47] D. Yennie, S. Frautschi, and H. Suura. The Infrared Divergences Phenomena and the High-Energy Processes. *Annals of Physics*, **13** 375, 1961

PARKING IN THE CITY: AN EXAMPLE OF LIMITED RESOURCE SHARING

PETR ŠEBA

*University of Hradec Králové
Hradec Králové – Czech Republic
E-mail: seba@fzu.cz*

During the attempt to park a car in the city the drivers have to share limited resources (the available roadside). We show that this fact leads to a predictable distribution of the distances between the cars that depends on the length of the street segment used for the collective parking. We demonstrate in addition that the individual parking maneuver is guided by generic psychophysical perceptual correlates. Both predictions are compared with the actual parking data collected in the city of Hradec Králové (Czech Republic).

Keywords: Transport systems

1. Introduction

Everyone knows that to park a car in the city center is problematic. The amount of the available places is limited and has to be shared between an increasing number of cars. There are several attempts to tackle this challenge by introducing parking charges, building underground garages, implementing parking zones, advertising for public transport and many others. The problem seems however to persist. Deeper understanding of the related processes is therefore of a common interest.

There have been several mathematical attempts to tackle the parallel parking process. The classical way to do so is the “random car parking model” introduced by Rényi [1] – see [2], [3] for review. In this model the cars park on randomly chosen places and once parked the cars do not leave the street. All cars are usually assumed to be of the same length l_0 and the process continues as long as all available free places are smaller than l_0 . The model leads to predictions that can be easily verified. First of all it gives a relation between the mean bumper-to-bumper distance $\overline{D}$ and the car length: $\overline{D} \sim 0.337 l_0$. Further: the probability density $p(D)$ of the car distances D behaves like [4], [5], [6], [7] $p(D) \approx -\ln(D)$. This means that small distances between cars are preferred.

To test this results real parking data were collected recently in the center of London [8]. The average distance between the parked cars was 152 cm which fits nicely with the relation $\overline{D} \sim 0.337 l_0$ for $l_0 = 450$ cm. The predicted probability density was however incompatible with the observed facts. The model leads to

$p(D) \rightarrow \infty$ for $D \rightarrow 0$ whereby the data from London display $p(D) \rightarrow 0$ as $D \rightarrow 0$. The same behavior has been found also in other cities [9].

This results show that the parking process is not so simple as assumed by the “random car parking model”. First of all: although the total available parking place remains unchanged the cars are reshuffled many times during the day since some parked cars leave and new cars park on the vacant places. Moreover the parking maneuver is not trivial. It is not just a simple positioning of the car to the parking lot.

Our aim here is to give an alternative approach to the parking process. We will show that it can be understood as a statistical partition of the limited parking space between competing persons trying to park. The partition is described by the Dirichlet distribution with a parameter g . We will also show that this parameter is in fact fixed by the capability of the driver to exploit small distances during the parking maneuver.

The paper is organized as follows: In the Section 2 we describe the mathematical model and derive a one parameter family of possible car clearance distributions. Section 3 contains the psychophysical background for positioning the car inside the parking lot. It is based on the estimated time to collision with the neighboring car. Its behavior on small distances picks exactly one distribution from the above mentioned one parameter family. The Section 4 contains finally the comparison of the theory with data collected for actually parked cars.

2. The parking process

We focus on the spacing distribution (bumper to bumper distances) between cars parked parallel to the curb. We will assume that the street segment used for parking starts and ends with some clear and nondisplaceable part unsuitable for parking. It can be a driveway or turning to a side street. Otherwise the parking segment is free of any kind of parking obstructions. We will assume that it has a length L . Moreover there are not marked parking lots or park meters inside it. So the drivers are free to park the car anywhere in the segment provided they find an empty space to do it. We suppose also that all cars have the same length l_0 .

Many cars are cruising for parking in this part of the city. So there are not free parking lots and a car can park only when another parked car leaves. To simplify the further formulation of the problem and to avoid troubling with the boundary effects we assume that the street segment under consideration form a circle. The car spacing distribution is obtained as a steady solution of the repeated car parking and car leaving process.

Due to the parking maneuver one needs a lot of a length $\approx 1.3l_0$ to park. Hence in a segment of length L the number of the parked cars equals to $N \approx [L/(1.3l_0)]$. Denoting by D_k the spacing between the car k and $k+1$ we get $\sum_{k=1}^N D_k = L - Nl_0$ and after a simple rescaling finally

$$\sum_{k=1}^N D_k = 1. \quad (1)$$

Since all parking lots are occupied the number of parked cars is supposed to be fixed. The repeated car parking and car leaving reshuffles however the distances D_k . We will treat them as independent random variables constrained by the simplex (1). The distance reshuffling goes as follows: In the first step one randomly chosen car leaves the street and the two adjoining lots merge into a single one. In the second step a new car parks into this empty space and splits it again into two smaller lots. Such fragmentation and coagulation processes were discussed intensively since they apply for instance to the computer memory allocation - see [10] for review. The related equations are simple. If a car leaves the street and the neighboring spacings - say the spacings D_n, D_{n+1} - merge into a single lot D we get

$$D = D_n + D_{n+1} + l_0. \quad (2)$$

When a new car parks to D it splits it into $\tilde{D}_n, \tilde{D}_{n+1}$:

$$\begin{aligned} \tilde{D}_n &= a(D - l_0) \\ \tilde{D}_{n+1} &= (1 - a)(D - l_0). \end{aligned} \quad (3)$$

where $a \in (0, 1)$ is a random variable with a probability density $q(a)$. The distribution $q(a)$ describes the parking preference of the driver. We assume that all drivers have identical preferences, i.e. identical $q(a)$. (The meaning of the variable a is straightforward. For $a = 0$ the car parks immediately in front of the car delimiting the parking lot from the left without leaving any empty space. For $a = 1/2$ it parks exactly to the center of the lot D and for $a = 1$ it stops exactly behind the car on the right.) Combining (2) and (3) gives the distance reshuffling

$$\begin{aligned} \tilde{D}_n &= a(D_n + D_{n+1}) \\ \tilde{D}_{n+1} &= (1 - a)(D_n + D_{n+1}). \end{aligned} \quad (4)$$

(The car length l_0 drops out.) The simplex (1) is of course invariant under this transformation.

The mappings (4) are regarded as statistically independent for various choices of n . Moreover all cars are equal. So in the steady situation the joint distance probability density $P(D_1, \dots, D_N)$ has to be exchangeable (i.e. invariant under the permutation of variables) and invariant with respect to (4). Its marginals $p(D)$ (the probability densities of the particular spacings) are identical:

$$p_k(D_k) = p(D_k) = \int_{D_1 + \dots + D_N = 1} P(D_1, \dots, D_N) dD_1 \dots dD_{k-1} dD_{k+1} \dots dD_N. \quad (5)$$

A standard approach to deal with the simplex (1) is to regard D_k as independent random variables normalized by a sum:

$$D_k = \frac{d_k}{\sum_{n=1}^N d_n}. \quad (6)$$

Here d_k are statistically independent and identically distributed and it is preferable to work with them. Moreover: it is obvious that the distribution of $\{D_1, \dots, D_N\}$ is invariant under the transform (4) merely when the distribution of $\{d_1, \dots, d_N\}$ is invariant. So let us apply the relation (4) on the variables d_n :

$$\begin{aligned}\tilde{d}_n &= a(d_n + d_{n+1}) \\ \tilde{d}_{n+1} &= (1-a)(d_n + d_{n+1})\end{aligned}\quad (7)$$

where a, d_n, d_{n+1} are independent and d_n, d_{n+1} identically distributed. Giving the distribution $q(a)$ of a we look for distributions of d_n such that the transformed variables $\tilde{d}_n, \tilde{d}_{n+1}$ preserve the distribution of d_n . The effort is to solve the equation

$$d \triangleq a(d + d') \quad (8)$$

where d' is an independent copy of the variable d and the symbol $\triangleq$ means that the left and right sides of (8) have identical statistical properties.

Distributional equations of this type are mathematically well studied – see for instance [12] – although not much is known about their exact solutions. In particular it is known that for a given distribution $q(a)$ (describing the parking habit) the equation (8) has an unique solution which can be obtained numerically. Since we are interested in explicit results we choose $q(a)$ from a two parametric class of β distributions. Then the solution of (8) results from the following statement [11]:

Statement: Let d_1, d_2 and a be independent random variables with distributions: $d_1 \sim \Gamma(a_1, 1)$, $d_2 \sim \Gamma(a_2, 1)$ and $a \sim \beta(a_1, a_2)$. Then $a(d_1 + d_2) \sim \Gamma(a_1, 1)$.

(The symbol $\sim$ means that the related random variable has the specified probability density. $\Gamma(g, 1), \beta(g_1, g_2)$ denote the standard gamma and beta distributions respectively.)

Since in our case the variables d_1, d_2 are equally distributed we have $g_1 = g_2 = g$ and $a \sim \beta(g, g)$. So the probability density of a is symmetric in this case, i.e. the variables a and $1-a$ have the same distribution. In other words the drivers are not biased to park more closely to a car adjacent from the behind or from the front.

The solution of (8) is in this case equal to $d \sim \Gamma(g, 1)$. The relation (6) returns the spacings D_k and we find that the joint probability density $P(D_1, \dots, D_N)$ is nothing but a one parameter family of the multivariate Dirichlet distributions on the simplex (1) [13]:

$$P(D_1, \dots, D_N) = \frac{\Gamma(Ng)}{\Gamma(g)^N} D_1^{g-1} D_2^{g-1} \dots D_N^{g-1} \quad (9)$$

Its marginal (5) is simply $D \sim \beta(g, (N-1)g)$. Normalizing the mean of D to 1 we are finally left with

$$p(D) = \frac{1}{N} \beta\left(g, (N-1)g, \frac{D}{N}\right). \quad (10)$$

Despite of the symmetrical parking maneuver this distribution is asymmetric. This is a consequence of the persistent parked car exchange and can be regarded as a collective phenomenon.

To summarize: for the parking maneuver described by $a \sim \beta(g, g)$ we were able to find a joint probability distribution of the car distances $P(D_1, \dots, D_N)$ that is invariant under the parked car reshuffling (4). It is given by the multivariate Dirichlet distribution (9). The normalized bumper to bumper distribution is described by the formula (10). Before proceeding to the next section let us stress that the formula (10) depends on the parameters g and N . The dependence on g is obvious since it reflects the parking maneuver. The sensitivity to N is, however, more subtle. It just says that the bumper to bumper distance statistics depends also on the length of the parking segment under consideration. (Note that all parking lots are occupied. So the total number N of the parked cars is simply proportional to the length L of the parking segment.) We will show later that this dependence is really observed when collecting data from places with different street lengths.

3. The parking maneuver

In this section we intend to study the parking maneuver more closely. It is described by the probability density $q(a)$ of the random variable a and defines the interchange of the parked cars by the relation (4). To ensure the solvability we assumed that this variable is β distributed: $q(a) = \beta(g, g, a)$, where g is a free parameter. We will show that there is in fact a natural choice of the parameter g leading to $g = 3$.

The point is that the behavior of $q(a)$ for small a reflects the capability of the driver to estimate small distances. The collision avoidance during the parking maneuver is guided visually and this ability is shared equally by all drivers. If it applies the behavior of $q(a)$ for small a has to be generic, i.e. independent on the particular city or parking situation. It is just fixed by the human perception of distance.

Distance perception is a complex task and there are several cues to do this. Some of them are monocular (linear perspective, monocular movement parallax etc.), others oculomotor (accommodation convergence) and finally binocular (i.e. based on stereopsis). All of them work simultaneously and are reliable under different conditions - see [14] for more details. For the parking maneuver however the crucial information is not the distance itself but the estimated time-to-collision between the bumper of the parking car and its neighbors. This time has to be evaluated using the knowledge of the distance and velocity. It has been argued in a seminal paper by Lee [15] that the estimated time to collision is psychophysically evaluated using a quantity named τ . It is defined as the inverse of the relative rate of expansion of the retinal image of the moving object. Behavioral experiments have indicated that τ is indeed controlling actions like contacting surfaces by flies, birds and mammals (including humans): see [16], [17], [18].

When the observer moves forward in the environment, the image on the retina expands. The rate of the expansion τ conveys information about the observer's speed and the time to collision. Psychophysical and physiological studies have provided abundant evidence that τ is processed by specialized neural mechanisms in the

brain [19]. We take τ to be the informative variable for the final braking - see [20] for review.

Let θ be the instantaneous angular size of the observed object (for instance the front of the car we are backing to during the parking maneuver). Then the estimated time to contact is given by

$$\tau = \frac{\theta}{d\theta/dt} \quad (11)$$

Since $\theta(t) = 2 \arctan(L_0/2D(t))$ with L_0 being the width of the approached object and $D(t)$ its instantaneous distance, we get

$$\tau(t) = -\frac{L_0^2 + 4D(t)^2}{2L_0(dD(t)/dt)} \arctan\left(\frac{L_0}{2D(t)}\right). \quad (12)$$

For $D \gg L_0$ and a constant approach speed $v = -dD/dt$ the quantity τ simply equals to the physical arrival time: $\tau = D/v$. For small distances (parking maneuver), however, $\tau \approx D^2/(vL_0)$ and the estimated time to contact decreases quadratically with the distance.

Let us return to the equation (4). For a fixed parking lot the final stopping distance is proportional to a . Assuming that the courage to exploit small distances is proportional to the estimated time to contact we finally get for the probability density $q(a)$: $q(a) \approx a^2$ for small a . Since $q(a) = \beta(g, g, a)$ the behavior $q(a) \approx a^2$ fix the parameter g to $g = 3$ and the normalized clearance distribution (10) reads

$$p(D) = \frac{1}{N} \beta\left(3, 3(N-1), \frac{D}{N}\right) = \left(\frac{1}{N}\right)^{3(N-1)} \frac{\Gamma(3N)}{2\Gamma(3(N-1))} D^2 (N-D)^{3N-4} \quad (13)$$

4. The measured data

We have argued in the above section that the parameter g in (10) is fixed to 3. But the distribution still depend on the number of cars N in the parking segment. It is a consequence of the constrain (1). For large number of cars, $N \gg 1$, the constrain (1) does not play a substantial role and $p(D)$ equals to $\Gamma(3, 1, D)$ (this is true in the limit $N \rightarrow \infty$). In the other extreme case with $N = 1$ (the parking segment is so short that it allows the parking of a single vehicle) the distribution $p(D)$ just reflects the parking maneuver and is equal to $q(D)$. For parking segments of intermediate size the theory predicts a dependence of the results on the segment length.

To verify the predictions of the model we measured the bumper to bumper distances between cars parked on two different streets in the center of Hradec Králové (Czech Republic). Both streets were located in a place with large parking demand and usually without any free parking lots. In addition one of these street (street 1) contained driveways to courtyards. This means that the actually available fixed parking segments were much shorter on this street. In the mean 3-4 cars were able to park among two subsequent driveways and we collected 773 car spacings under

this conditions. The second street (street 2) was free of any dividing elements. Here we measured altogether 699 spacings.

The probability distributions resulting from these data seems to be fairly compatible with the prediction of the model. First of all: the perceptual mechanism based on the estimated time to contact seems to be verified. We demonstrated that when the estimated time to contact is decisive for the final car stopping then the parameter g in (10) fix to $g = 3$. And exactly this value fits with the measured data. Moreover: the finite length of the street (the simplex (1)) leads to a dependence of the spacing distribution (13) on N . So the result obtained for short and long parking segments should be different. And this is indeed observed when the data from the street 1 and 2 are compared. We plot the results on the figure 1.

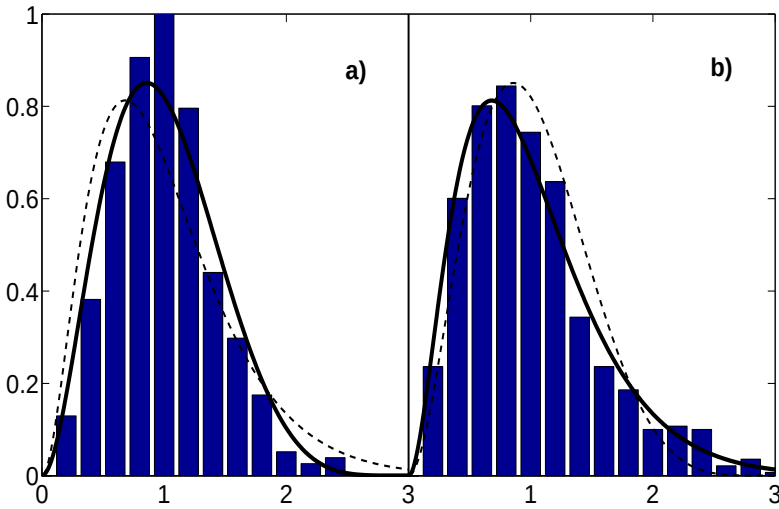


Fig. 1. The measured distance distributions for cars parking on the the street 1 and 2 in the city of Hradec Králové are compared with the prediction of the formula (13). The results for the street 1 and 2 are plotted on the panels a) and b) respectively. Bars mark the probability density extracted from the collected data. The full line stays for the formula (13) with $N = 3$ (panel a)) and $N=20$ (panel b)). To guide the eyes the result of (13) for $N = 20$ and $N = 3$ are plotted on the panels a) and b) as a dashed line.

The difference between the results is not large but it is nevertheless clearly visible (compare the full and dashed lines).

The distance estimation and parking problems are not restricted to human. In fact birds face the same problem when a flock tries to perch on an electric line. The available space is limited and has to be shared. Moreover the τ - mechanism works identically for man and for birds. As a consequence we expect that the spacing

distribution for the cars parked in parallel and for the birds perching on a power line should be the same. To demonstrate this fact we measured the clearance for cars parked in parallel and photographed the birds perching on a power line. In both cases the “parking segment” was full, i.e. there was not a free space for an additional participant. The segments under consideration were long and contained a large number of objects. So the constrain (1) can be omitted and the solvable variant of the model gives $p(D) = \Gamma(3, 1/3, D)$. We measured the bumper to bumper distances between the cars parked in the center of Hradec Králové (Czech Republic) - altogether 700 spacings. In the case of the birds we photographed flocks of starlings resting on the power lines during their flight to the south. The chosen line was “full”. Other starlings from the flock were forced to use the space vacated by another starling or to use another line to perch. The bird-to-bird distances were obtained by the photograph digitalization - altogether 1000 spacings. Both data sets were scaled to the mean distance equal to 1 and the results plotted on the Figure 2. The

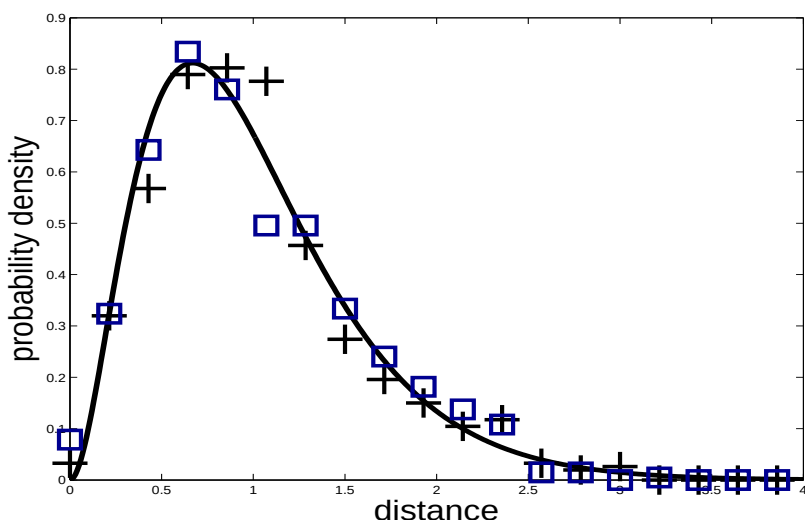


Fig. 2. The probability density of the distances between the parked cars (crosses) and perching starlings (squares) is compared with the prediction (13) of the theory (full line). The mean distance is normalized to 1.

probability distributions resulting from the two data sets seems to be identical and in a good agreement with the model prediction. This fact is interesting since the used “hardware” is fully different. The underlying psychophysical mechanism, is, however, identical.

Acknowledgement: The research was supported by the Czech Ministry of Education within the project LC06002. I appreciate the stimulating discussions with Balint Virag. The help of the PhD students Michal Musilek and Jan Fator,

who collected the parking data, is also gratefully acknowledged.

References

- [1] Renyi A.: Publ. Math. Inst. Hung. Acad. Sci. 3 (1958) 109.
- [2] Evans J.W.: Rev.Mod.Phys. 65 (4) (1993) 1281-1330
- [3] Cadilhe A., Araujo N.A.M. and Privman V.: J.Phys. Cond. Mat. 19 (2007) 065124
- [4] Araujo N.A.M., Coadilhe A.: Phys. Rev. E 73 (2006) 051602
- [5] D Orsogna M.R., Chou T.: J.Phys.A. 38 (2005) 531-542
- [6] Yang X.F., Knowles K.M.: J.Am.Ceram.Soc. 75 (1992) 141-147
- [7] Mackenzie J.K.: J.Chem.Phys. 37 (1962) 723-728
- [8] Rawal S., Rodgers G.J.: Physica A 246 (2005) 621-630
- [9] Seba P.: J.Phys.A 41 (2008) 122003
- [10] Bertoin J.: Random Fragmentation and Coagulation Processes. Cambridge University Press, Cambridge, 2006.
- [11] Dufresne D.: Adv. Appl. Math. 20 (1998) 285-299
- [12] Devroye L. and Neininger R.: Advances of Applied Probability, vol. 34 (2002) 441-468.
- [13] Wilks, S.S.: Mathematical Statistics. John Wiley & Sons, New York
- [14] Jacobs R.A.: Trends in Cognitive Sciences Vol.6 No.8 (2002) 345
- [15] Lee, D. N.: A theory of visual control of braking based on information about time-to-collision. Perception 5 (1976), 437-459.
- [16] van der Weel F.R., van der Meer L.H., Lee N.D.: Human Movement Science 15 (1996) 253-283
- [17] Hopkins B., Churchill A., Vogt S., Ronnqvist L.: Journal of Motor Behavior 36, Number 3 (2004) 3 - 12
- [18] Schrater P.R., Knill D.C., Simoncelli E.P.: Nature 410 (2001) 816
- [19] Farrow K., Haag J. and Borst A.: Nature Neuroscience 9 (2006) 1312 - 1320
- [20] Fajen B.R.: Journal of Experimental Psychology 31, No. 3 (2005) 480-501

HOT TOPICS IN COLD GASES

ROBERT SEIRINGER

*Department of Physics, Princeton University,
Princeton NJ 08544, USA
E-mail: rseiring@princeton.edu*

Since the first experimental realization of Bose-Einstein condensation in cold atomic gases in 1995 there has been a surge of activity in this field. Ingenious experiments have allowed us to probe matter close to zero temperature and reveal some of the fascinating effects quantum mechanics has bestowed on nature. It is a challenge for mathematical physicists to understand these various phenomena from first principles, that is, starting from the underlying many-body Schrödinger equation. Recent progress in this direction concerns mainly equilibrium properties of dilute, cold quantum gases. We shall explain some of the results in this article, and describe the mathematics involved in understanding these phenomena. Topics include the ground state energy and the free energy at positive temperature, the effect of interparticle interaction on the critical temperature for Bose-Einstein condensation, as well as the occurrence of superfluidity and quantized vortices in rapidly rotating gases.

Keywords: Bose-Einstein condensation; statistical mechanics; dilute gases; superfluidity.

1. Introduction

Bose-Einstein Condensation (BEC) was first experimentally realized in cold atomic gases in 1995 [2, 10]. In these experiments, a large number of (bosonic) atoms is confined to a trap and cooled to very low temperatures. Below a certain critical temperature condensation of a large fraction of particles into the same one-particle state occurs.

These Bose-Einstein condensates display various interesting quantum phenomena, like superfluidity and the appearance of quantized vortices in rotating traps, effective lower dimensional behavior in strongly elongated traps, etc. We refer to the review articles [5, 8, 9, 15] for an overview of the state-of-the-art of this subject and a list of references to the original literature.

BEC was predicted by Einstein in 1924 [13] from considerations of the non-interacting Bose gas, extending the work of Bose [7] to massive particles. The presence of particle interactions represents a major difficulty for a rigorous derivation of this phenomenon, however, as we shall discuss below.

1.1. The Bose Gas: A Quantum Many-Body Problem

The quantum-mechanical description of the Bose gas is given in terms of its Hamiltonian. For a gas of N bosons confined to a region $\Lambda \in \mathbb{R}^3$, and interacting via a repulsive pair-interaction potential v , it is given by

$$H = - \sum_{i=1}^N \Delta_i + \sum_{1 \leq i < j \leq N} v(\vec{x}_i - \vec{x}_j) \quad (1)$$

The kinetic energy of the particles is described by the Laplacian $\Delta = \vec{\nabla}^2$, and we choose Dirichlet boundary conditions on $\partial\Lambda$ for concreteness. Other boundary conditions could be used as well. The subscript i stands for the action in the i th particle coordinate $\vec{x}_i \in \mathbb{R}^3$. Units are chosen such that $\hbar = 2m = 1$, with m the mass of the particles.

The Hamiltonian H acts on the Hilbert space of permutation-symmetric wave functions $\Psi \in \bigotimes^N L^2(\mathbb{R}^3)$, as appropriate for bosons; i.e., square-integrable functions of N variables $\vec{x}_i \in \mathbb{R}^3$ satisfying

$$\Psi(\vec{x}_1, \dots, \vec{x}_N) = \Psi(\vec{x}_{\pi(1)}, \dots, \vec{x}_{\pi(N)})$$

for any permutation π of $(1, 2, \dots, N)$.

In the following, the interaction v will be assumed to be radial and non-negative. Moreover, it is sufficiently short range as to have a finite scattering length, which means that it is integrable outside some compact set. No other regularity assumptions will be made. In particular, v is allowed to have a hard core, which reduces the domain of definition of H to those functions Ψ that vanish whenever the distance between a pair of particle coordinates is smaller than the hard-sphere radius. A particular example of an interaction potential to keep in mind are pure hard spheres where, formally, $v(\vec{x}) = \infty$ for $|\vec{x}| \leq a$, and $v(\vec{x}) = 0$ for $|\vec{x}| > a$.

The present setup can be easily generalized to describe inhomogeneous systems in a trap. One simply adds a trap potential

$$\sum_{i=1}^N V(\vec{x}_i)$$

to H , where V is a real-valued, locally bounded function with $\lim_{|\vec{x}| \rightarrow \infty} V(\vec{x}) = \infty$. The latter condition guarantees that the particles are confined to the trap, even in case $\Lambda = \mathbb{R}^3$.

Similarly, rotating systems can be described by adding the term

$$\sum_{i=1}^N \vec{\Omega} \cdot \vec{L}_i$$

to the Hamiltonian H , with $\vec{\Omega} \in \mathbb{R}^3$ being the angular velocity and $\vec{L} = -i\vec{x} \wedge \vec{\nabla}$ the angular momentum operator. This term results from a transformation to the rotating frame of reference.

1.2. Quantities of Interest

In the following, we shall distinguish two types of questions that can be asked concerning the behavior of Bose gases described by the Hamiltonian (1) above.

- **Thermodynamic quantities**, like the ground state energy per unit volume, or the free energy density at positive temperature. Here one considers homogeneous systems and is interested in the thermodynamic limit $N \rightarrow \infty$, $\Lambda \rightarrow \mathbb{R}^3$ with the particle density $\varrho = N/|\Lambda|$ fixed.

Of particular interest is the notion of Bose-Einstein condensation, which concerns off-diagonal long-range order in the one-particle density matrix $\langle a^\dagger(x)a(y) \rangle$, and is expected to occur below a critical temperature.

- **Behavior of trapped systems** in the ground state. One observes interesting quantum phenomena, like effective one-dimensional behavior in strongly elongated traps, vortices in rotating systems, a bosonic analogue of the fractional quantum Hall effect in rapidly rotating gases, etc.

Of particular relevance is the **Gross-Pitaevskii** scaling, where the ratio of the scattering length a to the diameter of the trap is $O(N^{-1})$.

We shall discuss our current knowledge about answers to these questions, as far as mathematical physics is concerned, in the following sections.

2. Homogeneous Systems in the Thermodynamic Limit

2.1. The Ground State Energy of Homogeneous Bose Gases

Consider first the case of a homogeneous system in the absence of a trapping potential or rotation. The ground state energy density in the thermodynamic limit is given by

$$e(\varrho) = \lim_{\Lambda \rightarrow \mathbb{R}^3, N/|\Lambda| \rightarrow \varrho} \frac{1}{|\Lambda|} \inf \text{spec } H \quad (2)$$

with H as in (1). The existence of this thermodynamic limit is well understood for appropriate sequences of domains Λ approaching $\mathbb{R}^3$. See, e.g., Ruelle's book [34].

We will be particularly interested in the limit of low density, when the gas is dilute in the sense that $a^3\varrho \ll 1$, where a denotes the scattering length of the interaction potential v . It is defined as

$$4\pi a = \inf \left\{ \int_{\mathbb{R}^3} \left(|\vec{\nabla} \phi(|\vec{x}|)|^2 + \frac{1}{2} v(\vec{x}) \phi(|\vec{x}|)^2 \right) d\vec{x} : \phi \geq 0, \lim_{r \rightarrow \infty} \phi(r) = 1 \right\}. \quad (3)$$

For bosons at low density, one expects that

$$e(\varrho) \approx 4\pi a \varrho^2. \quad (4)$$

This formula is suggested by considering the ground state energy of two bosons in a large region Λ , which is $8\pi a/|\Lambda|$, as can be easily deduced from (3). Multiplying this by the number of pairs of bosons, $N(N-1)/2$, one arrives at (4). That this

simple heuristics is correct is far from obvious, however. It fails for two-dimensional systems, for instance [29, 35].

The investigation of the ground state energy density $e(\varrho)$ goes back to Bogoliubov [6] in the 40s, and Lee, Huang and Yang in the 50s [19]. Dyson [11] computed a rigorous upper bound that shows the correct leading order asymptotics (4) for hard spheres, but his lower bound was 14 times too small. His upper bound was later generalized to arbitrary repulsive interaction potentials in [24]. The correct lower bound was proved only in 1998 by Lieb and Yngvason [28]. We formulate this result as a theorem.

Theorem 2.1 (Bosons at $T = 0$). *As $\varrho \rightarrow 0$,*

$$\boxed{e(\varrho) = 4\pi a\varrho^2 + o(\varrho^2)} \quad (5)$$

Note that if one treats the interaction energy as a perturbation of the kinetic energy, naive perturbation theory would yield $\frac{1}{2} \int v$ instead of $4\pi a$. This is always too big, as (3) shows, and would even be infinite for hard spheres. In fact, the result (5) is non-perturbative in the sense that the scattering length a contains terms to arbitrary high order in the interaction potential v .

It remains an open problem to establish the leading order correction to (4), which is expected to be given by the Lee-Huang-Yang formula [19]

$$e(\varrho) \approx 4\pi a\varrho^2 \left(1 + \frac{128}{15\sqrt{\pi}} \sqrt{a^3\varrho} \right). \quad (6)$$

Recent progress in this direction was made in [17] and [27], where it was shown that (6) holds for certain density-dependent and appropriately scaled interaction potentials. The general question remains open, however.

2.2. Homogeneous Bose Gas at Positive Temperature

At positive temperature $T > 0$, the appropriate quantity to consider is the free energy density, which is defined as

$$f(\varrho, T) = -T \lim_{\Lambda \rightarrow \mathbb{R}^3, N/|\Lambda| \rightarrow \varrho} \frac{1}{|\Lambda|} \ln \text{Tr} \exp(-H/T). \quad (7)$$

For non-interacting bosons (i.e., $v \equiv 0$), it can be calculated explicitly. We denote it by $f_0(\varrho, T)$. It is given in terms of a Legendre transform as

$$f_0(\varrho, T) = \sup_{\mu < 0} \left[\mu\varrho + \frac{T}{(2\pi)^3} \int_{\mathbb{R}^3} d\vec{p} \ln (1 - \exp(-(\vec{p}^2 - \mu)/T)) \right]. \quad (8)$$

Note that it has the scaling property

$$f_0(\varrho, T) = \varrho^{5/3} f_0(1, T\varrho^{-2/3})$$

which follows from the fact that the only length scales in the problem are the mean particle spacing $\varrho^{-1/3}$ and the thermal wavelength $T^{-1/2}$, and hence f_0 depends, up to a prefactor, only on their ratio.

From (8) it is easy to see that f_0 is not an analytic function of ϱ (or T), and hence even a non-interacting Bose gas shows a phase transition. This transition is known as Bose-Einstein condensation, and occurs at a critical density

$$\varrho_c(T) = \zeta(\tfrac{3}{2}) \left(\frac{T}{4\pi} \right)^{3/2},$$

where ζ denotes the Riemann zeta-function. In fact, $\partial f_0(\varrho, T)/\partial \varrho = 0$ for $\varrho \geq \varrho_c$. For $\varrho > \varrho_c(T)$, $\varrho - \varrho_c(T)$ is interpreted as the density of the Bose-Einstein condensate.

For interacting gases, there are now three length scales to consider: the interaction range a , the mean particle distance $\varrho^{-1/3}$, and the thermal wavelength $T^{-1/2}$. For dilute systems, one considers the case

$$a \ll \varrho^{-1/3} \sim T^{-1/2}.$$

In this regime, the free energy turns out to be the given by the following expression.

Theorem 2.2 (Bosons at $T > 0$). *For $a^3 \varrho \ll 1$ we have*

$$f(\varrho, T) = f_0(\varrho, T) + 4\pi a \left(2\varrho^2 - [\varrho - \varrho_c(T)]_+^2 \right) + o(\varrho^2) \quad (9)$$

where $[t]_+ = \max\{t, 0\}$ denotes the positive part.

The lower bound in (9) was proved in [38]. An upper bound for smooth interacting potentials of rapid decay was later obtained in [40], the more general case being still open.

The error term in (9) is uniform in $T/\varrho^{2/3}$ for bounded $T/\varrho^{2/3}$, corresponding to the quantum regime. For $T/\varrho^{2/3} \rightarrow \infty$ one obtains a *classical* gas, whereas for $T/\varrho^{2/3} \rightarrow 0$ the system approaches the ground state.

Note that for $\varrho < \varrho_c(T)$, the leading order correction compared to the ideal Bose gas is $8\pi a \varrho^2$ instead of the $4\pi a \varrho^2$ at zero temperature. The additional factor 2 is a result of the symmetry requirements on the wave functions and can be interpreted as an exchange term; this symmetrization applies only to particles outside the condensate, however, and this explains the subtraction of the square of the condensate density in (9). We also remark that without restricting to symmetric functions, the leading order correction compared with an ideal gas would be $4\pi a \varrho^2$ at any $T > 0$, just like at $T = 0$.

The proof of Theorem 2.2 is long and technical and hence can not be reproduced here. One of the key issues to understand is a certain separation of energy scales in the two terms on the right side of (9). In momentum space, these are

- large momenta $|\vec{p}| \sim 1/a$ responsible for scattering of two particles at a distance $\sim a$ of each other.
- low momenta $|\vec{p}| \sim T^{1/2} \ll 1/a$, responsible for the thermal distribution distribution of the particles' kinetic energy.
- Bose-Einstein condensation at momentum $\vec{p} = 0$.

2.3. Critical Temperature for BEC

As discussed above, the ideal, non-interacting Bose gas displays a phase transition at a critical density. Equivalently, BEC in the ideal gas occurs below the critical temperature

$$T_c(\varrho) = \frac{4\pi}{\zeta(3/2)^{2/3}} \varrho^{2/3}.$$

A useful characterization of BEC, applicable also for interacting systems, is in terms of the one-particle density matrix of the system. This density matrix is defined as

$$\gamma = N \frac{1}{\text{Tr} e^{-H/T}} \text{Tr}^{(N-1)} e^{-H/T} \quad (10)$$

where Tr^{N-1} stands for the partial trace over $N - 1$ particle coordinates. Hence γ is an operator on the one-particle space $L^2(\mathbb{R}^3)$. Obviously $\gamma \geq 0$ and $\text{Tr} \gamma = N$, by definition. BEC is characterized by the fact that, in the thermodynamic limit, the integral kernel $\gamma(\vec{x}, \vec{y})$ of γ does not vanish as $|\vec{x} - \vec{y}| \rightarrow \infty$. This is also referred to as off-diagonal long range order. For non-interacting bosons, one can show that

$$\gamma(\vec{x}, \vec{y}) = [\varrho - \varrho_c(T)]_+ + \sum_{n \geq 0} \frac{e^{\bar{\mu}n/T}}{(4\pi n/T)^{3/2}} e^{-T|\vec{x} - \vec{y}|^2/(4n)}$$

in the thermodynamic limit, with $[t]_+ = \max\{t, 0\}$ denoting the positive part, and $\bar{\mu} \leq 0$ the μ where the maximum in (8) is achieved. Hence the kernel $\gamma(\vec{x}, \vec{y})$ has the following characteristics:

- For $T < T_c(\varrho)$, $\gamma(\vec{x}, \vec{y})$ does not decay. In fact, $\lim_{|\vec{x} - \vec{y}| \rightarrow \infty} \gamma(\vec{x}, \vec{y}) = \varrho - \varrho_c(T)$, the condensate density
- For $T > T_c(\varrho)$, $\gamma(\vec{x}, \vec{y})$ decays exponentially, like $e^{-\sqrt{-\bar{\mu}}|\vec{x} - \vec{y}|}$
- For $T = T_c(\varrho)$, $\gamma(\vec{x}, \vec{y})$ decays algebraically. In fact, $\gamma(\vec{x}, \vec{y}) \sim |\vec{x} - \vec{y}|^{-1}$ in this case.

These features are expected to hold also for interacting Bose gases, although with a different value of the critical temperature $T_c(\varrho)$. It is still an open problem to prove the existence of BEC for interacting gases, however. The only known case where BEC has been proved is the hard-core lattice gas at half filling [12], which is equivalent to the XY spin model. [31]

Although there is no proof that $T_c \neq 0$ in the interacting case, an upper bound can be derived rigorously [39]:

Theorem 2.3 (Upper bound on T_c). *For small $a^3\varrho$ and some $c > 0$,*

$$\boxed{\frac{T_c - T_c^{(0)}}{T_c^{(0)}} \leq c\sqrt{a\varrho^{1/3}}}$$

where $T_c^{(0)} = \frac{4\pi}{\zeta(3/2)^{2/3}} \varrho^{2/3}$ is the critical temperature for the ideal Bose gas.

More precisely, it is shown in [39] that $\gamma(\vec{x}, \vec{y})$ decays exponentially if $T > T_c^{(0)}(1 + c\sqrt{a\rho^{1/3}})$. The proof uses a well-known Feynman-Kac representation of the partition function in terms of integrals over paths and sums over cycles in permutations. [16]

There seems to be still no consensus in the physics literature concerning the correct power of the exponent of $a\rho^{1/3}$ in the shift in critical temperature, or even the sign of c ! Recent numerical simulations suggest that the shift should be linear in $a\rho^{1/3}$, with a positive c . This expected behavior of $T_c(\rho)$, as well as the upper bound of Theorem 2.3, are sketched in Figure 1.

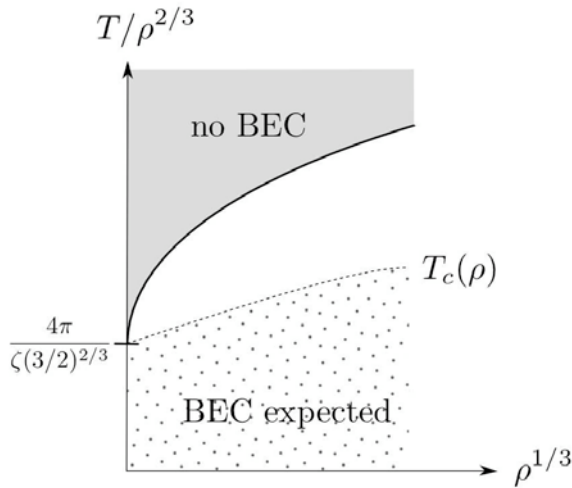


Fig. 1. The solid line shows the rigorous upper bound on the critical temperature for BEC. The dashed line corresponds to the expected behavior based on numerical simulations.

3. Trapped Bose Gases

In the previous chapter we considered homogeneous Bose gases in the thermodynamic limit. Recent experiments with cold atoms consider inhomogeneous gases in traps, however. That is, one can take Λ to be the whole of $\mathbb{R}^3$, but adds a trap potential $\sum_{i=1}^N V(\vec{x}_i)$ to the Hamiltonian (1). A typical example, which describes the experimental situation rather well, is a harmonic oscillator potential $V(\vec{x}) = \omega^2|\vec{x}|^2$, with $\omega > 0$ the trap frequency. More generally, the trapping frequencies in the three directions can be different, of course.

A characteristic feature of these trapped gases is their response to rotation. One observes the appearance of quantized vortices [14, 30], whose number increases with the rotation speed $|\vec{\Omega}|$.

Even a rotating Bose gas can be described in a time-independent way, by going to the rotating reference frame. The only effect on the Hamiltonian is to add the

term $\sum_{i=1}^N \vec{\Omega} \cdot \vec{L}_i$, as discussed in the Introduction. To ensure stability of the system, the trap potential V has to increase fast enough at infinity to compensate for the centrifugal force in the rotating system. More precisely, we have to assume that

$$\lim_{|\vec{x}| \rightarrow \infty} \left(V(\vec{x}) - \frac{1}{4} |\vec{\Omega} \wedge \vec{x}|^2 \right) = +\infty.$$

3.1. The Gross-Pitaevskii Equation

The previous considerations suggest that dilute Bose gases close to zero temperature should be well described by the Gross-Pitaevskii (GP) energy functional [18, 33]

$$\mathcal{E}^{\text{GP}}[\phi] = \left\langle \phi \left| -\Delta + V(\vec{x}) - \vec{\Omega} \cdot \vec{L} \right| \phi \right\rangle + 4\pi Na \int_{\mathbb{R}^3} |\phi(\vec{x})|^4 d\vec{x}. \quad (11)$$

Its ground state energy is

$$E^{\text{GP}}(Na, \vec{\Omega}) = \inf_{\|\phi\|_2=1} \mathcal{E}^{\text{GP}}[\phi],$$

and any minimizer satisfies the GP equation

$$-\Delta \phi(\vec{x}) + V(\vec{x})\phi(\vec{x}) - \vec{\Omega} \cdot \vec{L} \phi(\vec{x}) + 8\pi Na |\phi(\vec{x})|^2 \phi(\vec{x}) = \mu \phi(\vec{x}).$$

For a minimizer, $N|\phi(\vec{x})|^2$ is interpreted as the particle density of the system. Hence the last term in (11) is the natural generalization of the expression $4\pi a \rho^2$ to inhomogeneous systems.

For $\vec{\Omega} \neq 0$ and axially symmetric $V(\vec{x})$, the rotation symmetry can be broken due to the appearance of quantized vortices. More precisely, it was shown in [36, 37] that for all trap potentials $V(\vec{x})$ that grow faster than quadratically at infinity, there exists a $g_{\vec{\Omega}}$ such that for all $Na > g_{\vec{\Omega}}$ the GP minimizers necessarily are not axially symmetric. In particular, there are many (in fact, uncountably many) GP minimizers! The symmetry breaking is due to the appearance of quantized vortices which can not be arranged in a symmetric way. Many interesting results have been obtained concerning the nature and distribution of these vortices in GP minimizers. We refer to [1] and references therein.

3.2. Ground State Energy of Dilute Trapped Gases

In typical experiments on cold atomic gases, $N \gg 1$, $a \ll 1$ (the length scale of the trapping potential V), but $Na = O(1)$. To get to this dilute regime, one writes

$$v(\vec{x}) = \frac{1}{a^2} w(\vec{x}/a) \quad (12)$$

with w having scattering length 1. It is easy to see that $v(\vec{x})$ then has scattering length a . The scattering length thus becomes a parameter, and we can write

$$\inf \text{spec } H = E_0(N, a, \vec{\Omega}).$$

We note that the scaling (12) is of course equivalent to a rescaling of the trap potential V while keeping v fixed. This latter procedure may seem physically more natural (as the trap potential is easier to adjust experimentally than the interaction potential) but we find it more convenient to fix V instead and scale v as in (12). Our procedure corresponds to measuring all lengths in the system in units of the length scale of the trap potential.

For dilute systems, one expects that $E_0(N, a, \vec{\Omega}) \approx NE^{\text{GP}}(Na, \vec{\Omega})$. The proof of this fact was given in [23].

Theorem 3.1 (Ground State Energy of Trapped Gases). *For fixed $g \geq 0$ and $\vec{\Omega} \in \mathbb{R}^3$,*

$$\lim_{N \rightarrow \infty} \frac{E_0(N, g/N, \vec{\Omega})}{N} = E^{\text{GP}}(g, \vec{\Omega}) \quad (13)$$

This theorem was previously proved in [24] for the case $\vec{\Omega} = 0$. The main difficulty in the generalization to rotating systems comes from the fact that the permutation symmetry of the wave functions now becomes essential. While for non-rotating systems it is well known that the ground state for bosons coincides with the ground state without symmetry restrictions (as the latter is unique and positive, hence must be symmetric), this fact fails to hold for rotating systems. In fact one can show that (13) fails to hold, in general, if the left side is replaced by the absolute ground state energy of H (viewed as an operator on $L^2(\mathbb{R}^{3N})$, without symmetry restrictions) [37].

3.3. BEC for Rotating Trapped Gases

In the previous subsection it was argued that the ground state energy of the GP functional (11) is a good approximation to the ground state energy of H for dilute gases. For the corresponding ground state $\Psi_0(\vec{x}_1, \dots, \vec{x}_N)$, one would also expect that its one-particle reduced density matrix satisfies

$$\gamma_0 \equiv \text{Tr}^{(N-1)} |\Psi_0\rangle \langle \Psi_0| \approx |\phi\rangle \langle \phi| \quad (14)$$

with ϕ a minimizer of the GP functional. (For convenience, the normalization of γ_0 has been chosen differently here than we did previously in (10).) While this is indeed true in the non-rotating case [22], the rotating case is more complicated because of the non-uniqueness of the GP minimizers ϕ . The best one can hope for is to replace the right side of (14) by a convex combination of rank-one projections onto GP minimizers. This is indeed the content of Theorem 3.2 below, which was proved in [23].

To state the following results precisely, it is necessary to introduce the concept of an approximate ground state. We will call a sequence of N -particle density matrices (positive trace class operators on the N -particle space with trace equal to one) an approximate ground state if their energy equals the ground state energy to leading

order in N . Then we define the set Γ as the set of limit points of one-particle density matrices of such approximate ground states. More precisely,

$$\Gamma = \left\{ \gamma : \exists \text{ sequence } \gamma_N, \lim_{N \rightarrow \infty, Na \rightarrow g} \frac{1}{N} \text{Tr } H \gamma_N = E^{\text{GP}}(g, \vec{\Omega}), \lim_{N \rightarrow \infty} \gamma_N^{(1)} = \gamma \right\} \quad (15)$$

where $\gamma_N^{(1)} = \text{Tr}^{(N-1)} \gamma_N$ denotes the one-particle density matrix of γ_N .

Theorem 3.2 (BEC for Dilute Trapped Gases). *The set Γ in (15) has the following properties.*

- (i) $\Gamma \subset \mathcal{J}_1$ is compact and convex.
- (ii) The extreme points $\Gamma_{\text{ext}} \subset \Gamma$ are given by GP minimizers, i.e., $\Gamma_{\text{ext}} = \{ |\phi\rangle\langle\phi| : \mathcal{E}^{\text{GP}}[\phi] = E^{\text{GP}}(g, \vec{\Omega}) \}$.
- (iii) For every $\gamma \in \Gamma$ there exists a positive (regular Borel) measure $d\mu_\gamma$, supported in Γ_{ext} with $\int_{\Gamma_{\text{ext}}} d\mu_\gamma(\phi) = 1$, such that

$$\gamma = \int_{\Gamma_{\text{ext}}} d\mu_\gamma(\phi) |\phi\rangle\langle\phi|. \quad (16)$$

Eq. (16) is the natural generalization of (14) to the case of multiple GP minimizers. It says that the one-particle density matrix of any approximate ground state is close (in trace class norm) to the convex combination of projections onto GP minimizers.

Theorem 3.2 represents also a proof of the spontaneous breaking of the rotation symmetry in rotating Bose gases. An infinitesimal perturbation, e.g. of the trap potential V , leads to a unique GP minimizer and hence to 100% condensation, since the set Γ consists of only one element in this case. The quantized vortices are visible in the GP minimizer; they are a typical feature of superfluids. Theorem 3.2 can therefore also be interpreted as a proof of the superfluid behavior of rotating Bose gases [25].

3.4. Rapid Rotation

Consider now the special case of a harmonic trapping potential

$$V(\vec{x}) = \frac{1}{4} |\vec{x}|^2.$$

As discussed above, H is bounded below only for $|\vec{\Omega}| \leq 1$. The results in the previous subsections are valid for fixed $|\vec{\Omega}| < 1$. The question we would like to address in this final section is what happens as $|\vec{\Omega}| \rightarrow 1$? Denoting $\vec{e}_\Omega = \vec{\Omega}/|\vec{\Omega}|$ the unit vector in the direction of $\vec{\Omega}$, we can write

$$-\Delta + \frac{1}{4} |\vec{x}|^2 - \vec{\Omega} \cdot \vec{L} = \underbrace{\left(-i\vec{\nabla} - \frac{1}{2} \vec{e}_\Omega \wedge \vec{x} \right)^2 + \frac{1}{4} |\vec{e}_\Omega \cdot \vec{x}|^2}_{h} + (\vec{e}_\Omega - \vec{\Omega}) \cdot \vec{L}. \quad (17)$$

The operator h has eigenvalues $\frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \dots$, each of which is infinitely degenerate.

For low energies it makes sense to restrict the allowed wave functions to the kernel of $h - \frac{3}{2}$. This kernel is given by the Bargmann space [3]

$$\{f(z)e^{-|\vec{x}|^2/4}, f: \mathbb{C} \rightarrow \mathbb{C} \text{ analytic}\} \subset L^2(\mathbb{R}^3) \quad (18)$$

where we identify the complex variable z with the plane perpendicular to $\vec{\Omega}$. In particular, $|\vec{x}|^2 = |z|^2 + |\vec{e}_\Omega \cdot \vec{x}|^2$. Since the Gaussian factor is fixed, it is convenient to absorb it into the measure and think of the Hilbert space as a space of analytic functions only. It can easily be checked that the angular momentum operator $\vec{e}_\Omega \cdot \vec{L}$ acts on f as $z\partial_z$. In particular, its eigenfunctions are z^n , with eigenvalues $n \in \{0, 1, 2, \dots\}$.

We note that if one interprets $\vec{e}_\Omega$ as a homogeneous magnetic field, the Bargmann space (18) corresponds to the lowest Landau level in the perpendicular direction, multiplied by a fixed Gaussian in the longitudinal direction.

Now that we have identified the Bargmann space (18) as the appropriate one-particle Hilbert space for rapidly rotating bosons, we have to come up with an effective Hamiltonian describing this system. The only term left in the one-particle energy (17), except for a trivial factor $\frac{3}{2}$, is the angular momentum term $(\vec{e}_\Omega - \vec{\Omega}) \cdot \vec{L}$. If the range of the interaction potential is much shorter than the “magnetic length” 1, it makes sense to approximate the interaction potential by a δ -function, which becomes a bounded operator when projected to the Bargmann space. Writing the prefactor of the δ -function as $8\pi a$, in accordance with previous considerations, we arrive at the effective Hamiltonian

$$H^{\text{LLL}} := (1 - |\vec{\Omega}|) \sum_{i=1}^N z_i \partial_{z_i} + 8\pi a \sum_{1 \leq i < j \leq N} \delta_{ij}. \quad (19)$$

It acts on the space of permutation-symmetric analytic functions $f(z_1, \dots, z_N)$ which are square-integrable with respect to the measure $\prod_{i=1}^N e^{-|z_i|^2/4} dz_i$. We denote this space by $\mathcal{B}^{\otimes N}$. The operator $\delta_{12} : \mathcal{B}^{\otimes 2} \rightarrow \mathcal{B}^{\otimes 2}$ acts as

$$(\delta_{12}f)(z_1, z_2) = \frac{1}{(2\pi)^{3/2}} f\left(\frac{1}{2}(z_1 + z_2), \frac{1}{2}(z_1 + z_2)\right)$$

which takes analytic functions into analytic functions. It is obtained by projecting $\delta(\vec{x}_1 - \vec{x}_2)$ onto $\mathcal{B}^{\otimes 2}$.

Concerning the effective Hamiltonian (19), the following questions arise naturally:

- (1) Can one derive H^{LLL} rigorously from the full Hamiltonian H as $|\vec{\Omega}| \rightarrow 1$ and $a \rightarrow 0$? Such a rigorous derivation was indeed achieved in [21], where it was shown that if $|\vec{\Omega}| \rightarrow 1$ for fixed N and fixed $a/(1 - |\vec{\Omega}|)$ (i.e., also $a \rightarrow 0$), then the ratio of the ground state energy of H , minus the trivial term $\frac{3}{2}N$, to the ground state energy of H^{LLL} goes to 1. Similarly, one can show that also eigenfunctions converge in the same limit. Uniformity in the particle number N is still an open problem, however.

- (2) What are the properties of H^{LLL} , in particular concerning its spectrum and corresponding eigenfunctions? Certain features of H^{LLL} are expected to show some similarities to the fractional quantum Hall effect which occurs in fermionic systems.

Concerning the latter question, let us first note that the two terms $\mathcal{L}_N$ and Δ_N in H^{LLL} commute:

$$H^{\text{LLL}} = (1 - |\vec{\Omega}|) \underbrace{\sum_{i=1}^N z_i \partial_{z_i}}_{\mathcal{L}_N} + 8\pi a \underbrace{\sum_{1 \leq i < j \leq N} \delta_{ij}}_{\Delta_N}.$$

Hence the ground state energy $E_0^{\text{LLL}}(N, a, \vec{\Omega}) = \inf \text{spec } H^{\text{LLL}}$ is obtained from the joint spectrum of these two operators. Of particular relevance is the yrast curve, which is defined as the lowest eigenvalue of Δ_N in the sector of total angular momentum L :

$$\Delta_N(L) = \inf \text{spec } \Delta_N \upharpoonright_{\mathcal{L}_N=L}.$$

It is explicitly known for $L \leq N$ [4, 20, 32] (see also [21] for a simple proof)

$$\Delta_N(L) = (2\pi)^{-3/2} \begin{cases} \frac{1}{2}N(N-1) & \text{for } L \in \{0, 1\} \\ \frac{1}{2}N(N-1-\frac{1}{2}L) & \text{for } 2 \leq L \leq N. \end{cases}$$

Moreover, $\Delta_N(L) = 0$ for $L \geq N(N-1)$. The eigenfunction for $L = N(N-1)$ corresponding to the eigenvalue 0 of Δ_N is the bosonic Laughlin wave function

$$\prod_{1 \leq i < j \leq N} (z_i - z_j)^2.$$

Little is known about $\Delta_N(L)$ for $N < L < N(N-1)$, except for numerical simulations for small particle number. The only rigorous result concerns the limit $N \gg 1$ and $L \ll N^2$ where one can show that the Gross-Pitaevskii approximation is exact [26]. I.e., in this regime the convex hull of $\Delta_N(L)$ is given by

$$\inf \left\{ \frac{1}{2} \langle f \otimes f | \delta_{12} | f \otimes f \rangle : f \in \mathcal{B}, \|f\|^2 = N, \langle f | z \partial_z | f \rangle = L \right\}.$$

The qualitative behavior of $\Delta_N(L)$ is sketched in Figure 2.

Acknowledgments

This article was written during a visit at the Erwin-Schrödinger Institute in Vienna, whose hospitality is gratefully acknowledged. This work was partially supported by the U.S. National Science Foundation under grant No. PHY-0652356.

References

- [1] A. Aftalion, *Vortices in Bose-Einstein condensates*, Progress in Nonlinear Differential Equations and their Applications **67**, Birkhäuser (2006).

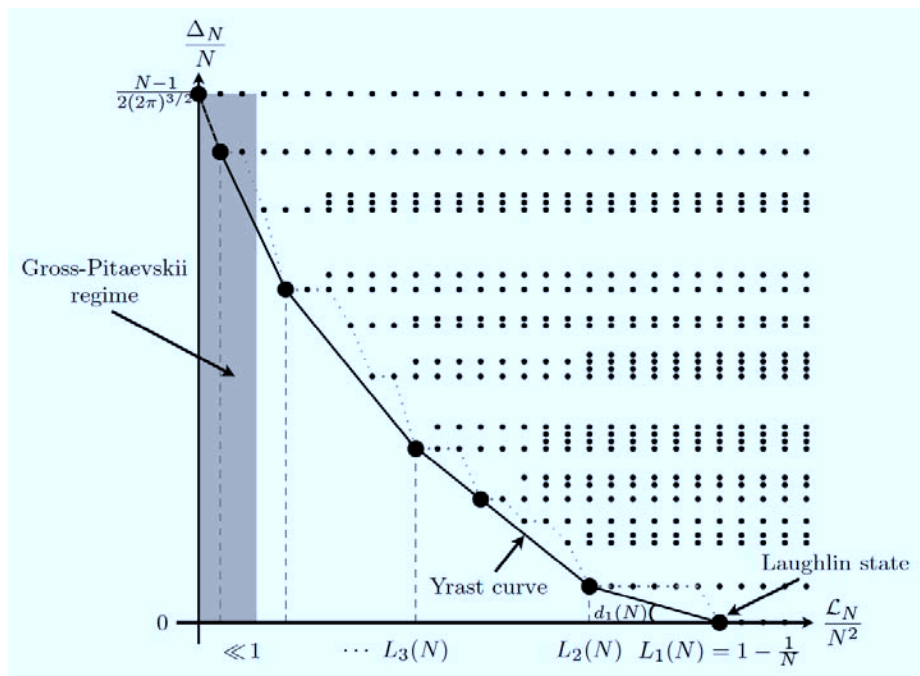


Fig. 2. (taken from [21]) A sketch of the joint spectrum of $\mathcal{L}_N$ and Δ_N . The dotted line is the Yrast curve, the solid line its convex hull. The bold dots correspond to the possible ground states of H^{LLL} as one varies $(1 - |\vec{\Omega}|)/a$. The grey area on the left shows the validity regime of the Gross-Pitaevskii equation. For $L \geq N(N-1)$, the interaction energy is zero.

- [2] M.H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, E.A. Cornell, *Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor*, Science **269**, 198–201 (1995).
- [3] V. Bargmann, *On the Representations of the Rotation Group*, Rev. Mod. Phys. **34**, 829 (1962).
- [4] G. Bertsch, T. Papenbrock, *Yrast line for weakly interacting trapped bosons*, Phys. Rev. Lett. **83**, 5412–5414 (1999).
- [5] I. Bloch, J. Dalibard, W. Zwerger, *Many-body physics with ultracold gases*, Rev. Mod. Phys. **80**, 885 (2008).
- [6] N.N. Bogoliubov, *On the theory of superfluidity*, Izv. Akad. Nauk USSR, **11**, 77 (1947). Eng. Trans. J. Phys. (USSR), **11**, 23 (1947). See also *Lectures on quantum statistics*, vol. 1, Gordon and Breach (1967).
- [7] S.N. Bose, *Plancks Gesetz und Lichtquantenhypothese*, Z. Phys. **26**, 178–181 (1924).
- [8] N.R. Cooper, *Rapidly rotating atomic gases*, Adv. Phys. **57**, 539–616 (2008).
- [9] F. Dalfovo, S. Giorgini, L.P. Pitaevskii, S. Stringari, *Theory of Bose-Einstein condensation in trapped gases*, Rev. Mod. Phys. **71**, 463–512 (1999).
- [10] K.B. Davis, M.O. Mewes, M.R. Andrews, N.J. van Druten, D.S. Durfee, D.M. Kurn, W. Ketterle, *Bose-Einstein Condensation in a Gas of Sodium Atoms*, Phys. Rev. Lett. **75**, 3969–3973 (1995).
- [11] F.J. Dyson, *Ground-State Energy of a Hard-Sphere Gas*, Phys. Rev. **106**, 20–26 (1957).

- [12] F.J. Dyson, E.H. Lieb, B. Simon, *Phase Transitions in Quantum Spin Systems with Isotropic and Nonisotropic Interactions*, J. Stat. Phys. **18**, 335–383 (1978).
- [13] A. Einstein, *Quantentheorie des einatomigen idealen Gases*, Sitzber. Kgl. Preuss. Akad. Wiss., 261–267 (1924), and 3–14 (1925).
- [14] P. Engels, I. Coddington, P.C. Haljan, V. Schweikhard, E.A. Cornell, *Observation of Long-Lived Vortex Aggregates in Rapidly Rotating Bose-Einstein Condensates*, Phys. Rev. Lett. **90**, 170405 (2003).
- [15] A.L. Fetter, *Rotating trapped Bose-Einstein condensates*, Rev. Mod. Phys. **81**, 647 (2009).
- [16] J. Ginibre, *Some Applications of Functional Integration in Statistical Mechanics*, in: Mécanique Statistique et Théorie Quantique des Champs, Les Houches, 1970, Gordon and Breach, New York, 327–427 (1971).
- [17] A. Giuliani, R. Seiringer, *The Ground State Energy of the Weakly Interacting Bose Gas at High Density*, J. Stat. Phys. **135**, 915–934 (2009).
- [18] E.P. Gross, *Structure of a Quantized Vortex in Boson Systems*, Nuovo Cimento **20**, 454–466 (1961). *Hydrodynamics of a superfluid condensate*, J. Math. Phys. **4**, 195–207 (1963).
- [19] K. Huang, C.N. Yang, *Quantum-Mechanical Many-Body Problem with Hard-Sphere Interaction*, Phys. Rev. **105**, 767–775 (1957); T.D. Lee, K. Huang, C.N. Yang, *Eigenvalues and Eigenfunctions of a Bose System of Hard Spheres and Its Low-Temperature Properties*, Phys. Rev. **106**, 1135–1145 (1957).
- [20] M. Hussein, O. Vorov, *Generalized yrast states of a Bose-Einstein condensate in a harmonic trap for a universality class of interactions*, Phys. Rev. A **65**, 035603 (2002).
- [21] M. Lewin, R. Seiringer, *Strongly correlated phases in rapidly rotating Bose gases*, preprint arXiv:0906.0741, J. Stat. Phys. (in press).
- [22] E.H. Lieb, R. Seiringer, *Proof of Bose-Einstein Condensation for Dilute Trapped Gases*, Phys. Rev. Lett. **88**, 170409-1–4 (2002).
- [23] E.H. Lieb, R. Seiringer, *Derivation of the Gross-Pitaevskii Equation for Rotating Bose Gases*, Commun. Math. Phys. **264**, 505–537 (2006).
- [24] E.H. Lieb, R. Seiringer, J. Yngvason, *Bosons in a trap: A rigorous derivation of the Gross-Pitaevskii energy functional*, Phys. Rev. A **61**, 043602-1–13 (2000).
- [25] E.H. Lieb, R. Seiringer, J. Yngvason, *Superfluidity in dilute trapped Bose gases*, Phys. Rev. B **66**, 134529-1–6 (2002).
- [26] E.H. Lieb, R. Seiringer, J. Yngvason, *Yrast line of a rapidly rotating Bose gas: Gross-Pitaevskii regime*, Phys. Rev. A **79**, 063626 (2009).
- [27] E.H. Lieb, J.P. Solovej, in preparation.
- [28] E.H. Lieb, J. Yngvason, *Ground State Energy of the low density Bose Gas*, Phys. Rev. Lett. **80**, 2504–2507 (1998).
- [29] E.H. Lieb, J. Yngvason, *The Ground State Energy of a Dilute Two-dimensional Bose Gas*, J. Stat. Phys. **103**, 509 (2001).
- [30] K.W. Madison, F. Chevy, W. Wohlleben, J. Dalibard, *Vortex Formation in a Stirred Bose-Einstein Condensate*, Phys. Rev. Lett. **84**, 806–809 (2000).
- [31] T. Matsubara, H. Matsuda, *A lattice model of liquid helium*, Prog. Theor. Phys. **16**, 569–582 (1956).
- [32] T. Papenbrock, G.F. Bertsch, *Rotational spectra of weakly interacting Bose-Einstein condensates*, Phys. Rev. A **63**, 023616 (2001).
- [33] L.P. Pitaevskii, *Vortex lines in an imperfect Bose gas*, Sov. Phys. JETP. **13**, 451–454 (1961).
- [34] D. Ruelle, *Statistical Mechanics. Rigorous Results*, World Scientific (1999).
- [35] M. Schick, *Two-Dimensional System of Hard Core Bosons*, Phys. Rev. A **3**, 1067–1073

(1971).

- [36] R. Seiringer, *Gross-Pitaevskii Theory of the Rotating Bose Gas*, Commun. Math. Phys. **229**, 491–509 (2002).
- [37] R. Seiringer, *Ground state asymptotics of a dilute, rotating gas*, J. Phys. A: Math. Gen. **36**, 9755–9778 (2003).
- [38] R. Seiringer, *Free Energy of a Dilute Bose Gas: Lower Bound*, Commun. Math. Phys. **279**, 595–636 (2008).
- [39] R. Seiringer, D. Ueltschi, *Rigorous upper bound on the critical temperature of dilute Bose gases*, Phys. Rev. B **80**, 014502 (2009).
- [40] J. Yin, *Free energies of dilute Bose gases*, preprint arXiv:0906.1322.

VORTEX PATTERNS IN GINZBURG-LANDAU MINIMIZERS

SYLVIA SERFATY

*Courant Institute, New York University, and Université Pierre et Marie Curie Paris 6
Laboratoire Jacques-Louis Lions
175 rue du Chevaleret, 75013 Paris France.
E-mail: serfaty@ann.jussieu.fr*

ETIENNE SANDIER

*Département de mathématiques
Université Paris Est,
61, Avenue du Général de Gaulle
94010 Créteil Cedex, France.
E-mail: sandier@univ-paris12.fr*

We present a survey of results obtained with Etienne Sandier on vortices in the minimizers of the 2D Ginzburg-Landau energy of superconductivity with an applied magnetic field, in the asymptotic regime of large κ where vortices become point-like. We describe results which characterize the critical values of the applied field for which vortices appear, their numbers and locations. If the applied field is large enough, it is observed in experiments that vortices are densely packed and form triangular (hexagonal) lattices named Abrikosov lattices. Part of our results is the rigorous derivation of a mean field model describing the optimal density of vortices at leading order in the energy, and then the derivation of a next order limiting energy which governs the positions of the vortices after blow-up at their inter-distance scale. This limiting energy is a logarithmic-type interaction between points in the plane. Among lattice configurations it is uniquely minimized by the hexagonal lattice, thus providing a first justification of the Abrikosov lattice in this regime.

Keywords: Ginzburg-Landau, superconductivity, vortices, Abrikosov lattice, Gamma-convergence.

1. Introduction

We are interested in describing mathematical results on vortices in the two-dimensional Ginzburg-Landau model, a celebrated physics model.^a

Superconductivity consists in the complete loss of resistivity of certain metals and alloys below a certain critical temperature. The consequences of it are the possibility of permanent *superconducting currents* and the particular behavior that,

^aIt has earned Ginzburg the 2003 Physics Nobel Prize, jointly with Abrikosov for his work on explaining vortex lattices, and Legett for his modelling of superfluidity. Experimental discoveries on superconductivity and Bose-Einstein condensates have also won other physics Nobel prizes.

when the material is submitted to an external magnetic field, that field gets expelled from it. A striking phenomenon, predicted by A. Abrikosov, is the possibility of a *mixed state* in type II superconductors where triangular (or hexagonal) vortex lattices appear. Vortices can be described roughly as a quantized amount of vorticity of the superconducting current localized near a point. For reference on the subject, one can refer to standard physics texts, such as [10, 18, 25].

These phenomena have been the object of a huge amount of both experimental and theoretical studies on the physics side. In the last 15 years, the subject has also rapidly developed as a field of interest for mathematicians, who have derived and proved rigorously part of the phenomena observed in the physics, while providing additional information on it (for a survey of that literature, one can see Chapter 14 in [19]).

In addition to its importance in the modelling of superconductivity, the Ginzburg-Landau model turns out to be the simplest case of a gauge theory, and vortices to be the simplest case of topological solitons (for these aspects see [12, 14] and the references therein); moreover, it is mathematically extremely close to the Gross-Pitaevskii model for superfluidity, and models for rotating Bose-Einstein condensates in which quantized vortices are also essential objects, and to which the Ginzburg-Landau techniques have been successfully exported.

Here our aim is to survey some results we have obtained on minimizers of the Ginzburg-Landau model in the “large kappa limit”, in the regimes where vortices are expected to appear. For the sake of clarity we will be rather informal in this presentation, however complete proofs can be found in the papers referenced below. We will provide a description of

- the values of the critical fields for which vortices appear
- the vortex patterns for energy minimizers (ground states)
- the limiting energies which govern their interaction

1.1. The model

Consider a domain Ω in $\mathbb{R}^3$. In the Ginzburg-Landau model, the energy of a superconductor occupying Ω in the presence of a constant applied field H_{ex} , when the exterior region is insulating, is

$$G(u, A) = G_0 + \int_{\mathbb{R}^3} \frac{|\text{curl } A - H_{\text{ex}}|^2}{8\pi} + \int_{\Omega} \frac{1}{2m^*} \left| \left(\hbar \nabla - \frac{ie^*}{c} A \right) u \right|^2 + \alpha |u|^2 + \beta |u|^4. \quad (1)$$

In this expression, $u : \Omega \rightarrow \mathbb{C}$ is the *order parameter* (generally denoted ψ) whose physical meaning is that of a “wave function” for superconducting electron pairs and $A : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is the electromagnetic vector potential, whose curl is the induced magnetic field. Besides the physical constants $\hbar$ and c , additional constants m^* and e^* are present (see [25] for an explanation of these constants) as well as two quantities α and β that depend on the temperature T and on the superconducting material. Near the so-called critical temperature T_c , it is assumed that β

is a positive constant and α is proportional to $T - T_c$ and has the same sign. The quantity G_0 represents the energy of the normal state and does not depend on u or A . From then on, we consider we are below the critical temperature T_c . After some nondimensionalizing procedure (described for example in [19], Chap. 2) and reduction to a two-dimensional domain, the energy functional can be reduced to

$$G_\varepsilon(u, A) = \frac{1}{2} \int_\Omega |\nabla_A u|^2 + |\operatorname{curl} A - h_{\text{ex}}|^2 + \frac{1}{2\varepsilon^2} (1 - |u|^2)^2 \quad (2)$$

where Ω is now a bounded simply connected domain of $\mathbb{R}^2$.

Here A is now a function from Ω to $\mathbb{R}^2$. The magnetic field in the sample is deduced by $h = \operatorname{curl} A = \partial_1 A_2 - \partial_2 A_1$, it is thus a real-valued function in Ω . The notation ∇_A denotes the covariant gradient $\nabla - iA$; $\nabla_A u$ is thus a vector with complex components. u indicates the local state of the material or the phase in the Landau theory approach of phase transitions: $|u|^2$ is the density of “Cooper pairs” of superconducting electrons. With our normalization $|u| \leq 1$ and where $|u| \simeq 1$ the material is in the superconducting phase, while where $|u| = 0$, it is in the normal phase (i.e. behaves like a normal conductor), the two phases being able to coexist in the sample.

The *superconducting current* that we will denote j is a real vector given by $j = \langle iu, \nabla_A u \rangle$ where $\langle \cdot, \cdot \rangle$ denotes the scalar-product in $\mathbb{C}$ identified with $\mathbb{R}^2$, it may also be written as

$$\frac{i}{2} (u \overline{\nabla_A u} - \bar{u} \nabla_A u),$$

where the bar denotes the complex conjugation.

The parameter $h_{\text{ex}} > 0$ represents the intensity of the applied field (assumed to be perpendicular to the plane of Ω). Finally, the parameter ε is the inverse of the “Ginzburg-Landau parameter” usually denoted κ , a non-dimensional parameter depending on the material only. We will be interested in the regime of small ε , corresponding to high- κ (or extreme type-II) superconductors. The limit $\varepsilon \rightarrow 0$ or $\kappa \rightarrow \infty$ that we will consider is also called the London limit. In this limit, the characteristic size of the vortices, which is ε , tends to 0 and vortices become point-like. In this limit, h_{ex} will be a function of ε and not an independent parameter.

The stationary states of the system are the critical points of G_ε , or the solutions of the Ginzburg-Landau equations :

$$(GL) \left\{ \begin{array}{ll} -(\nabla_A)^2 u = \frac{1}{\varepsilon^2} u(1 - |u|^2) & \text{in } \Omega \\ -\nabla^\perp h = \langle iu, \nabla_A u \rangle & \text{in } \Omega \\ h = h_{\text{ex}} & \text{on } \partial\Omega \\ \nu \cdot \nabla_A u = 0 & \text{on } \partial\Omega, \end{array} \right.$$

where $\nabla^\perp$ denotes the operator $(-\partial_2, \partial_1)$, and ν the outer unit normal to $\partial\Omega$.

The Ginzburg-Landau equations and functional are invariant under $\mathbb{U}(1)$ -gauge-transformations (it is an Abelian gauge-theory):

$$\begin{cases} u \mapsto ue^{i\Phi} \\ A \mapsto A + \nabla\Phi \end{cases} \quad (3)$$

The physically relevant quantities are those that are gauge-invariant, such as the energy G_ε , $|u|$, h , etc...

In this paper we will only focus on *global minimizers of the energy* G_ε , in other words ground states. Other results for (stable or unstable) nonminimizing configurations can be found in [19].

1.2. Formal look at the solutions, vortices and critical fields

We start with some very formal considerations on solutions to the (GL) equations (further formal computations can be found in [10] or [19], Chapter 2).

1.2.1. Types of solutions

Three types of solutions (or states) can be found:

- (1) the normal solution : ($u \equiv 0, \text{curl } A \equiv h_{\text{ex}}$). This is a true solution to (GL) and its energy is very easily computed: it is $\frac{|\Omega|}{4\varepsilon^2}$.
- (2) the Meissner solution (or superconducting solution) : ($u \equiv 1, A \equiv 0$). This is a true solution if $h_{\text{ex}} = 0$, and a solution close to this one (i.e. with $|u| \simeq 1$ everywhere) persists if h_{ex} is not too large. Its energy is approximately $G_\varepsilon(1, 0) = \frac{h_{\text{ex}}^2}{2}|\Omega|$. By comparing these energies, we see that the Meissner solution is more favorable when h_{ex} is small, while the normal solution is more favorable when h_{ex} is large enough (depending on ε), more precisely when $h_{\text{ex}} > \frac{1}{\varepsilon\sqrt{2}}$.
- (3) the vortex solutions: there is another state, with vortices, called the *mixed state* where normal and superconducting phases co-exist, and which is more favorable for intermediate values of h_{ex} .

1.2.2. Vortices and their benefit

What are vortices? A vortex is an object centered at an isolated zero of u , around which the phase of u has a nonzero winding number, called the *degree of the vortex*. When ε is small, it is clear from (2) that any discrepancy between $|u|$ and 1 is strongly penalized, and a scaling argument hints that $|u|$ is different from 1 only in regions of characteristic size ε . A typical vortex centered at a point x_0 “looks like” $u = \rho e^{i\varphi}$ with $\rho(x_0) = 0$ and $\rho = f(\frac{|x-x_0|}{\varepsilon})$ where $f(0) = 0$ and f tends to 1 as $r \rightarrow +\infty$, i.e. its characteristic core size is ε , and

$$\frac{1}{2\pi} \int_{\partial B(x_0, R\varepsilon)} \frac{\partial \varphi}{\partial \tau} = d \in \mathbb{Z}$$

is an integer, called the *degree of the vortex*. For example $\varphi = d\theta$ where θ is the polar angle centered at x_0 yields a vortex of degree d .

True radial solutions in $\mathbb{R}^2$ of the Ginzburg-Landau equations of degree n , of the form

$$u_n(r, \theta) = f_n(r)e^{in\theta}, \quad A_n(r, \theta) = g_n(r)(-\sin \theta, \cos \theta)$$

have been shown to exist [5, 15, 16]. In a bounded domain, there are solutions with several such vortices glued together, for example arranged along a triangular lattice (their existence is proved at least as a bifurcation from the normal solution, see [3, 8]).

For a configuration with vortices we have the important (formal) relation

$$\operatorname{curl} \nabla \varphi = 2\pi \sum_i d_i \delta_{a_i} \quad (4)$$

where the a_i 's are the centers of the vortices and the d_i 's their degrees. This hints at the reason why vortices allow to gain energy. Indeed the first term in the energy $\int |\nabla_A u|^2$ can be written $\int |u|^2 |\nabla \varphi - A|^2$ so this term prefers $\nabla \varphi$ to be close to A , while the second term in the energy prefers h to be close to h_{ex} . But since $\operatorname{curl} A = h$ this means that $\operatorname{curl} \nabla \varphi$ should be close to h_{ex} . With the above relation (4), this translates into

$$2\pi \sum_i d_i \delta_{a_i} \text{ is as close as possible to the constant value } h_{\text{ex}}. \quad (5)$$

One sees that if there are no vortices the sum on the left-hand side is zero hence this will fail to be achieved if h_{ex} is large enough; on the contrary one sees how vortices can “help” energetically. Then, the question of understanding what (5) may exactly mean and in what sense, and what configurations of points a_i satisfy this assertion, is the core of the matter of our study. We can say right away that the conjecture is that (just like in the optimal packing problem) the configuration of vortex points a_i (with degrees $d_i = 1$) which optimizes (5), i.e. which best approximates the constant value h_{ex} , is the hexagonal Abrikosov lattice of density h_{ex} . However, as we shall see below, the situation is a bit more subtle than that, since there are some boundary effects which we have neglected in this formal discussion.

1.2.3. Critical fields

Let us now be more precise about the situation as a function of h_{ex} , as is known from the physics literature and the mathematical studies.

There are three main critical values of h_{ex} or *critical fields* H_{c_1} , H_{c_2} , and H_{c_3} , for which phase-transitions occur.

- For $h_{\text{ex}} < H_{c_1}$ the energy minimizer is the superconducting (Meissner) solution. H_{c_1} is of order of $|\log \varepsilon|$ as $\varepsilon \rightarrow 0$.
- For $h_{\text{ex}} = H_{c_1}$ the first vortex(es) appear.

- For $H_{c_1} < h_{\text{ex}} < H_{c_2}$ the superconductor is in the mixed phase i.e. there are vortices. The higher $h_{\text{ex}} > H_{c_1}$, the more vortices there are. The vortices repel each other so they tend to arrange in these hexagonal Abrikosov lattices in order to minimize their repulsion.
- For $h_{\text{ex}} = H_{c_2} \sim \frac{1}{\varepsilon^2}$, the vortices are so densely packed that they overlap each other, and a second phase transition occurs, after which $|u| \sim 0$ inside the sample, i.e. all superconductivity in the bulk of the sample is lost.
- For $H_{c_2} < h_{\text{ex}} < H_{c_3}$ superconductivity persists only near the boundary, this is called *surface superconductivity*.
- For $h_{\text{ex}} > H_{c_3} = O(\frac{1}{\varepsilon^2})$ (defined in decreasing fields), the sample is completely in the normal phase $u \equiv 0$.

In Section 2 we give a precise mathematical description of that picture for all h_{ex} much smaller than H_{c_2} , at leading order of the energy. In Section 3 we present more recent results where we refine this study in the cases with a large number of vortices, to derive a next order interaction energy which governs the vortex patterns.

2. Main results for global minimization at the leading order

In all that follows the notation $a = O(b)$ will mean that a/b remains bounded, and $a \ll b$ that $a/b \rightarrow 0$, in the limit $\varepsilon \rightarrow 0$.

2.1. The vorticity measures

Recall that a complex-valued map u can be written in polar coordinates $u = \rho e^{i\varphi}$ with a phase φ which can be multi-valued. Given a configuration (u, A) , we define its *vorticity* by

$$\mu(u, A) = \text{curl} \langle iu, \nabla_A u \rangle + \text{curl} A = \text{curl} j + h, \quad (6)$$

where we recall that $j = \langle iu, \nabla_A u \rangle$ is the superconducting current. Formally, considering that $\rho = |u| \simeq 1$ we have $\mu(u, A) = \text{curl}(\rho^2(\nabla\varphi - A)) + \text{curl} A \simeq \text{curl} \nabla\varphi$, so using (4), we have the approximate (formal) relation

$$\mu(u, A) = 2\pi \sum_i d_i \delta_{a_i} \quad (7)$$

where a_i 's are the vortices of u and d_i 's their degrees, and δ the Dirac mass. This relation can be justified rigorously in some appropriate sense, see [13] or [19], Theorem 6.1. Thus we see why the gauge-invariant quantity $\mu(u, A)$ is appropriate as a proxy for the vortices of u (it is formally like the vorticity for fluids).

Remark 2.1. When the second of the Ginzburg-Landau equations (*GL*) is satisfied, taking its curl, we find that the vorticity and the induced field are linked by the London equation (this time an identity and not an approximation)

$$\begin{cases} -\Delta h + h = \mu(u, A) & \text{in } \Omega \\ h = h_{\text{ex}} & \text{on } \partial\Omega. \end{cases} \quad (8)$$

Thus the knowledge of the vorticity is equivalent to that of the induced field h .

2.2. Global minimizers of G_ε close to H_{c_1}

Let us introduce h_0 the solution of

$$\begin{cases} -\Delta h_0 + h_0 = 0 & \text{in } \Omega \\ h_0 = 1 & \text{on } \partial\Omega \end{cases} \quad (9)$$

and

$$\lambda_\Omega = (2 \max |h_0 - 1|)^{-1}. \quad (10)$$

We also introduce the set $\Lambda = \{x \in \Omega / h_0(x) = \min h_0\}$ and we will assume here for simplicity that it is reduced to only one point called $\bar{p}$, and denote $Q(x) = \langle D^2 h_0(\bar{p})x, x \rangle$ its second order differential, assumed to be definite positive.

The first vortices will appear near the point $\bar{p}$, and in order to describe them, we will perform blow-ups around $\bar{p}$ at suitable scales.

With these notation, a first essential result is the asymptotic formula for H_{c_1} (confirming physical predictions that H_{c_1} is of the order of $|\log \varepsilon|$ as $\varepsilon \rightarrow 0$):

$$H_{c_1} = \lambda_\Omega |\log \varepsilon| + cst. \quad (11)$$

Theorem 2.1 (cf. [19, 23]). *There exists an increasing sequence of values*

$$H_n = \lambda_\Omega |\log \varepsilon| + (n-1)\lambda_\Omega \log \frac{|\log \varepsilon|}{n} + \text{constant order terms}$$

such that if $h_{\text{ex}} \leq H_{c_1} + O(\log |\log \varepsilon|)$ and $h_{\text{ex}} \in (H_n, H_{n+1})$, then global minimizers of G_ε have exactly n vortices of degree 1, at points $a_i^\varepsilon \rightarrow \bar{p}$ as $\varepsilon \rightarrow 0$, and the $\tilde{a}_i^\varepsilon = \sqrt{\frac{h_{\text{ex}}}{n}}(a_i^\varepsilon - \bar{p})$ converge as $\varepsilon \rightarrow 0$ to a minimizer of

$$w_n(x_1, \dots, x_n) = -\pi \sum_{i \neq j} \log |x_i - x_j| + \pi n \sum_{i=1}^n Q(x_i). \quad (12)$$

Through this theorem we see that the behavior is as expected: below $H_{c_1} = H_1$ there are no vortices in energy minimizers, then at H_{c_1} the first vortex becomes favorable, close to the point $\bar{p}$. Then, there is a sequence of additional critical fields $H_2, H_3, \dots$ separated by increments of $\log |\log \varepsilon|$, for which a second, third, $\dots$, vortex becomes favorable. Each time the optimal vortices are located close to $\bar{p}$ as $\varepsilon \rightarrow 0$ (cf. Fig. 1) and after blowing-up at the scale $\sqrt{\frac{h_{\text{ex}}}{n}}$ around $\bar{p}$, they converge to configurations which minimize w_n in $\mathbb{R}^2$. Now, w_n , which appears as a limiting energy (after that rescaling) contains a repulsion and a confinement term. It is a standard two-dimensional interaction, however rigorous results on its minimization are hard to obtain as soon as $n \geq 3$. When Q has rotational symmetry, numerical minimization (see Gueron-Shafrir [11]) yields very regular shapes (regular polygons for $n \leq 6$, regular stars) which look very much like the birth of a triangular lattice as n becomes large (their density tends to be uniform supported in a fixed disc of $\mathbb{R}^n$ as $n \rightarrow \infty$), see Fig. 2.

All these results are in very good agreement with experimental observations.

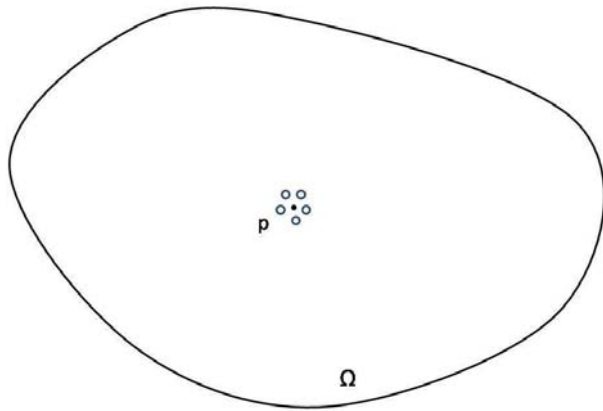


Fig. 1. Minimizers with small number of vortices

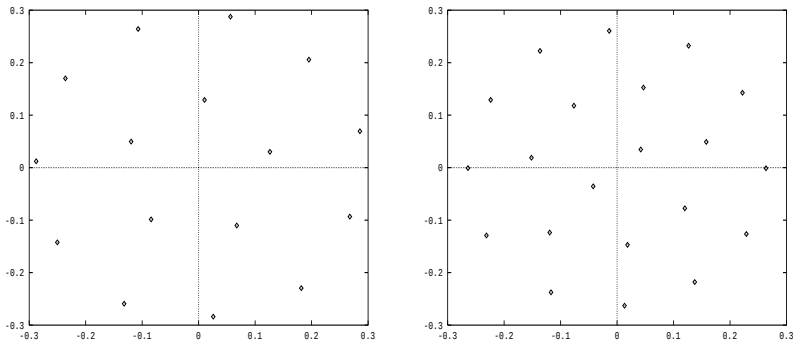


Fig. 2. Results of the numerical optimization of [11] for w_n , $n = 16$ and $n = 21$.

Remark 2.2. It was proved in [24] that for $h_{\text{ex}} < H_{c1}$, the energy-minimizer is unique and has no vortex.

2.3. Global minimizers in the intermediate regime

In the next higher regime of applied field, the result is the following:

Theorem 2.2 (cf. [19]). Assume h_{ex} satisfies as $\varepsilon \rightarrow 0$,

$$\log |\log \varepsilon| \ll h_{\text{ex}} - H_{c1} \ll |\log \varepsilon|$$

then there exists $1 \ll n_\varepsilon \ll h_{\text{ex}}$ such that

$$h_{\text{ex}} \sim \lambda_\Omega \left(|\log \varepsilon| + n_\varepsilon \log \frac{|\log \varepsilon|}{n_\varepsilon} \right)$$

and if $(u_\varepsilon, A_\varepsilon)$ minimizes G_ε , then

$$\frac{\tilde{\mu}(u_\varepsilon, A_\varepsilon)}{2\pi n_\varepsilon} \rightarrow \quad \text{in the weak sense of measures}$$

where $\tilde{\mu}(u_\varepsilon, A_\varepsilon)$ is the push-forward of the vorticity measure $\mu(u_\varepsilon, A_\varepsilon)$ under the blow-up $x \mapsto \sqrt{\frac{h_{\text{ex}}}{n_\varepsilon}}(x - \bar{p})$, and μ_0 is the unique minimizer over probability measures of

$$I(\mu) = -\pi \int_{\mathbb{R}^2 \times \mathbb{R}^2} \log |x - y| d\mu(x) d\mu(y) + \pi \int_{\mathbb{R}^2} Q(x) d\mu(x). \quad (13)$$

Here, n_ε corresponds to the expected optimal number of vortices. Note that from (7) we have

$$\tilde{\mu}(u_\varepsilon, A_\varepsilon) \simeq 2\pi \sum_i d_i \delta_{\tilde{a}_i}$$

where the $\tilde{a}_i$'s are the images of the true vortices of $(u_\varepsilon, A_\varepsilon)$ after the blow-up.

The problem of minimizing I is a classical one in potential theory. Its minimizer μ_0 is a probability measure of constant density over a subdomain of $\mathbb{R}^2$ (typically a disc or an ellipse). This result is in continuous connection with Theorem 2.1, except $n_\varepsilon \gg 1$. Again, vortices in the minimizers converge to $\bar{p}$ as $\varepsilon \rightarrow 0$, and when one blows up at the right scale $\sqrt{\frac{h_{\text{ex}}}{n_\varepsilon}}$ around $\bar{p}$, one obtains a uniform density of vortices in a subdomain of $\mathbb{R}^2$ (a disc if $D^2 h_0$ has rotational symmetry).

2.4. Global minimizers in the regime n_ε proportional to h_{ex}

This happens in the next regime: $h_{\text{ex}} \sim \lambda |\log \varepsilon|$ with $\lambda > \lambda_\Omega$. Let us define

$$E_\lambda(\mu) = \frac{1}{2\lambda} \int_\Omega |\mu| + \frac{1}{2} \int_\Omega |\nabla h_\mu|^2 + |h_\mu - 1|^2,$$

over bounded Radon measures such that $E_\lambda(\mu) < \infty$, where $\int_\Omega |\mu|$ is the total mass of the measure μ and

$$\begin{cases} -\Delta h_\mu + h_\mu = \mu & \text{in } \Omega \\ h_\mu = 1 & \text{on } \partial\Omega \end{cases} \quad (14)$$

is the magnetic field associated to μ as in (8).

In this regime, the region filled up with vortices is no longer concentrating just around the point $\bar{p}$ but is spread out at finite distance from $\bar{p}$, so there is no need to blow-up in order to distinguish the vortices. However, we still need to divide or normalize the vorticity $\mu(u, A)$ by the order of the expected number of vortices, which is h_{ex} , and blows up with $\varepsilon \rightarrow 0$.

Theorem 2.3 (cf. [19, 21]). Assume $h_{\text{ex}} = \lambda |\log \varepsilon|$ where $\lambda > 0$ is a constant independent of ε . If $(u_\varepsilon, A_\varepsilon)$ minimizes G_ε , then as $\varepsilon \rightarrow 0$

$$\frac{\mu(u_\varepsilon, A_\varepsilon)}{h_{\text{ex}}} \rightarrow \mu_* \quad \text{in the weak sense of measures}$$

where μ_* is the unique minimizer of E_λ .

Observe also that E_λ can be rewritten

$$E_\lambda(\mu) = \frac{1}{2\lambda} \int_\Omega |\mu| + \frac{1}{2} \int_{\Omega \times \Omega} G(x, y) d(\mu - 1)(x) d(\mu - 1)(y) \quad (15)$$

where G is a Green's function, the solution to $-\Delta G + G = \delta_y$ with $G = 0$ on $\partial\Omega$. That way, the similarity with I is more apparent.

There remains to minimize E_λ . It turns out that this problem is dual in the sense of convex duality to an obstacle problem:

Proposition 2.1. μ minimizes E_λ if and only if h_μ is the minimizer for the variational problem

$$\min_{\substack{h \geq 1 - \frac{1}{2\lambda} \\ \eta=1 \text{ on } \partial\Omega}} \int_\Omega |\nabla h|^2 + h^2. \quad (16)$$

Now, the solution of the obstacle problem (16) is well-known, and given by a variational inequality. Obstacle problems are a particular type of *free-boundary problems*, the free-boundary here being the boundary of the so-called “coincidence set”

$$\Omega_\lambda = \left\{ x \in \Omega / h_{\mu_*}(x) = 1 - \frac{1}{2\lambda} \right\} = \text{supp } \mu_*. \quad (17)$$

Then h_{μ_*} verifies $-\Delta h_{\mu_*} + h_{\mu_*} = 0$ outside of ω_λ , so ω_λ is really the support of μ_* , on which μ_* is equal to the constant density $1 - \frac{1}{2\lambda}$ so we may write $\mu_* = (1 - \frac{1}{2\lambda}) \mathbf{1}_{\omega_\lambda}$, see Fig. 3. μ_* is thus completely characterized (ω_λ is itself a set uniquely determined by λ via (16)).

An easy analysis of this obstacle problem yields the following:

- (1) $\omega_\lambda = \emptyset$ (hence $\mu_* = 0$) if and only if $\lambda < \lambda_\Omega$, where λ_Ω was given by (10). (This corresponds to the case $h_{\text{ex}} < H_{c_1}$.)
- (2) For $\lambda = \lambda_\Omega$ then $\omega_\lambda = \{\bar{p}\}$. This is the case when $h_{\text{ex}} \sim H_{c_1}$ at leading order. In the scaling chosen here $\mu_* = 0$ but the true behavior of the vorticity is ambiguous unless going to the next order term as done in Theorems 2.1 and 2.2.
- (3) For $\lambda > \lambda_\Omega$, the measure of ω_λ is nonzero, so the limiting vortex density $\mu_* \neq 0$. Moreover, as λ increases (i.e. as h_{ex} does), the set ω_λ increases. When $\lambda = +\infty$, ω_λ becomes Ω and $\mu_* = 1$, this corresponds to the case $h_{\text{ex}} \gg |\log \varepsilon|$ of the next subsection.

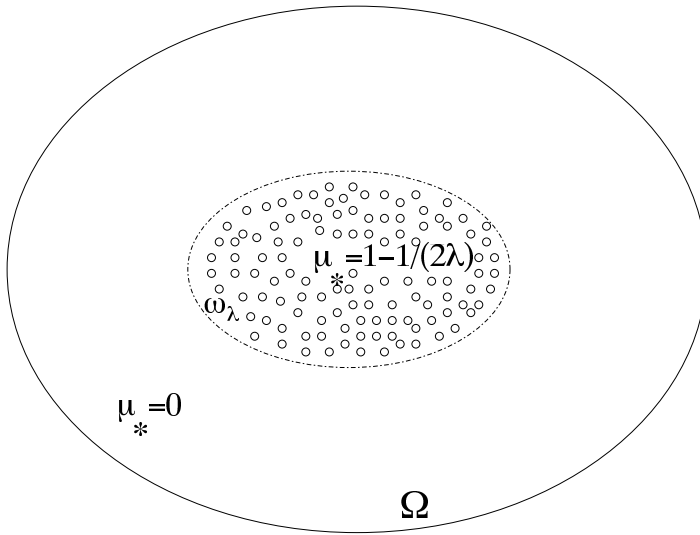


Fig. 3. Optimal density of vortices according to the obstacle problem.

2.5. Global minimizers in the regime $|\log \varepsilon| \ll h_{\text{ex}} \ll \varepsilon^{-2}$

For applied fields larger than $\frac{1}{\varepsilon^2}$ but below H_{c_2} , the parameter $\lambda = \frac{h_{\text{ex}}}{|\log \varepsilon|}$ above is formally $+\infty$ and we find $\min E_\infty = 0$ and $\mu_* = 1$. The vortex density is thus found to be uniform but the expansion $\min G \sim h_{\text{ex}}^2 E_\infty(\mu_*)$ is degenerate. However, even though the number of vortices becomes very large, the minimization problem becomes local and can be solved by blowing-up and using Theorem 2.3. The minimal energy density can then be evaluated:

Theorem 2.4 (cf. [19, 20]). *Assume, as $\varepsilon \rightarrow 0$, that $|\log \varepsilon| \ll h_{\text{ex}} \ll 1/\varepsilon^2$. Then, letting $(u_\varepsilon, A_\varepsilon)$ minimize G_ε , and letting $g_\varepsilon(u, A)$ denote the energy-density $\frac{1}{2} (|\nabla_A u|^2 + |h - h_{\text{ex}}|^2 + \frac{1}{2\varepsilon^2} (1 - |u|^2)^2)$, we have*

$$\frac{2g_\varepsilon(u_\varepsilon, A_\varepsilon)}{h_{\text{ex}} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}}} \rightarrow dx \quad \text{as } \varepsilon \rightarrow 0$$

in the weak sense of measures, where dx denotes the two-dimensional Lebesgue measure; and thus

$$\min G_\varepsilon(u, A) \sim \frac{|\Omega|}{2} h_{\text{ex}} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \quad \text{as } \varepsilon \rightarrow 0,$$

where $|\Omega|$ is the area of Ω . Moreover

$$\frac{\mu(u_\varepsilon, A_\varepsilon)}{h_{\text{ex}}} \rightarrow dx \quad \text{in the weak sense of measures.}$$

2.6. Methods of the proofs

The method we use to obtain those results is the scheme of “ Γ -convergence”, consisting in evaluating very precisely the minimal energy via upper bounds (obtained by an explicit construction and computation) and matching lower bounds which are *Ansatz-free*. This requires being able to describe very precisely the energy carried by vortices even if there is a very large number of them. The method for that originates in the book [6] where situations with bounded numbers of vortices were studied in a simplified context (without magnetic field). Later on, a more generally applicable “technology” was developed by Jerrard and Sandier independently to obtain estimates of self-interaction energy of vortices even when their number if unbounded. The main estimate is that each vortex of degree d carries an energy at least $\pi|d|\log \frac{r}{\varepsilon}$ where r is the distance to its nearest neighbors (for more precise estimates see Theorem 4.1 in [19]). The rest of our analysis then consists in comparing via appropriate splittings of the Ginzburg-Landau functional, this self-interaction energy cost and the vortex-repulsion cost to the energy “benefit” of the vortices due to their interaction with the magnetic field.

2.7. Summary

In the results above we have identified the critical fields and regimes for which vortices appear, and have characterized the optimal vortex densities *at leading order*, i.e. derived either limiting interaction energies or mean field models in the cases of Theorems 2.3 and 2.4. More precisely we have identified the following regimes:

- (1) When $h_{\text{ex}} < H_{c_1}$ there are no vortices.
- (2) At $h_{\text{ex}} = H_{c_1} = H_1$ one vortex of degree 1 appears, near the point $\bar{p}$. As h_{ex} crosses $H_2, H_3, \dots, H_n$ an n -th vortex of degree 1 appears, also near $\bar{p}$. After blow-up around $\bar{p}$ these n vortices tend to arrange according to regular shapes minimizing w_n (see again Fig. 1, 2).
- (3) When $\log |\log \varepsilon| \ll h_{\text{ex}} - H_{c_1} \ll |\log \varepsilon|$ then the number of vortices n is no longer bounded with respect to ε , but remains $\ll h_{\text{ex}}$. There is then a “cloud” of vortices around the point $\bar{p}$, and when blown-up at the scale $\sqrt{\frac{h_{\text{ex}}}{n}}$ this cloud appears as a uniform density supported in an ellipse.
- (4) When $h_{\text{ex}} = \lambda |\log \varepsilon|$ with $\lambda > \lambda_\Omega$ there is a cloud of vortices with uniform density (proportional to h_{ex}) over the subdomain ω_λ (completely determined by λ), and essentially no vortices outside ω_λ , as in Fig. 3.
- (5) When $|\log \varepsilon| \ll h_{\text{ex}} \ll \frac{1}{\varepsilon^2}$ there is a uniform cloud of vortices covering up the whole domain Ω with a constant density h_{ex} .

3. Next-order behavior of vortices

Going back to the conclusions above, the behavior of individual vortices is completely understood in the regimes corresponding to items 1,2 where the number of vortices is finite and fixed. In items 3,4,5 the number of vortices blows up as $\varepsilon \rightarrow 0$,

and we only know their optimal average density. Such a constant optimal density is in agreement with the Abrikosov lattice where the average density of vortices is also constant, however it is far from justifying the presence of a lattice. Many other patterns (starting with non-hexagonal lattices) are admissible. In other words the mean field description above is insensitive to the precise pattern formed by vortices. This pattern is in fact, as we shall see, selected by the minimization of the next term in the asymptotic expansion of the energy as $\varepsilon \rightarrow 0$. This is the object of what follows. Proving this requires more sophisticated tools to obtain more precise estimates on the energy of vortices than the $\pi|d|\log \frac{r}{\varepsilon}$ equivalent mentioned above (in fact we really need to estimate this vortex energy up to a $o(1)$ error). In order to understand the vortex patterns, which are really driven by the next order interaction term, one then needs to blow up (i.e. zoom in) the solutions in space at the scale where one sees individual well-separated vortices.

3.1. *Splitting of the energy and blow-up*

The next order expansion of the energy is achieved by finding a splitting of the energy which separates, via an identity, the leading order term found above from a remainder. This splitting states (roughly) that

$$G_\varepsilon(u, A) = h_{\text{ex}}^2 E_\lambda(\mu_*) + G_1(u, A) \quad (18)$$

where $G_1(u, A)$ is a new functional but of order h_{ex} , hence a next order correction, since $h_{\text{ex}} \rightarrow +\infty$ as $\varepsilon \rightarrow 0$ in the regime of interest. In the case $h_{\text{ex}} \gg |\log \varepsilon|$ this should be replaced by

$$G_\varepsilon(u, A) = \frac{|\Omega|}{2} h_{\text{ex}} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} + G_1(u, A) \quad (19)$$

where again $G_1(u, A)$ is of order h_{ex} . Minimizers of G are obviously the same as minimizers of G_1 hence there remains to minimize G_1 .

The results we obtain below are valid for all the regimes 3,4,5 above. However, for the sake of simplicity of the presentation, we will sometimes reduce to case 5, where we have a uniform density of vortices in Ω , and to the simpler periodic setting.

We explain the proof of the energy splitting (18)–(19) in the periodic situation where it is much simpler: assuming we are in a torus and all the gauge-invariant quantities are periodic, let $h = \text{curl } A$ and set $h_1 = h - h_{\text{ex}} + \frac{1}{2} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}}$. Inserting into (2) and expanding the square terms we find

$$G_\varepsilon(u, A) = \frac{1}{2} \int_\Omega |\nabla_A u|^2 + h_1^2 + \frac{(1 - |u|^2)^2}{2\varepsilon^2} + \frac{1}{8} \left(\log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \right)^2 |\Omega| - \frac{1}{2} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \int_\Omega h_1.$$

On the other hand we have the London equation (8) $-\Delta h + h = \mu(u, A)$ so

$$-\Delta h_1 + h_1 = \mu(u, A) - h_{\text{ex}} + \frac{1}{2} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \quad (20)$$

and since we are in a periodic situation it follows that

$$\int_{\Omega} h_1 = (-h_{\text{ex}} + \frac{1}{2} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}}) |\Omega| + \int_{\Omega} \mu(u, A).$$

Inserting into the above we find

$$G_{\varepsilon}(u, A) = \frac{|\Omega|}{2} h_{\text{ex}} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} - \frac{1}{8} \left(\log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \right)^2 |\Omega| + G_1(u, A) \quad (21)$$

where

$$G_1(u, A) = \left[\frac{1}{2} \int_{\Omega} |\nabla_A u|^2 + h_1^2 + \frac{(1 - |u|^2)^2}{2\varepsilon^2} \right] - \frac{1}{2} \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \int_{\Omega} \mu(u, A). \quad (22)$$

This is actually also the functional G_1 that we find in the non-periodic case. Note that the constant leading order term in (21) is the same as that found in Theorem 2.4.

Let us continue to restrict for simplicity to this periodic case with $h_{\text{ex}} \gg |\log \varepsilon|$. Since the optimal vortex density is h_{ex} , the typical intervortex distance is $\frac{1}{\sqrt{h_{\text{ex}}}}$ so we should blow up the configurations by $\sqrt{h_{\text{ex}}}$ to see well-separated vortices in the plane, with an average density equal to 1. Then, let us take a point x at random in the domain, and blow-up at the scale $\sqrt{h_{\text{ex}}}$ around this point. More precisely set $H(y) = h_1(x + \frac{y}{\sqrt{h_{\text{ex}}}})$. Scaling appropriately in the equation (20) we find (using that $\log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \ll h_{\text{ex}}$) that the limit of H as $\varepsilon \rightarrow 0$ should satisfy

$$-\Delta H = 2\pi \sum_i d_i \delta_{p_i} - 1 \quad (23)$$

where this time the p_i 's are the images of the true vortices a_i of u under the blow up around x at scale $\sqrt{h_{\text{ex}}}$. These should now be well-separated points in the whole plane $\mathbb{R}^2$, and the relation (23) holds in all $\mathbb{R}^2$, the sum being infinite. The complication is that we have one such relation for each choice of blow-up center x .

3.2. The “renormalized” energy

The next question is then to understand how the energy G_1 governs the interaction between these limiting blown-up vortex points p_i . One can notice that (22) is the difference between two terms: a first positive energy which looks very much like a Ginzburg-Landau energy with zero external magnetic field, and a second term which from (7) is roughly equal to $-\pi \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \sum_i d_i$ where the d_i 's are the degrees of the vortices. Moreover as mentioned above it is expected (see [19]) that in the positive Ginzburg-Landau part, each vortex at distance $\frac{1}{\sqrt{h_{\text{ex}}}}$ from its nearest neighbours has a cost (or self-interaction energy) $\sim \pi d_i^2 \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}}$. So this cost gets cancelled by the other term, provided $d_i = 1$. Thus if we do not have $d_i = 1$ the total cost of one vortex blows up, so energetically only vortices of degrees ± 1 are favorable. So without loss of generality we may assume that all the d_i 's above are ± 1 . Then the cost of each vortex should be exactly compensated by the $-\pi \log \frac{1}{\varepsilon \sqrt{h_{\text{ex}}}} \sum_i d_i$

term. If this can all be made rigorous, what should be left after this subtraction of two “infinite” costs should be an energy of lower order, corresponding just to the interaction cost between the vortices, or a “renormalized energy” cost (we use this expression by analogy with [6]). Extracting this energy thus requires very precise energy estimates. The second thing one can notice is that from the second (GL) equation we have $j = -\nabla^\perp h = -\nabla^\perp h_1$ with $j = |u|^2(\nabla\varphi - A)$ and so $\int_\Omega |\nabla_A u|^2$ is approximately equal to $\int_\Omega |\nabla h|^2 = \int_\Omega |\nabla h_1|^2$. So the energy left in G_1 should be expressible in terms of $\int_\Omega |\nabla H|^2$. It is in fact more convenient below to express things in terms of j .

We are now in a position to present the renormalized energy that we derive through this procedure.

Definition 3.1. Let j be a vector field in $\mathbb{R}^2$ and ν a positive measure on $\mathbb{R}^2$. We say (j, ν) belongs to the admissible class $\mathcal{A}$ if

$$\nu = 2\pi \sum_{p \in \Lambda} \delta_p \quad \text{for some discrete set } \Lambda \subset \mathbb{R}^2,$$

$$\lim_{R \rightarrow \infty} \frac{\nu(B_R)}{|B_R|} = 1 \quad (24)$$

and

$$\operatorname{curl} j = \nu - 1 = 2\pi \sum_{p \in \Lambda} \delta_p - 1 \quad \operatorname{div} j = 0. \quad (25)$$

We denote by χ_{B_R} cutoff functions associated to the family of balls B_R centered at the origin and of radius R such that

$$|\nabla \chi_{B_R}| \leq 2, \quad \chi_{B_R}(x) \equiv 1 \text{ in } B_{R-1}, \quad \chi_{B_R}(x) = 0 \text{ outside } B_R. \quad (26)$$

Definition 3.2. The renormalized energy W is defined, for $(j, \nu) \in \mathcal{A}$, by

$$W(j) = \limsup_{R \rightarrow \infty} \frac{W(j, \chi_{B_R})}{|B_R|}, \quad (27)$$

(we denote the dependence as of j only since ν can be deduced from j via (25)) where

$$W(j, \chi_{B_R}) = \lim_{\eta \rightarrow 0} \left(\frac{1}{2} \int_{\mathbb{R}^2 \setminus \bigcup_{p \in \Lambda} B(p, \eta)} \chi_{B_R} |j|^2 + \pi \log \eta \sum_{p \in \Lambda} \chi_{B_R}(p) \right). \quad (28)$$

Observe that the limiting j (limit of the superconducting current j_ε after blow-up and $\varepsilon \rightarrow 0$) should be equal to $-\nabla^\perp H$ where H satisfies (23). But solutions to (23) in dimension 2 have a logarithmic singularity at each point p_i , thus $\int |\nabla H|^2 = \int |j|^2$ diverges near each vortex point p . This is why in (28) small holes need to be cut out around each p . Adding the expected logarithmic divergence $\pi \log \eta$ for each vortex and letting η tend to 0 is like “renormalizing” and taking the regular part of a

Green's function, hence the need for the parameter η . The cut-off χ_R is just there to avoid boundary effects, since otherwise $W(j, \mathbf{1}_{B_R})$ would oscillate wildly between $+\infty$ and $-\infty$ as a point p tends to ∂B_R .

In the end W is a logarithmic interaction between points in the plane. It behaves like $-\pi \log |p_i - p_j|$ when $|p_i - p_j| \rightarrow 0$. However it is not only an interaction between the points p 's but rather the interaction between these points acting like charged particles and between them and the fixed background constant "charge" -1 (see again (23)), averaged over larger and larger balls, see Fig. 4.

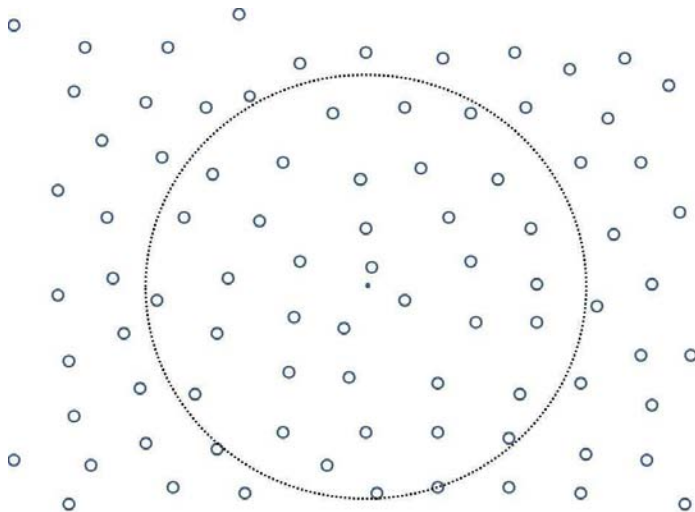


Fig. 4. The energy of a collection of blown-up vortices in the whole plane (with density 1 at infinity) is calculated by averaging over larger and larger balls

We will comment more on the minimization of W after the statement of our last main theorem.

3.3. Main result

For the statement of the theorem we return to a bounded domain and general values of h_{ex} . The blow-up procedure should now be made around any point of ω_λ , the support of the limiting vortex density, defined in (17). Moreover, since the expected limiting vortex density is no longer h_{ex} but $(1 - \frac{1}{2\lambda})h_{\text{ex}}$, the blow-up should be made at a slightly modified scale.

Theorem 3.1 (cf. [22]). *Assume that*

$$\lambda \in (\lambda_\Omega, +\infty] \text{ and } h_{\text{ex}} = \lambda |\log \varepsilon|, \quad (29)$$

and if $\lambda = +\infty$ that $h_{\text{ex}} \ll \frac{1}{\varepsilon^2}$. Let $(u_\varepsilon, A_\varepsilon)$ be a minimizer of G_ε and for $x \in \omega_\lambda$ let

$$\tilde{j}_{\varepsilon,x}(\cdot) = \frac{1}{\sqrt{(1 - 1/(2\lambda))h_{\text{ex}}}} j(u_\varepsilon, A_\varepsilon) \left(x + \frac{\cdot}{\sqrt{(1 - 1/(2\lambda))h_{\text{ex}}}} \right)$$

be the blow-up image of the superconducting current at the appropriate scale. There exists a probability measure P on vector fields on $\mathbb{R}^2$ such that the following hold:

(1) up to extraction, for any continuous function Φ on the space of vector fields on $\mathbb{R}^2$, we have

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{|\omega_\lambda|} \int_{\omega_\lambda} \Phi(\tilde{j}_{\varepsilon,x}) dx = \int \Phi(j) dP(j) \quad (30)$$

(2) P -almost every j satisfies

$$\text{curl } j = 2\pi \sum_{p \in \Lambda} \delta_p - 1, \quad \text{div } j = 0,$$

for some discrete subset Λ of $\mathbb{R}^2$

(3) P -almost every j minimizes the function W of (27).

Moreover, we have

$$\min G_1 = \left(1 - \frac{1}{2\lambda}\right) h_{\text{ex}} |\omega_\lambda| \min W + o(h_{\text{ex}})$$

so $\min G$ may be computed up to $o(h_{\text{ex}})$.

The above theorem may seem a bit abstract however its concrete meaning is the following : if one considers an energy minimizer and picks a blow up point at random and looks at the blown-up profile of vortices around that point, then almost surely, one sees a minimizer of W .

The need for probability measures is due to the difficulty in localizing energy lower bounds since after blown-up one is on an unbounded domain. The probability measure approach allows to do this via the use of the ergodic theorem. Note that from the characterization (30) this probability measure is like a Young measure, but in contrast with Young measures, it is not a probability measure on values taken by the functions, but rather a probability measure on the whole limiting profile around a point, so it contains more information. Our approach can be seen as an alternate to the related approach in [2].

Looking again at (23) and viewing W as a “renormalized” computation of $\int |\nabla H|^2$, we see that W measures in some sense the size of $2\pi \sum_i \delta_{p_i} - 1$. So we are back (but at the blown-up scale) to the question of (5): W is a measure of how close this sum of Diracs is to a constant, and W should be minimized. The open question is then to know whether, like in packing problems, W is minimized by configurations of points which form a triangular lattice. This is a question of crystallisation, for a logarithmic type of interaction, and as such, it is known to be very difficult to answer. Even proving that minimizers have some periodicity seems out of reach.

However, we can answer an easier question: what are minimizers of W if one restricts to pure lattice configurations?

Theorem 3.2 (cf. [22]). *The minimum of W over lattices of volume 2π is uniquely achieved by the hexagonal (or triangular) lattice.*

A simple proof can be found using results from number theory. Indeed let H be a solution to $-\Delta H = \delta_0 - 1$ on a torus of volume 1 of arbitrary shape. The expression (28) simplifies quite a bit under this periodicity assumption and we can Fourier transform the explicit expression for W in that case to make it a function of the lattice. It then becomes a regularisation of the divergent series $\sum_{p \in \Lambda} \frac{1}{|p|^2}$. By some transformations on modular functions, minimizing W becomes equivalent to minimizing the Epstein zeta function $\zeta(s) = \sum_{p \in \Lambda} \frac{1}{|p|^s}$ with $s > 2$, over lattices Λ with fixed volume. Results from number theory from the 60's [7, 17] say that this is uniquely minimized by the hexagonal lattice.

So we conclude that at least W allows to distinguish between lattices. We have thus derived a nontrivial variational problem for the emergence of Abrikosov lattices, and provided to our knowledge, the first rigorous justification of the Abrikosov lattice in this regime: at least the hexagonal lattice is the best among perfect lattice configurations (it beats the square lattice for example). Note that analogous results were obtained in [1, 4] but for the regime of h_{ex} very near H_{c_2} which is a very different regime where the problem becomes essentially linear.

These results are obtained once more by upper bounds on the minimal energy obtained via an explicit construction, combined with matching and ansatz-free lower bounds, made possible by the energy-splitting presented above and by the ergodic theorem approach (which allows in particular to control the number of vortices per unit volume after blow-up, around most blow-up centers).

References

- [1] Aftalion, A; Serfaty, S. Lowest Landau level approach in superconductivity for the Abrikosov lattice close to H_{c_2} , *Selecta Math.* **13**, (2007), No 2, 183–202.
- [2] Alberti, G.; Müller, S. A new approach to variational problems with multiple scales. *Comm. Pure Appl. Math.* **54**, (2001), No 7, 761–825.
- [3] Almog, Y. On the bifurcation and stability of periodic solutions of the Ginzburg-Landau equations in the plane. *SIAM J. Appl. Math.* **61** (2000), no. 1, 149–171.
- [4] Almog, Y., Abrikosov lattices in finite domains, *Comm. Math. Phys.* **262**, (2006), No. 3, 677–702.
- [5] Berger, M. S.; Chen, Y. Y. Symmetric vortices for the nonlinear Ginzburg-Landau of superconductivity, and the nonlinear desingularization phenomenon. *J. Funct. Anal.* **82** (1989), no. 2, 259–295.
- [6] Bethuel, F.; Brezis, H.; Hélein, F. *Ginzburg-Landau vortices*. Birkhäuser Boston, Boston, 1994.
- [7] Cassels, On a problem of Rankin about the Epstein zeta-function *Proc. Glasgow Math. Assoc.* (1959).
- [8] Chapman, S. J. Nucleation of superconductivity in decreasing fields. *Eur. J. Appl. Math.* **5** (1994), part 1, 449–468; part 2, 468–494.

- [9] Chapman, S. J.; Rubinstein, J.; Schatzman, M. A mean-field model of superconducting vortices. *Eur. J. Appl. Math.* **7** (1996), no. 2, 97–111.
- [10] DeGennes, P. G. *Superconductivity of metal and alloys*. Benjamin, New York and Amsterdam, 1966.
- [11] Gueron, S.; Shafrir, I. On a Discrete Variational Problem Involving Interacting Particles. *SIAM J. Appl. Math.* **60** (2000), no. 1, 1–17.
- [12] Jaffe, A.; Taubes, C. *Vortices and monopoles*. Progress in Physics, 2. Birkhäuser, Boston, 1980.
- [13] Jerrard, R. L.; Soner, H. M. The Jacobian and the Ginzburg-Landau energy. *Calc. Var. Partial Differential Equations* **14** (2002), no. 2, 151–191.
- [14] Manton, N.; Sutcliffe, P. *Topological solitons*. Cambridge monographs on mathematical physics. Cambridge University Press, 2004.
- [15] Plohr, B. The existence, regularity, and behaviour at infinity of isotropic solutions of classical gauge field theories. Doctoral Dissertation, Princeton University, 1980.
- [16] Plohr, B. The Behavior at Infinity of Isotropic Vortices and Monopoles B. Plohr, *J. Math. Phys.*, **22** (1981), 2184–2190.
- [17] Rankin, A. minimum problem for the Epstein zeta function, *Proc. Glasgow Math. Assoc.* **1**, (1953). 149–158.
- [18] Saint-James, D.; Sarma, G.; Thomas, E. J. *Type-II superconductivity*. Pergamon Press, Oxford, 1969.
- [19] Sandier, E; Serfaty, S. *Vortices in the Magnetic Ginzburg-Landau Model*, Birkhäuser, 2007.
- [20] Sandier, E.; Serfaty, S. On the energy of type-II superconductors in the mixed phase. *Rev. Math. Phys.* **12** (2000), no. 9, 1219–1257.
- [21] Sandier, E.; Serfaty, S. A rigorous derivation of a free-boundary problem arising in superconductivity. *Ann. Sci. École Norm. Sup. (4)* **33** (2000), no. 4, 561–592.
- [22] Sandier, E.; Serfaty, S., From the Ginzburg-Landau Model to Vortex Lattice Problems, to appear.
- [23] Serfaty, S. Local minimizers for the Ginzburg-Landau energy near critical magnetic field, part I. *Comm. Contemp. Math.* **1** (1999), no. 2, 213–254; part II, 295–333.
- [24] Serfaty, S. Stable configurations in superconductivity: uniqueness, multiplicity and vortex-nucleation. *Arch. Ration. Mech. Anal.* **149** (1999), 329–365.
- [25] Tinkham, M. *Introduction to superconductivity*. Second edition. McGraw-Hill, New York, 1996.

QUANTIZATION OF INTEGRABLE SYSTEMS

AND

FOUR DIMENSIONAL GAUGE THEORIES

NIKITA A. NEKRASOV

IHES, Le Bois-Marie, 35 route de Chartres, Bures-sur-Yvette, 91440, France
Simons Center for Geometry and Physics, Stony Brook University, NY 11794 USA
E-mail: nikitastring@gmail.com

SAMSON L. SHATASHVILI

School of Mathematics, Trinity College, Dublin 2, Ireland
Hamilton Mathematics Institute, Trinity College, Dublin 2, Ireland
IHES, Le Bois-Marie, 35 route de Chartres, Bures-sur-Yvette, 91440, France
E-mail: samson@maths.tcd.ie

We study four dimensional $\mathcal{N} = 2$ supersymmetric gauge theory in the Ω -background with the two dimensional $\mathcal{N} = 2$ super-Poincare invariance. We explain how this gauge theory provides the quantization of the classical integrable system underlying the moduli space of vacua of the ordinary four dimensional $\mathcal{N} = 2$ theory. The ϵ -parameter is identified with the Planck constant, the twisted chiral ring maps to quantum Hamiltonians, the supersymmetric vacua are identified with Bethe states of quantum integrable systems. This four dimensional gauge theory in its low energy description has two dimensional twisted superpotential on Σ which becomes the Yang-Yang function of the integrable system. We present the thermodynamic-Bethe-Ansatz like formulae for these functions and the spectra of commuting Hamiltonians following the direct computation in gauge theory. The general construction is illustrated at the examples of the many-body systems, such as the periodic Toda chain, the elliptic Calogero-Moser system, and their relativistic versions, for which we present a complete characterization of the L^2 -spectrum. We very briefly discuss the quantization of Hitchin system.

Keywords: gauge theory; instantons; Bethe Ansatz; many-body systems; finite size corrections; S-matrix

1. Introduction

It has been realized in the recent years [1–5] that there exists an intimate connection between the vacua of the supersymmetric gauge theories and the quantum integrable systems. This connection is quite general and applies to gauge theories in various spacetime dimensions. In the short review articles [4, 5] a large class of the two

dimensional gauge theories were shown to correspond to the finite dimensional spin chains (and their various limits usually studied in literature on integrable models).

We report here on the new developments - we establish the connection between the four dimensional supersymmetric gauge theories and the quantum many-body systems. More precisely, we consider the $\mathcal{N} = 2$ supersymmetric gauge theories in four dimensions. We subject it to the Ω -background [6] in two out of four dimensions. The Ω -background is a particular background of $\mathcal{N} = 2$ supergravity in four dimensions. The general Ω -background is characterized by two complex parameters $\varepsilon_1, \varepsilon_2$. These parameters were introduced in [1, 7] and used in [8] to regularize the integrals over the instanton moduli spaces which arise in the supersymmetric gauge theories and the bound state problems in the supersymmetric quantum mechanics. It was suggested in [7] that by deforming the Donaldson supercharge Q to its equivariant version $Q + V_\varepsilon^\mu G_\mu$ the instanton partition functions would become computable (in fact, an example of an instanton integral was proposed in [7]) and could ultimately test the Seiberg-Witten solution [9] of the four dimensional $\mathcal{N} = 2$ theory. This program was completed in [6]. It turns out that both the ideology and the specific examples of the integral formulae of [7] are important in our current developments related to Bethe Ansatz of the quantum many-body systems.

When both parameters $\varepsilon_1, \varepsilon_2$ of the Ω -background are non-zero the super-Poincare invariance is broken down to a superalgebra with two fermionic and two bosonic generators corresponding to the rotations in $\mathbf{R}^4$. If one of these parameters, e.g. ε_2 vanishes, $\varepsilon_2 = 0$, then the resulting theory has the two dimensional $\mathcal{N} = 2$ super-Poincare invariance. This is the theory we study in the present paper. It is characterized by a single complex parameter, which we shall denote simply by ε .

Our main claim is: *the supersymmetric vacua of this gauge theory are the eigenstates of the quantum integrable system obtained by the quantization of the classical algebraic integrable system underlying the geometry of the moduli space $\mathcal{M}_v$ of undeformed $\mathcal{N} = 2$ theory. The Planck constant, the parameter of the quantization, is identified with the deformation parameter ε .*

Recall that the four dimensional gauge theory with $\varepsilon_1 = \varepsilon_2 = 0$ is characterized, at low energy, by the prepotential $\mathcal{F}(a)$ which is known to be related to some classical algebraic integrable system. For example, the pure $SU(N)$ $\mathcal{N} = 2$ super-Yang-Mills corresponds to the periodic A_{N-1} Toda chain [10]. As shown in [11] the vector multiplet part $\mathcal{M}_v$ of the moduli space of vacua, the Coulomb branch, of any $\mathcal{N} = 2$ supersymmetric gauge theory, is a base of the Liouville fibration of some classical algebraic integrable system. Once the classical integrable system is identified one can write down the gauge theory low energy effective action. When $\varepsilon_1 = \varepsilon \neq 0$, $\varepsilon_2 = 0$ the rôle of $\mathcal{F}(a)$ in the gauge theory is played, in a certain sense, by the two dimensional twisted superpotential $\mathcal{W}(a; \varepsilon)$. We show it also has a meaning in the algebraic integrable system, albeit in the quantum one. It is identified with the Yang-Yang (YY) counting function [12] governing the spectrum of the quantum system. One does not always know a priori how to quantize an algebraic integrable system, just knowing its classical version does not suffice. However, $\mathcal{W}(a; \varepsilon)$ is computable

by the gauge theory methods (and so is $\mathcal{F}(a)$), so the situation got reversed – the gauge theory helps to learn about the quantization of an integrable system.

This is an important part of the general program, whose details will appear in the longer version [13] which in addition (and in particular) will combine the results reported here and those in [4, 5] to form a unified picture.

There are numerous applications of the correspondence [1–5]: the gauge theory applications, the study of quantum cohomology, (infinite-dimensional) representation theory, harmonic analysis, the many-body quantum mechanics.

In a sense the most general non-relativistic algebraic integrable system is the so-called Hitchin system. Its quantization is an interesting and important problem, whose special cases, corresponding to the degenerate Riemann surfaces, are by now old and classical problems. In this paper we illustrate the power of our methods at the examples of two such degenerations, the quantum elliptic Calogero-Moser system, and its limit, the periodic Toda chain. We shall present a complete characterization of the L^2 -spectrum of these systems.

2. Review of the Bethe/gauge correspondence

2.1. Twisted chiral ring and quantum integrability

The space of SUSY vacua of a theory with four supersymmetries carries a representation of a commutative associative algebra, the so-called (twisted) chiral ring, see e. g. book [14].

For example, in two space-time dimensions, the $\mathcal{N} = 2$ supersymmetry is generated by the fermionic charges $Q_{\pm}, \bar{Q}_{\pm}$, which obey the anticommutation relations:

$$\{Q_{\pm}, \bar{Q}_{\pm}\} = 2(H \pm P),$$

$$\{Q_+, Q_-\} = \{\bar{Q}_+, \bar{Q}_-\} = \{Q_+, \bar{Q}_-\} = \{\bar{Q}_+, Q_-\} = 0 \quad (1)$$

the last two lines being valid in the absence of the central extension, induced, e.g. by some global symmetry charges.

The twisted chiral ring is generated by the operators $\mathcal{O}_k$, $k = 1, 2, \dots$ which (anti)-commute with the operator $\mathcal{Q}_A = Q_+ + \bar{Q}_-$; $\{\mathcal{Q}_A, \mathcal{Q}_A^{\dagger}\} = H$. Analogously one defines the chiral ring, whose generators (anti)commute with the operator: $\mathcal{Q}_B = Q_+ + Q_-$; $\{\mathcal{Q}_B, \mathcal{Q}_B^{\dagger}\} = H$. In our work we concentrate on the $\mathcal{Q}_A$ -cohomology and assume that the possible central extension of (1) leaves $\mathcal{Q}_A$ nilpotent $\mathcal{Q}_A^2 = 0$. The local operators $\mathcal{O}_k(x)$ are independent up to the $\mathcal{Q}_A$ -commutators of their location x . Their operator product expansion defines a commutative associative ring,

$$\mathcal{O}_i \mathcal{O}_j = c_{ij}^k \mathcal{O}_k + \{\mathcal{Q}_A, \dots\} \quad (2)$$

If $|0\rangle$ is a vacuum state of the Hamiltonian H , $H|0\rangle = 0$, then so is $\mathcal{O}_i|0\rangle = |i\rangle$, and moreover the space of vacua is the representation of the twisted chiral ring.

Thus the space of supersymmetric vacua, which can be effectively studied using the cohomology of the operator $\mathcal{Q}_A$ (or $\mathcal{Q}_B$), is the space of states of some quantum

integrable system:

$$\mathbf{H}^{\text{quantum}} = \ker \mathcal{Q}_A / \text{im } \mathcal{Q}_A \quad (3)$$

The operators $\mathcal{O}_k$ and more generally the functions of $\mathcal{O}_k$'s are the quantum Hamiltonians. These operators can be chosen *under the assumption of the absence of massless charged matter fields* to be the gauge invariant polynomials of the complex scalar σ in the vector supermultiplet,

$$\mathcal{O}_k = \frac{1}{k!(2\pi i)^k} \text{Tr } \sigma^k \quad (4)$$

One is looking for the common eigenstates of these Hamiltonians:

$$\Psi_\lambda \in \mathbf{H}^{\text{quantum}}, \quad \mathcal{O}_k \Psi_\lambda = E_k(\lambda) \Psi_\lambda \quad (5)$$

where $E_k(\lambda)$ are the corresponding eigenvalues, and λ are some labels. In general they are complex, $E_k(\lambda) \in \mathbf{C}$.

The important, or at least the interesting, problem is to identify the quantum integrable system given an $\mathcal{N} = 2$ gauge theory, or to solve the converse problem – to find the $\mathcal{N} = 2$ theory given a quantum integrable system. For a large class of models on both sides this problem has been solved in [1–5].

In most interesting cases the supersymmetric gauge theory at low energies has an effective two dimensional abelian gauge theory description (with four supercharges), the so-called theory on the Coulomb branch. The supersymmetric vacua are determined in terms of the exactly calculable effective twisted superpotential $\tilde{W}^{\text{eff}}(\sigma)$. Loosely speaking, given a vector of electric fluxes $(n_1, \dots, n_r)$, with $n_i \in \mathbf{Z}$ the vacua are given by the critical points of the shifted superpotential $\tilde{W}^{\text{eff}}(\sigma) - 2\pi i \sum_{i=1}^r n_i \sigma^i$:

$$\frac{1}{2\pi i} \frac{\partial \tilde{W}^{\text{eff}}(\sigma)}{\partial \sigma^i} = n_i; \quad \Leftrightarrow \quad \exp \left(\frac{\partial \tilde{W}^{\text{eff}}(\sigma)}{\partial \sigma^i} \right) = 1 \quad (6)$$

as follows from the consideration of the effective potential

$$U_{\text{ecn}}(\sigma) = \frac{1}{2} g^{ij} \left(-2\pi i n_i + \frac{\partial \tilde{W}^{\text{eff}}}{\partial \sigma^i} \right) \left(+2\pi i n_j + \frac{\partial \tilde{W}^{\text{eff}}}{\partial \sigma^j} \right) \quad (7)$$

2.2. Topological field theory

The two dimensional $\mathcal{N} = 2$ gauge theory can be topologically twisted. In the twisted version the supercharge $\mathcal{Q}_A$ plays the rôle of the BRST operator. The vacua of the physical theory are the physical states of the topological theory.

The action of the topological field theory can be brought to the simple form

$$S = \sum_{i=1}^r \int_{\Sigma} \frac{\partial \tilde{W}^{\text{eff}}(\sigma)}{\partial \sigma_i} F_A^i + \frac{1}{2} \sum_{i,j=1}^r \int_{\Sigma} \frac{\partial^2 \tilde{W}^{\text{eff}}(\sigma)}{\partial \sigma_i \partial \sigma_j} \psi^i \wedge \psi^j \quad (8)$$

by the so-called quartet mechanism (one adds the anti-twisted superpotential $\bar{t} \text{Tr } \bar{\sigma}^2$, and sends $\bar{t} \rightarrow \infty$). Here F_A^i is a curvature of abelian gauge field A^i and

ψ^i is anti-commuting 1-form on Σ , the super-partner of A^i . In this form the gauge theory becomes a contour integral over the σ field (the field $\bar{\sigma}$ being eliminated). The original supersymmetric theory contains information hidden in the so-called D -terms, which ultimately leads to the wall-crossing phenomena [15].

The canonical quantization of the theory (8) on the cylinder $\Sigma = \mathbf{R} \times \mathbf{S}^1$ is simple. Indeed, the only physical degree of freedom of this theory is the monodromy $\exp i\vartheta^i = \exp \oint_{\mathbf{S}^1} A^i$ of the gauge field around the circle $\mathbf{S}^1$ and the momentum conjugate to ϑ^i , $I_i = \partial \tilde{W}^{\text{eff}} / \partial \sigma_i$. Since ϑ^i takes values in a circle (due to the large gauge transformations) I_i is quantized, leading to the equations (6).

Our conclusion is that one can study the vacuum sector of the $\mathcal{N} = 2$ gauge theory or one can study the topologically twisted version of the same theory – ultimately one deals with the Bethe states of some quantum integrable system. This correspondence benefits all three subjects involved.

2.3. Yang-Yang function and quantum spectrum

In [2, 3] the form (8) of the effective action and the quantization argument above for the two dimensional theory studied in [1] was used to identify $\tilde{W}^{\text{eff}}(\sigma)$ with the YY function of the N -particle Yang integrable system. However both the quantization argument and the identification of $\tilde{W}^{\text{eff}}$ with a YY function are not restricted to this case only.

It is a remarkable feature of the quantum integrable systems that their spectrum can be sometimes studied using Bethe Ansatz. The spectrum of the quantum system is determined by the equations on the (quasi)momentum variables (rapidities) λ_i , which enter the parametrization of the eigenfunctions. It is even more remarkable that the quasimomenta are determined by the equations which have a potential:

$$\frac{1}{2\pi i} \frac{\partial Y(\lambda)}{\partial \lambda_i} = n_i \in \mathbf{Z} \quad (9)$$

The function $Y(\lambda)$ is called the counting, or YY function [12], and (9) is the corresponding Bethe equation [16].

It has been demonstrated in [4, 5] that for many interesting gauge theories in two, three (compactified on $\mathbf{S}^1$) or four (compactified on $\mathbf{T}^2$) dimensions the equation (6) coincides with Bethe equation, for a large class of interesting quantum integrable systems. The effective twisted superpotential of the gauge theory equals the YY function of the quantum integrable system (e. g. for $G = U(N)$):

$$\tilde{W}^{\text{eff}}(\sigma) = Y(\lambda); \quad \sigma_i = \lambda_i, \quad i = 1, \dots, N \quad (10)$$

when the Coulomb branch moduli σ_i are identified with the spectral parameters λ_i of the quantum integrable system. The expectation values of the twisted chiral ring operators $\mathcal{O}_k$ in the vacuum $|\lambda\rangle$ given by the solution $\sigma = \lambda$ of (6) coincide with the eigenvalues of quantum Hamiltonians of integrable system:

$$\langle \lambda | \mathcal{O}_k | \lambda \rangle = \mathcal{E}_k(\lambda); \quad \mathbf{H}_k \Psi_\lambda = \mathcal{E}_k(\lambda) \Psi_\lambda \quad (11)$$

Most of these models of [4, 5] are intrinsically quantum. Their Hilbert spaces are finite dimensional. The Planck constant in these cases is not a continuous variable. These theories are not very natural from the point of view of the quantization procedure, where one is given a classical integrable system and is asked to construct the quantum integrable system.

2.4. Quantization from four dimensions

In the current paper we study a novel type of theories, which originate in four dimensions. They have a continuous parameter ε , which becomes the continuous Planck constant of a quantum integrable system. The corresponding quantum integrable systems have infinite-dimensional Hilbert spaces. For example, we shall give a solution to the quantum periodic Toda chain (pToda), an elliptic Calogero-Moser system (eCM) and their relativistic analogues. We shall also make some remarks on the general Hitchin system. Our strategy is the following:

- (1) Start with a four dimensional $\mathcal{N} = 2$ gauge theory (for example, one may take a pure $\mathcal{N} = 2$ theory with some gauge group G or the $\mathcal{N} = 2^*$ theory, the theory with one massive hypermultiplet in the adjoint representation); its low energy effective Lagrangian is determined in terms of a single multi-valued analytic function of the Coulomb moduli $(a_1, \dots, a_r)$, $r = \text{rank}(G)$, called the prepotential, $\mathcal{F}(a; m, \tau)$. Here m, τ etc. are the parameters of the gauge theory.
- (2) The theory in the ultra-violet has the observables $\mathcal{O}_k$

$$\mathcal{O}_k = \frac{1}{(2\pi i)^k k!} \text{Tr} \phi^k$$

where ϕ is the complex scalar in the vector multiplet. The observables $\mathcal{O}_k$ correspond to the holomorphic functions u_k on $\mathcal{M}_V$. These observables are singled out by their (anti-)commutation with the supercharge $\mathcal{Q}_A$ (this supercharge becomes the Q -operator of the Donaldson-Witten theory in the standard $\mathcal{N} = 2$ twist). One can formally deform the theory in the ultra-violet

$$\mathcal{F}^{\text{tree}} \rightarrow \mathcal{F}^{\text{tree}} + \sum_k t_k \mathcal{O}_k \quad (12)$$

In the low energy the deformed theory is given by the family of prepotentials $\mathcal{F}(a; \mathbf{t}; m, \tau)$, $\mathbf{t} = (t_k)$. For some models the deformed prepotentials were computed in [15, 17, 18]. For our purposes the t_k deformation can be studied formally, i.e. without addressing the convergence issues. However, some of the couplings in $\mathbf{t}$ can be interpreted as shifting the ultraviolet complexified gauge coupling τ . For these couplings a finite deformation can be studied. For other couplings only the first order deformation is needed (and therefore the contact term problem [15] does not arise). The deformed prepotential $\mathcal{F}(a; \mathbf{t}; m, \tau)$ generates the vevs of $\mathcal{O}_k$'s: $\frac{\partial \mathcal{F}(a; \mathbf{t}; m, \tau)}{\partial t_k} = \langle \mathcal{O}_k \rangle_a$.

- (3) The prepotential $\mathcal{F}(a, m, \tau)$ has an interpretation in terms of a classical algebraic integrable system where it plays the rôle of the generating function of a

Lagrangian submanifold L of a symplectic complex vector space relating two types of complex action variables, a^i and $a_{D,i}$. The $\mathbf{t}$ -couplings deform L to another Lagrangian submanifold $L_{\mathbf{t}}$ [15].

- (4) Finally, we introduce one more deformation parameter, ε , which corresponds to subjecting the theory to the Ω -background $\mathbf{S}^1 \times \mathbf{R}^1 \times \mathbf{R}_{\varepsilon}^2$, or $\mathbf{R}^2 \times \mathbf{R}_{\varepsilon}^2$ (see the next subsection). The F -terms of the low energy Lagrangian are now two dimensional with the twisted superpotential which we denote by $\mathcal{W}(a; \mathbf{t}; \varepsilon)$. For the discussion of the supersymmetric vacua only these twisted F -terms are relevant, therefore the effective description of the low energy physics is two dimensional. In particular, one derives the equation determining the vacua and spectrum (of twisted chiral observables), as

$$\frac{\partial \mathcal{W}(a, \mathbf{t}, \varepsilon)}{\partial a_i} = 2\pi i n_i; \quad \mathcal{E}_k = \left. \frac{\partial \mathcal{W}(a, \mathbf{t}; \varepsilon)}{\partial t_k} \right|_{\mathbf{t}=0}; \quad n_i \in \mathbf{Z} \quad (13)$$

We shall note that the equation (13) can be written, in the examples studied in this paper and probably in more general situations, in terms of the factorized S -matrix of the associated “hyperbolic” many body system plus “finite-size” corrections (e. g. corrections in $e^{i\tau}$ in case of $\mathcal{N} = 2^*$, in Λ for pure $\mathcal{N} = 2, \dots$):

$$\frac{\partial \mathcal{W}(a, \mathbf{t} = 0, \varepsilon)}{\partial a_i} = i \frac{\tau a_i}{\varepsilon} + \sum_{j \neq i} \log S(a_i - a_j) + \text{finite size} \quad (14)$$

- (5) As in [1–5] the equations (13) determine the spectrum of the quantum integrable system, where ε plays the rôle of the (complexified) Planck constant. We identify this system with the quantization of the classical algebraic integrable system describing the low energy effective theory of the four dimensional $\mathcal{N} = 2$ theory.

Remark. Note the unfortunate notational conflict. The order parameters, σ_i , the eigenvalues of the complex scalar in the vector multiplet, are denoted traditionally by σ_i in the context of two dimensional gauge theories, by a^i in the context of four dimensional gauge theories, by ϕ_i in the context of topological gauge theories. The same parameters exhibit themselves as the Bethe roots in our correspondence with the quantum integrable systems. In that world they are denoted by λ_i . In the context of the periodic Toda chain these variables are denoted by δ_i in [19], and by t_α in [20]. We failed to propose a unified notation and followed the traditional way of denoting the same object by different letters corresponding to the different context. We hope the reader will not be too confused by this.

3. Four dimensional gauge theory

3.1. The Ω -background and twisted masses

A quantum field theory in $k + 2$ space-time dimensions can be viewed, formally, as a two dimensional theory with an infinite number of fields. If this theory is studied on flat $k + 2$ dimensional space-time, or on a space-time fibered over a

two-dimensional manifold Σ with the Euclidean fibers $\mathbf{R}^k$, it has (in addition to the global symmetries of the $k+2$ -dimensional theory) a global symmetry group $\mathbf{E}(k)$ of isometries of the k -dimensional Euclidean space $\mathbf{R}^k$. Accordingly, if the theory has an $\mathcal{N} = 2$ supersymmetry on Σ , then one may deform it by turning on the twisted masses corresponding to the global symmetry $\mathbf{E}(k)$. Unlike the conventional global symmetries which typically form a compact Lie group with the unique, up to a conjugation, maximal torus, the group $\mathbf{E}(k)$ has several, for $k > 1$, inequivalent Cartan subgroups. Thus there exists several physically inequivalent deformations of the $k+2$ dimensional theory.

For example, one may choose a subgroup $\mathbf{R}^{k-2l} \times SO(2)^l$ of translations in $\mathbf{R}^{k-2l}$ and rotations in the l orthogonal two-planes. The theory with twisted masses corresponding to the translations in $\mathbf{R}^{k-2l}$ is equivalent to the ordinary Kaluza-Klein compactification on a torus $\mathbf{T}^{k-2l}$. Such theories are studied in [4, 5].

3.1.1. Four dimensional theory on $\mathbf{R}_\varepsilon^2$

Another possibility is to consider the twisted masses corresponding to the rotational symmetry. In this case one gets the theory in the Ω -background [6].

Consider a four dimensional $\mathcal{N} = 2$ theory on a four manifold M^4 fibered over a two dimensional base Σ with the Ω -background along the fibers $\mathbf{R}^2$. Somewhat schematically we shall denote the fibers by $\mathbf{R}_\varepsilon^2$. The base Σ of the fibration could be a two-plane $\mathbf{R}^{1,1}$ or a cylinder $\mathbf{S}^1 \times \mathbf{R}^1$. One can also study the twisted theory for which Σ could be an arbitrary Riemann surface.

In our main examples, corresponding to the periodic Toda and elliptic Calogero-Moser systems our starting point would be the pure $\mathcal{N} = 2$ super-Yang-Mills theory, or the $\mathcal{N} = 2^*$ theory, corresponding to the gauge theory with a single adjoint massive hypermultiplet. In the limit of vanishing mass of the adjoint matter fields the latter becomes an $\mathcal{N} = 4$ theory.

Now let us give some details on the Ω -deformation. Let us denote the coordinates on the fiber $\mathbf{R}_\varepsilon^2$ by (x^2, x^3) , and the coordinates on the base Σ by (x^0, x^1) . Introduce the vector field $U = x^2 \partial_3 - x^3 \partial_2$ generating the $U(1)$ rotation in $\mathbf{R}^2$. Let $\varepsilon \in \mathbf{C}$ be a complex parameter and let $V = \varepsilon U, \bar{V} = \bar{\varepsilon} U$ be the complex vector fields on $\mathbf{R}^2$.

The bosonic part of the pure $\mathcal{N} = 2$ super-Yang-Mills Lagrangian on $\mathbf{R}_\varepsilon^2$ is:

$$\begin{aligned} L = & -\frac{1}{4g_0^2} \text{tr} F \wedge \star F + \text{tr} (D_A \phi - \varepsilon \iota_U F) \wedge \star (D_A \bar{\phi} - \bar{\varepsilon} \iota_U F) + \\ & + \frac{1}{2} \text{tr} ([\phi, \bar{\phi}] + \iota_U D_A (\varepsilon \bar{\phi} - \bar{\varepsilon} \phi))^2 + \frac{\theta_0}{2\pi} \text{tr} F \wedge F \end{aligned} \quad (15)$$

It is clear that the Poincare invariance in the (x^0, x^1) directions is unbroken, and it is possible to show that in fact the two dimensional $\mathcal{N} = 2$ super-Poincare invariance is preserved. Thus there are four supercharges.

The only (twisted) F -terms of the low energy effective theory are two-dimensional and can be represented as the non-trivial twisted superpotential $\mathcal{W}(a; \varepsilon)$

(where as before a denotes the complex scalar in abelian vector multiplet).

If we send ε back to zero then the low energy theory is fully four-dimensional with a continuous moduli space $\mathcal{M}_v$ of vacua and the low energy effective Lagrangian is described in terms of the prepotential $\mathcal{F}(a)$.

For non-zero ε the theory has a discrete set of vacua given by the minima of the potential (7). These vacua are therefore the solutions to the equation (13), which we shall later on identify with the Bethe equation of some quantum integrable system.

3.1.2. Calculation of the twisted superpotential

Our discussion would have had a rather limited significance were it not for the possibility of exact computation of $\mathcal{W}(a; \varepsilon)$. Let us briefly explain our strategy. Consider the four dimensional theory in the general Ω -background, with both rotation parameters $\varepsilon_1, \varepsilon_2$ non-zero. Then the effective theory has a prepotential $\mathcal{F}(a, \varepsilon_1, \varepsilon_2)$ which is analytic in $\varepsilon_1, \varepsilon_2$ near zero and becomes exactly the prepotential of the low energy effective four dimensional theory in the limit $\varepsilon_1, \varepsilon_2 \rightarrow 0$:

$$S_{4d}^{\text{eff}} = \int \mathcal{F}^{(4)}(a, \varepsilon_1, \varepsilon_2) + \{Q, \dots\} \quad (16)$$

where we denote collectively by a all the vector multiplet scalars as well as background scalars, such as the masses of matter fields. At the same time, had we started with a two dimensional theory which is characterized by some twisted superpotential $\mathcal{W}(a)$, and had we subjected it to the two dimensional Ω -background with the parameter ε_2 , the effective action would have had the form:

$$S_{2d}^{\text{eff}} = \int \mathcal{W}^{(2)}(a, \varepsilon_2) + \{Q, \dots\} \quad (17)$$

where $\mathcal{W}(a, \varepsilon_2) \rightarrow \mathcal{W}(a)$ as $\varepsilon_2 \rightarrow 0$, and Q is a certain supercharge which we use to study the vacuum states of our theory. In the Eqs. (16), (17) the notations $F^{(4)}$, $W^{(2)}$ refer to the cohomological descendents of the local operators F , W , etc. Now the standard manipulations with equivariant cohomology give:

$$\int_{\mathbf{R}^4} \mathcal{F}^{(4)}(a, \varepsilon_1, \varepsilon_2) = \frac{1}{\varepsilon_1} \int_{\mathbf{R}^2} \mathcal{F}^{(2)}(a, \varepsilon_1, \varepsilon_2) = \frac{1}{\varepsilon_1 \varepsilon_2} \mathcal{F}(a, \varepsilon_1, \varepsilon_2) \quad (18)$$

modulo Q -exact terms at each step. By carefully manipulating the Q -exact terms we can connect the computation of the gauge theory partition function in the ultraviolet, which is given by a one-loop perturbative and a series of exact instanton corrections, to the computation in the infrared, using the Wilsonian effective action, where the effective energy scale can be sent all the way to zero:

$$\exp \frac{1}{\varepsilon_1 \varepsilon_2} \mathcal{F}(a, \varepsilon_1, \varepsilon_2; q) = \mathcal{Z}(a, \varepsilon_1, \varepsilon_2; q) = \mathcal{Z}^{\text{pert}}(a, \varepsilon_1, \varepsilon_2; q) \times \mathcal{Z}^{\text{inst}}(a, \varepsilon_1, \varepsilon_2; q) \quad (19)$$

where we restored the dependence on the complexified bare gauge coupling

$$q = \exp 2\pi i \tau \quad \tau = \frac{\vartheta}{2\pi} + \frac{4\pi i}{g^2}$$

Now, by comparing the Eqs. (16), (17), (19) we conclude:

$$\mathcal{W}(a, \varepsilon; q) = \text{Limit}_{\varepsilon_2 \rightarrow 0} [\varepsilon_2 \log \mathcal{Z}(a, \varepsilon_1 = \varepsilon, \varepsilon_2; q)] = \mathcal{W}^{\text{pert}}(a, \varepsilon; q) + \mathcal{W}^{\text{inst}}(a, \varepsilon; q) \quad (20)$$

Here the instanton part has an expansion in the powers of q : $\mathcal{W}^{\text{inst}}(a, \varepsilon; q) = \sum_{k=1}^{\infty} q^k \mathcal{W}_k^{\text{inst}}(a, \varepsilon)$, while the perturbative part has a tree level term, proportional to $\log(q)$ and the one-loop term, which is q -independent.

It follows that in the limit $\varepsilon \rightarrow 0$ the twisted superpotential $\mathcal{W}(a, \varepsilon)$ behaves as:

$$\mathcal{W}(a, \varepsilon; q) = \frac{\mathcal{F}(a; q)}{\varepsilon} + \dots \quad (21)$$

with $\dots$ denoting the regular in ε terms, and now the equations on the supersymmetric vacua assumes the *Bethe* (13) form with this superpotential (21).

Our course is now pretty much set. We shall use the gauge theory knowledge of the instanton partition functions $\mathcal{Z}(a, \varepsilon_1, \varepsilon_2; q)$ to extract, via (20), the twisted superpotential $\mathcal{W}(a, \varepsilon; q)$. The details of this procedure are reviewed in the **Section 6** where various ways of writing $\mathcal{Z}(a, \varepsilon_1, \varepsilon_2; q)$ and extracting $\mathcal{W}(a, \varepsilon; q)$ are presented. For the purposes of the current discussion, the function $\mathcal{W}(a, \varepsilon; q)$ is known and the logic above gives the desired equation for the supersymmetric vacua Bethe form.

By the philosophy of [1–5] this superpotential serves as the YY function of some quantum integrable system. In the present context the parameter ε plays the rôle of the Planck constant, and can be tuned to zero. In this limit we shall be able to use the quasiclassical asymptotics (21) to identify the classical integrable system, whose quantization (in ε) is the quantum integrable system in question.

Now we remind the rôle of prepotential in the world of classical integrability [21].

Remark 1. Note that in the derivation [22] of the Seiberg-Witten prepotential $\mathcal{F}(a)$ from the direct instanton counting one evaluates the small $\varepsilon_1, \varepsilon_2 \rightarrow 0$ asymptotics of $\mathcal{Z}(a, \varepsilon_1, \varepsilon_2; q)$ by a discrete version of the saddle point method, which connects nicely the theory of instanton integrals to the theory of limit shapes and random geometries. In our story we need to go beyond that analysis, see **Section 6**. Our results suggest that the “quantum theory of the limit shape” is related to the thermodynamic Bethe Ansatz [12, 23–27].

Remark 2. The simplest case of the Ω -background is in two dimensions. The two dimensional Ω -background is characterized by the single parameter ε . The partition function $I_{\alpha\beta}(\varepsilon)$ depends on two (discrete) parameters: the choice β of the boundary condition at infinity, and the choice of the twisted chiral ring operator $\mathcal{O}_\alpha$ inserted at the origin. These partition functions were studied in the context of the (equivariant) Gromov-Witten theory [28], where the ε -dependence comes from the coupling to the two dimensional topological gravity. When the theory is deformed in a way analogous to (12), the partition function becomes a matrix-valued function $I_{\alpha\beta}(\mathbf{t}; \varepsilon)$ which solves the *quantum differential equation*:

$$\varepsilon \frac{\partial}{\partial t^\gamma} I_{\alpha\beta}(\mathbf{t}; \varepsilon) = C_{\alpha\gamma}^\kappa(\mathbf{t}) I_{\kappa\beta}(\mathbf{t}; \varepsilon) \quad (22)$$

where $C_{\alpha\beta}^\gamma$ are the structure constants of the twisted chiral ring.

4. Integrable systems

In this section we remind a few relevant notions in the theory of integrable systems and in particular explain the role of prepotential in classical algebraic integrable system. We introduce our main examples and t -deformations along the lines of [15].

4.1. The classical story

The classical Hamiltonian integrable system is the collection $(\mathcal{P}, \omega, \mathbf{H})$, where $\mathcal{P}$ is a $2n$ -dimensional smooth manifold endowed with the non-degenerate closed two-form ω and a collection $\mathbf{H} = (H_1, H_2, \dots, H_n)$ of (generically) functionally independent functions $\mathbf{H} : \mathcal{P} \rightarrow \mathbf{R}^n$, which mutually Poisson-commute: $\{H_i, H_j\} = 0$, $i \neq j$. Here $\{A, B\} = (\omega^{-1})^{\mu\nu} \partial_\mu A \partial_\nu B$. We can view $\mathcal{P}$ as a Lagrangian fibration $\mathbf{H} : \mathcal{P} \rightarrow \mathcal{U} \subset \mathbf{R}^n$.

The classical Liouville-Arnold theorem states that if the common level set $\mathbf{H}^{-1}(h)$ is compact, then it is diffeomorphic to the n -dimensional torus $\mathbf{T}^n$. Moreover, if $\mathbf{H}^{-1}(h)$ is compact for any h in a neighborhood $\mathcal{U}$ of a point $h_0 \in \mathbf{R}^n$, then $\mathbf{H}^{-1}(\mathcal{U})$ is symplectomorphic to the neighborhood of a zero section in $T^*\mathbf{T}^n$. One can then find the special Darboux coordinates $(\mathbf{I}, \varphi)$, $\mathbf{I} = (I_1, I_2, \dots, I_n)$, $\varphi = (\varphi_1, \dots, \varphi_n)$, called the *action-angle* variables, s.t. the Hamiltonians H_i , $i = 1, \dots, n$, depend only on $\mathbf{I}$, $H_i(\mathbf{I})$, while φ_i are the periodic angular coordinates on $\mathbf{T}^n$ with the period 2π . The *action* variables are given by the periods

$$I_i = \frac{1}{2\pi} \oint_{A_i} \mathbf{p} d\mathbf{q} \quad (23)$$

of the one-form $\mathbf{p} d\mathbf{q} = d^{-1}\omega$ (one can give a more invariant definition), over some $\mathbf{Z}$ -basis in $H_1(\mathbf{T}^n, \mathbf{Z})$.

The notion of the classical *real* integrable system has an interesting *complex* analogue, sometimes known as the *algebraic integrable system*. The data $(\mathcal{P}, \omega, \mathbf{H})$ now consists of the complex manifold $\mathcal{P}$, the holomorphic non-degenerate closed $(2, 0)$ form ω , and the holomorphic map $\mathbf{H} : \mathcal{P} \rightarrow \mathbf{C}^n$ whose fibers $\mathcal{J}_h = \mathbf{H}^{-1}(h)$ are Lagrangian polarized abelian varieties. The polarization is a Kähler form ϖ , whose restriction on each fiber is an integral class $[\varpi] \in H^2(J_h, \mathbf{Z}) \cap H^{1,1}(J_h)$. The image $\mathcal{B} = \mathbf{H}(\mathcal{P})$ is an open domain in $\mathbf{C}^n$. It has a special Kähler geometry, with the metric

$$ds^2 = \frac{1}{\pi} \sum_{i=1}^n \text{Im} (da^i \otimes d\bar{a}_{D,i}) \quad (24)$$

where the *special coordinates* a^i , $a_{D,i}$ are given by the periods:

$$a^i = \frac{1}{2\pi} \oint_{A_i} \mathbf{p} d\mathbf{q}, \quad a_{D,i} = \frac{1}{2\pi} \oint_{B_i} \mathbf{p} d\mathbf{q} \quad (25)$$

over the A and B -cycles, which are the Lagrangian (with respect to the intersection form given by $[\varpi]$) subspaces in $H_1(J_h, \mathbf{Z})$. It follows that the two-form $\sum_i da^i \wedge da_{D,i}$ vanishes on $\mathcal{B}$ thereby embedding the covering $\mathcal{U}$ of the complement $\mathcal{B} \setminus \Sigma$ to

the discriminant $\Sigma \subset \mathcal{B}$ of the singular fibers to the first cohomology $H^1(J_{h_0}, \mathbf{C})$ of the fiber over some distinguished point $h_0 \in \mathcal{B}$, as a Lagrangian submanifold $\mathcal{L}$. As such, it comes with the function $\mathcal{F} : \mathcal{L} \rightarrow \mathbf{C}$ which can be locally viewed as a function of a^i , such that

$$a_{D,i} = \frac{\partial \mathcal{F}}{\partial a^i} \quad (26)$$

The comparison of the Eqs. (25) and (23) suggests both a^i and $a_{D,i}$ are the *complex action variables*. Since the $2n$ -dimensional symplectic manifold has at most n functionally independent Poisson-commuting functions, there ought to be a relation between a^i and $a_{D,i}$'s. It is remarkable that this relation has a potential function. The action variables a^i come with the corresponding angle variables $\phi_i = \alpha_i + \tau_{ij}\beta^j$, $\tau_{ij} = \partial_{ij}^2 \mathcal{F}$, while $a_{D,i}$ correspond to $\phi^{D,i} = (\tau^{-1})^{ij} \phi_j$:

$$\omega = \sum_i da^i \wedge d\phi_i = \sum_i da_{D,i} \wedge d\phi^{D,i} \quad (27)$$

Here $\alpha_i, \beta^i \in \mathbf{R}/2\pi\mathbf{Z}$ are the real angular coordinates on the Liouville torus.

4.1.1. The $\mathbf{t}$ -deformation

An algebraic classical integrable system $(\mathcal{P}, \omega, \mathbf{H})$ can be deformed in the following way. Consider a family $(\mathcal{P}_{\mathbf{t}}, \omega_{\mathbf{t}}, \mathbf{H}_{\mathbf{t}})$ of complex symplectic manifolds with the Lagrangian fibration given by the “Hamiltonians” $\mathbf{H}_{\mathbf{t}} : \mathcal{P}_{\mathbf{t}} \rightarrow \mathbf{C}^n$, over a (formal) multidimensional disk $\mathcal{D}$, parameterized by $\mathbf{t}$. Then the variation of the symplectic form $\omega_{\mathbf{t}}$ in $\mathbf{t}$ can be described by the equation:

$$\frac{\partial}{\partial t_k} \omega_{\mathbf{t}} = \mathcal{L}_{V_k} \omega^{\mathbf{R}}; \quad \omega^{\mathbf{R}} = \sum_{i=1}^n d\alpha_i \wedge d\beta^i + \sum_{i,j} \text{Im} \tau_{ij} da^i \wedge d\bar{a}^j \quad (28)$$

is the Kähler $(1,1)$ -form; V_k is the holomorphic Hamiltonian (in the $\omega_{\mathbf{t}}$ symplectic structure) vector field corresponding to the Hamiltonian H_k of the original system.

4.1.2. Quantization

The quantization of the classical integrable system is a (possibly discrete) family $(\mathcal{A}_{\varepsilon}, \mathcal{H}_{\varepsilon}, \hat{\mathbf{H}})$, of the associative algebras $\mathcal{A}_{\varepsilon}$, which deform the algebra of functions on the Poisson manifold (X, ω^{-1}) , the (Hilbert) vector spaces $\mathcal{H}_{\varepsilon}$, with the action of $\mathcal{A}_{\varepsilon}$, and the operators $\hat{\mathbf{H}} = (\hat{H}_1, \dots, \hat{H}_n)$, $\hat{H}_i \in \mathcal{A}_{\varepsilon}$, which mutually commute $[\hat{H}_i, \hat{H}_j] = 0$, and generate $\mathcal{H}_{\varepsilon}$ in the following sense: *the common spectral problem* $\hat{H}_i \Psi = \mathcal{E}_i \Psi$ *defines a basis in* $\mathcal{H}_{\varepsilon}$. Here ε has the meaning of the Planck constant.

The construction of the common eigenstates and the spectrum of the operators $\hat{H}_i$ is a problem of the coordinate Bethe Ansatz, quantum inverse scattering method (Algebraic Bethe Ansatz [29–31]), quantum separation of variables [32], Baxter equation [33] and various other versions of the Bethe Ansatz.

In this paper we take a different route – via the supersymmetric gauge theory, along the lines of [1–5]. Gauge theory allows to find the exact spectrum of common eigenstates which is an invariant information independent of what coordinates we use in quantization procedure.

In our story the algebra $\mathcal{A}_\varepsilon$ is the deformation of the algebra of holomorphic functions on $\mathcal{P}$. We shall assume the existence of the global coordinates p_i, x_i (we do not assume them to be the globally defined holomorphic functions on $\mathcal{P}$, but on some covering space). The algebra $\mathcal{A}_\varepsilon$ is generated by $\hat{p}_i, \hat{x}_i$, obeying $[\hat{p}_i, \hat{x}_j] = \varepsilon \delta_{ij}$, $\mathcal{H}_\varepsilon$ is defined as the space of appropriate holomorphic functions of x_i , and the representation of $\mathcal{A}_\varepsilon$ is given by: $\hat{x}_i = x_i$, $\hat{p}_i = \varepsilon \frac{\partial}{\partial x_i}$. The next question is the construction of the Hilbert space where $\mathcal{A}_\varepsilon$ is represented. If $\mathcal{P}$ were a cotangent bundle to a complex manifold M , the algebra $\mathcal{A}_\varepsilon$ would be isomorphic to the algebra of holomorphic differential operators on M . However, it is rarely the case that there are interesting holomorphic differential operators which are defined globally on M . Moreover, the naive complexification of the quantization of the quantization of $T^*M_{\mathbf{R}}$ (which produces the differential operators on $M_{\mathbf{R}}$ acting in the space of half-densities $L^2(M_{\mathbf{R}}; K_M^{1/2})$) produces the $K_M^{1/2}$ – twisted differential operators. It may well happen that the space of global sections of the $K_M^{1/2}$ – twisted differential operators is non-zero yet the space of global sections of $K_M^{1/2}$ where these operators would have acted is empty. This is the situation with the quantization of Hitchin system as discussed in [34].

In our story the algebra $\mathcal{A}_\varepsilon$ is the noncommutative deformation of the algebra of holomorphic functions on $\mathcal{P}$, yet it is represented in the regular L^2 -sections of some line bundle on a real middle dimensional submanifold $\mathcal{P}_{\mathbf{R}}$ of $\mathcal{P}$. The choice of $\mathcal{P}_{\mathbf{R}}$ is apparently made by the boundary conditions in the gauge theory. We do not have a complete understanding of this issue yet, but let us make an important:

Remark: The equation (6) and the discussion in **Section 2** show that in our approach we quantize, for the type A model (defined below) the real submanifold which projects on the locus $\text{Re}(\partial \mathcal{W}(a)/\partial a^i) = 0$ on the base, and cuts out a middle dimensional real torus in the Liouville fiber; for the type B model it projects on the locus $\text{Im}(a^i/\varepsilon) = 0$.

5. Examples

5.1. The periodic Toda chain

5.1.1. The classical system

The periodic Toda chain is the system of N particles $q_1, \dots, q_N$ on the real line interacting with the potential:

$$U(x_1, \dots, x_N) = \Lambda^2 \left(\sum_{i=1}^{N-1} e^{x_i - x_{i+1}} + e^{x_N - x_1} \right) \quad (29)$$

The phase space of this model is $\mathcal{P}_{\mathbf{R}} = T^*\mathbf{R}^N$, with the coordinates $(p_i, x_i)_{i=1}^N$, where $p_i, x_i \in \mathbf{R}$, the symplectic form $\omega = \sum_{i=1}^N dp_i \wedge dx_i$, and the Hamiltonians

$$H_1 = \sum_i p_i, \quad H_2 = \frac{1}{2} \sum_i p_i^2 + U(x_1, \dots, x_N), \quad \dots \quad H_k = \frac{1}{k!} \sum_i p_i^k + \dots \quad (30)$$

The complexified Toda chain has the phase space $\mathcal{P} = T^*(\mathbf{C}^\times)^N$, with the coordinates $(p_i, x_i)_{i=1}^N$ where $p_i \in \mathbf{C}, x_i \in \mathbf{C}/(2\pi i)\mathbf{Z}$.

To describe this model as the algebraic integrable system we introduce the Lax operator $\Phi(z)$ (the explicit form of this matrix is not important) and define the Hamiltonians $h_1, \dots, h_N$ as the coefficients of the characteristic polynomial:

$$\text{Det}(x - \Phi(z)) = -\Lambda^{2N} e^z - e^{-z} + x^N + h_1 x^{N-1} + h_2 x^{N-2} + \dots + h_N \quad (31)$$

The Hamiltonians h_1, h_2 are then given by

$$h_1 = -\sum_{i=1}^N p_i; \quad h_2 = -\sum_{i < j} p_i p_j + U(x_1, \dots, x_N) \quad (32)$$

which differs from H_2 defined in (30) by the term $\frac{1}{2}h_1^2$.

Define the *spectral curve* $\mathcal{C}_h \subset \mathbf{C} \times \mathbf{C}^\times$, $x \in \mathbf{C}, z \in \mathbf{C}/2\pi i$, as the zero locus of the characteristic polynomial (31). For each value $h = (H_1, H_2, \dots, H_N)$ this is a curve which is a genus $N - 1$ hyperelliptic curve with two points where $x = \infty$ deleted. The fiber $\mathbf{H}^{-1}(h)$ is given by the product $\mathbf{C} \times \mathcal{J}_h$. The $\mathbf{C}$ -factor corresponds to the center-of-mass mode $\sum_i x_i$, the compact factor $\mathcal{J}_h = \text{Jac}(\overline{\mathcal{C}_h})$ is the Jacobian of the compactified curve $\mathcal{C}_h$. The complex action variables $a^i, a_{D,i}$ can be computed as the periods of the differential $\lambda = \frac{1}{2\pi} x dz$.

5.1.2. The gauge theory

The gauge theory significance of the periodic Toda chain is that the potential $\mathcal{F}(a)$ in the Eq. (26) defined using the family of spectral curves (31) and the above differential coincides with the prepotential of the low energy effective Lagrangian of the pure $\mathcal{N} = 2$ gauge theory with the gauge group $U(N)$.

5.1.3. The quantum system

The quantization of the periodic Toda chain is achieved by promoting h_k 's defined by (31) to the differential operators acting on functions of $(x_1, \dots, x_N)$, $p_i = \varepsilon \partial_{x_i}$. It is possible to show that the potential normal ordering ambiguities do not arise in this case. One is looking for the eigenfunctions of the form $(\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i)$:

$$\Psi(x_1, \dots, x_N) = e^{Nk\bar{x}} \psi_k(x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_N - \bar{x}) \quad (33)$$

$\psi_k \in L^2(\mathbf{R}^{N-1})$. When $\varepsilon = -i\hbar$ and $\Lambda, \hbar \in \mathbf{R}$ this problem has a discrete real spectrum (for fixed k). We are after the effective characterization of this spectrum. It turns out that this model allows a complex analytic continuation in the parameters

ε , Λ , so that the spectrum remains discrete, yet in general complex. We shall call this spectral problem the type A quantum periodic Toda.

The quantum periodic Toda admits another, somewhat unconventional (for $N > 2$) formulation, which also leads to the discrete yet complex spectrum. In this formulation we make the differential operators $\hat{H}_k$ act on functions $\Psi(x_1, \dots, x_N)$ which are $2\pi i$ -periodic, and non-singular for some fixed value of $\operatorname{Re} x_1, \dots, \operatorname{Re} x_N$. We shall call this spectral problem the type B quantum periodic Toda.

Note that for $N = 2$ case the type B periodic Toda is equivalent to finding the (quasi)-periodic solutions of the canonical Mathieu's differential equation, while the type A model corresponds to the L^2 solutions of Mathieu's modified differential equation.

Remark. The quasi-periodic solutions of the differential equations are parametrized by the so-called Bloch-Floquet multipliers, which in our correspondence come from the two-dimensional theta angle ϑ , which might come from the B -field in four dimensions, or from the peculiar deformation of the four dimensional tree level prepotential $\propto \varepsilon \vartheta \operatorname{tr} \Phi$, in the Ω -background $\mathbf{R}_\varepsilon^2$. Of course we can set $\vartheta = 0$ and discuss the periodic wavefunctions. This remark applies to all the many-body systems.

5.2. Elliptic Calogero-Moser system

5.2.1. The classical system

The elliptic Calogero-Moser system (eCM) is the system of N particles $x_1, x_2, \dots, x_N$ on the circle of circumference β , i.e. $x_i \sim x_i + \beta$, which interact with the pair-wise potential

$$U(x_1, x_2, \dots, x_N) = m^2 \sum_{i < j} u(x_i - x_j), \quad (34)$$

$$u(x) = C(\beta) + \sum_{k \in \mathbf{Z}} \frac{1}{\sinh^2(x + k\beta)} = -\partial_x^2 \log \Theta(x)$$

where Θ is the odd theta function on the elliptic curve E_τ with the modular parameter $\tau = \frac{i\beta}{\pi}$,

$$\Theta(x) = - \sum_{k \in \mathbf{Z} + \frac{1}{2}} (-1)^k q^{\frac{k^2}{2}} e^{2kx}, \quad q = \exp 2\pi i \tau$$

and $C(\beta)$ is some constant which depends on β .

Again, there exists a Lax representation [35] of the eCM system:

$$\Phi_{ij}(z) = p_i \delta_{ij} + m \frac{\Theta(z + x_i - x_j) \Theta'(0)}{\Theta(x_i - x_j) \Theta(z)} (1 - \delta_{ij}) \quad (35)$$

In the limit $\beta \rightarrow \infty$, $m \rightarrow \infty$, such that $\Lambda^{2N} = m^{2N} q$ is kept finite, the eCM system becomes the periodic Toda chain [36], where

$$x_i^{\text{eCM}} = \frac{i}{N} \beta + x_i^{\text{pToda}}$$

5.2.2. The gauge theory

The spectral curve $\text{Det}(\Phi(z) - x) = 0$, which is an N -sheeted ramified cover of the elliptic curve E_τ where z lives, is the Seiberg-Witten curve of a remarkable four dimensional $\mathcal{N} = 2$ theory. This is an $SU(N)$ gauge theory with a massive adjoint hypermultiplet. The parameter m in (35) is the mass of the adjoint hypermultiplet, the complex structure of the curve E_τ is determined by the complexified bare gauge coupling of the ultraviolet theory (which is in fact the superconformal $\mathcal{N} = 4$ super-Yang-Miills). We shall call this theory the $\mathcal{N} = 2^*$ theory.

5.2.3. The quantum system

$$\hat{H}_2 = \frac{\varepsilon^2}{2} \sum_{i=1}^N \frac{\partial^2}{\partial x_i^2} - m(m + \varepsilon) \sum_{i < j} \wp(x_i - x_j) \quad (36)$$

When $\varepsilon = -i\hbar$, $m = -\nu\varepsilon$, and $\hbar, \nu, \beta \in \mathbf{R}_+$ the spectral problem of the operator (36) is well-known: one is looking for the (quasi-)periodic (in β) symmetric functions $\Psi(x_1, \dots, x_N)$, which are non-singular in the fundamental domain, and vanish at:

$$\Psi(x_1, \dots, x_N) \sim (x_i - x_j)^\nu \quad (37)$$

It follows from our results that this problem has an analytic continuation where ε, m, β all become complex. The monodromy of this continuation is an extremely interesting problem which we shall not be able to cover in this exposition.

Just like in the case of the periodic Toda chain one may consider various spectral problems for the Hamiltonian $\hat{H}_2$ and its higher order counterparts. The pleasant bonus of having an elliptic potential is the similarity of these problems. The type A quantum elliptic Calogero-Moser is, therefore, a problem which we just described: finding the β -periodic L^2 (on the subspace with fixed $\text{Im}(x_i/\beta)$) wavefunctions, with the (37) behavior. The type B problem is a problem of finding the πi - periodic L^2 wavefunctions (on the subspace with fixed $\text{Re}(x_i)$). The $SL_2(\mathbf{Z})$ transformation $\tau \rightarrow -\frac{1}{\tau}$ maps the type A problem to the type B problem and vice versa.

Superficially, however, one may worry that the type A and the type B problems are quite different in nature. In the limit $\beta \rightarrow \infty$ the type A problem becomes that of the hyperbolic Sutherland model, which has a continuous spectrum. In the same limit the type B problem becomes that of the trigonometric Sutherland, which is well-studied, has a discrete spectrum, and Jack polynomials (up to a ground state factor) as the eigenfunctions. Our approach connects all these models.

In the $N = 2$ case we are solving the celebrated Lamé's differential equation [37].

5.3. Hitchin system

The previous systems are the degenerate examples of the so-called Hitchin integrable system. Its phase space $\mathcal{P}$ is the partial resolution of the cotangent bundle to the

moduli space $\mathcal{M}$ of semistable holomorphic bundles on a complex curve C . More precisely, the phase space is the space of pairs $(\nabla_{\bar{\partial}}, \Phi = \Phi_z dz)$, where $\nabla_{\bar{\partial}} = \bar{\partial} + \bar{A}$, $\bar{A} = A_{\bar{z}} d\bar{z}$, is the $(0, 1)$ -part of a connection on a vector bundle E over C , and Φ_z is the Higgs field, a $(1, 0)$ -form valued in the endomorphisms of the E . The vector bundle E gets a holomorphic bundle structure by declaring the solutions to $\nabla_{\bar{z}} \chi = 0$ to be the holomorphic sections. The Higgs field Φ_z must obey:

$$\nabla_{\bar{\partial}} \Phi = 0, \text{ i.e. } \bar{\partial}_{\bar{z}} \Phi_z + [A_{\bar{z}}, \Phi_z] = 0 \quad (38)$$

Divide by the complex gauge transformations, or make a symplectic quotient with respect to the compact gauge transformations. In the latter case one imposes the real moment map condition:

$$F_A + [\Phi, \bar{\Phi}] = 0 \quad (39)$$

The holomorphic symplectic structure is induced from

$$\omega = \int_C \text{tr } \delta \Phi \wedge \delta \bar{A} \quad (40)$$

The Hamiltonians are obtained as follows: consider the $(j, 0)$ -differentials on C , given by $\text{tr} \Phi^j$. Due to (38) these are holomorphic j -differentials, and (for $j > 1$) there are $(2j - 1)(g - 1)$ linearly independent (over $\mathbf{C}$) such differentials. The base $\mathcal{B}$ of the corresponding integrable system is the vector space:

$$\mathcal{B} = \oplus_{j=1}^N H^0(C, K_C^{\otimes j}) \approx \mathbf{C}^{1+N^2(g-1)} \quad (41)$$

while the fiber over a point $h \in \mathcal{B}$ is the Jacobian of the spectral curve:

$$\mathcal{C} \subset T^*C : \quad 0 = \text{Det}(\Phi - x) \quad (42)$$

(where we view x as the canonical Liouville one-form on T^*C). The homology class $[\mathcal{C}]$ spectral curve is equal to $N[C]$. It follows that the self-intersection number $2g(\mathcal{C}) - 2 = \mathcal{C} \cdot \mathcal{C} = N^2 C \cdot C = N^2(2g - 2)$, therefore the genus

$$g(\mathcal{C}) = 1 + N^2(g - 1) \quad (43)$$

is equal to the dimension of $\mathcal{B}$. The polarization comes from the Kahler form:

$$\varpi = \int_C \text{tr} (\delta A \wedge \delta \bar{A} + \delta \Phi \wedge \delta \bar{\Phi}) \quad (44)$$

Let us now explain that the systems we considered so far are the degenerate cases of Hitchin system. Imagine we study a Hitchin system for the group $U(N)$ on a curve of genus two. Now let us degenerate the curve, in such a way that it becomes a union of two elliptic curves, connected by a long neck. Then the equations (38) can be solved on both elliptic curves independently, the only memory of the original curve being a boundary condition at the puncture representing the neck. The limiting Hitchin system would actually split as a union of invariant submanifolds, labelled by the coadjoint orbits $\mathcal{O}$ of the complexified gauge group, attached to the puncture

(as a hyperkähler manifold it fibers over $(\mathfrak{t} \otimes \mathbf{R}^3)/W$) [38]. The limiting equations (38) look like:

$$\bar{\partial}_{\bar{z}} \Phi_z + [A_{\bar{z}}, \Phi_z] = J \delta^{(2)}(z, \bar{z}) \quad (45)$$

J represents a (complex) moment map of the group action on a coadjoint orbit $\mathcal{O}$. Consider a very special vase, where the orbit $\mathcal{O}$ is diffeomorphic to $T^*\mathbf{CP}^{N-1}$. The corresponding moment map J is the $N \times N$ complex traceless matrix of which $N-1$ eigenvalues coincide: $J = m \operatorname{diag}(N-1, -1, \dots, -1)$. By solving the equation (45) one gets (up to an irrelevant gauge transformation) the Lax operator (35) in the gauge where $A_{\bar{z}} = \operatorname{diag}(x_1, \dots, x_N)$ is a constant diagonal matrix [39].

5.3.1. The quantum system and the gauge theory

We do not have much to say about these two topics in the case of Hitchin system for compact Riemann surface. The supersymmetric field theory realization of this system involves the six dimensional $(2, 0)$ theory compactified on the Riemann surface C . Using the results of [40] one may hope to formulate the result as the four dimensional $\mathcal{N} = 2$ supersymmetric gauge theory, at least for the complex curves C which are close to the degeneration locus. This gauge theory can be further subject to the Ω -background. As a result one would get a quantization of Hitchin system, together with an expression for its YY function. It would be interesting to understand the relation of our approach to the constructions of [41].

5.4. Relativistic systems

So far we considered the systems which have a Galilean symmetry group, the space translations are generated by $\hat{H}_1$ while the time translations are generated by $\hat{H}_2$, and there is a generator of the the Galilean boost $\hat{S}$, such that $[\hat{S}, \hat{H}_2] = \hat{H}_1$. The point is that the Hamiltonians $\hat{H}_1, \hat{H}_2, \dots, \hat{H}_N$ are polynomials in the momenta $\hat{p}_i$.

There exists [42, 43] a one-parametric deformation of the eCM integrable system, whose Hamiltonians are trigonometric (hyperbolic) functions of the momenta. In particular, the Hamiltonians $\hat{P} = \hat{H}_1, \hat{E} = \hat{H}_2$ become

$$\hat{P} = \sum_{i=1}^N \sinh(\beta p_i) f_i(x); \quad \hat{E} = \sum_{i=1}^N \cosh(\beta p_i) f_i(x); \quad f_i(x) = \prod_{j \neq i} \sqrt{1 - \frac{\wp(x_i - x_j)}{\wp(\beta m)}} \quad (46)$$

The quantization of the system (46) produces a family of commuting difference operators. These operators, in the trigonometric limit, reduce (after a similarity transformation), to the famous Macdonald difference operators [44], whose eigenfunctions are the so-called Macdonald polynomials. In the context of a physical problem, the polynomial eigenfunctions correspond to the compact model, where the particles $x_1, \dots, x_N$ live on a circle. The problem of interest for us reduces, in the trigonometric limit, to the hyperbolic system where the particles live on a real line, and the spectrum, as in the non-relativistic case, becomes continuous. As

far as we know, there is no satisfactory treatment of this model in the literature. Therefore our results, which we present below, might be of additional interest for the community working in the domain of harmonic analysis.

In our analysis, the relativistic systems (46) correspond to the five dimensional gauge theory compactified on a circle, subject further to the Ω -background in $\mathbf{R}_\varepsilon^2$.

6. Superpotential/Yang-Yang function $\mathcal{W}(a, \varepsilon; \mathbf{q})$

6.1. Thermodynamic Bethe Ansatz

The perturbative part of $\mathcal{W}$ is written using the special function $\varpi_\varepsilon(x)$. It obeys

$$\frac{d}{dx}\varpi_\varepsilon(x) = \log \Gamma\left(1 + \frac{x}{\varepsilon}\right).$$

The instanton part of $\mathcal{W}$ is given by the critical value of the following functional:

$$\mathcal{W}^{\text{inst}}(a; \mathbf{q}) = \text{Crit}_{\rho, \varphi} \frac{1}{2} \int_{\mathcal{C} \times \mathcal{C}} \rho(x) \rho(y) G(x-y) + \int_{\mathcal{C}} \left[\rho(x) \varphi(x) + \text{Li}_2 \left(\mathbf{q} Q(x) e^{-\varphi(x)} \right) \right] \quad (47)$$

Equivalently:

$$\mathcal{W}^{\text{inst}}(a; \mathbf{q}) = \int_{\mathcal{C}} \left[-\frac{1}{2} \varphi(x) \log \left(1 - \mathbf{q} Q(x) e^{-\varphi(x)} \right) + \text{Li}_2 \left(\mathbf{q} Q(x) e^{-\varphi(x)} \right) \right] \quad (48)$$

where $\varphi(x)$ solves a TBA-like equation:

$$\varphi(x) = \int_{\mathcal{C}} G(x-y) \log \left(1 - \mathbf{q} Q(y) e^{-\varphi(y)} \right) \quad (49)$$

The solution is determined by the choice of the contour $\mathcal{C}$, and the functions $G(x)$, $Q(x)$. Given this data, it is straightforward to solve (49) recursively:

$$\varphi(x) = \sum_{k=1}^{\infty} \mathbf{q}^k \varphi_k(x), \quad \rho(x) = \sum_{k=1}^{\infty} \mathbf{q}^k \rho_k(x) \quad (50)$$

6.2. The examples

Here we give the functions $Q(x)$, $G(x)$ and the contour $\mathcal{C}$, for our examples.

6.2.1. The elliptic Calogero-Moser system

Let $P(x) = \prod_{l=1}^N (x - a_l)$, and

$$Q(x) = \frac{P(x-m)P(x+m+\varepsilon)}{P(x)P(x+\varepsilon)}; \quad G(x) = \frac{d}{dx} \log \frac{(x+m+\varepsilon)(x-m)(x-\varepsilon)}{(x-m-\varepsilon)(x+m)(x+\varepsilon)} \quad (51)$$

Then: $\mathcal{W}(a, \varepsilon; \mathbf{q}) = \mathcal{W}^{\text{pert}} + \mathcal{W}^{\text{inst}}$,

$$\mathcal{W}^{\text{pert}} = \frac{\pi i \tau}{2\varepsilon} \sum_{n=1}^N a_n^2 + \sum_{l,n=1}^N (\varpi_\varepsilon(a_l - a_n) - \varpi_\varepsilon(a_l - a_n - m - \varepsilon)) \quad (52)$$

and $\mathcal{W}^{\text{inst}}$ is given by (48) with the contour $\mathcal{C}$ which comes from infinity, surrounds in the complex plane the points $a_l + k\varepsilon$, $l = 1, \dots, N$, $k = 0, 1, 2, \dots$, and goes back to infinity. It separates these points and the points $a_l + lm + k\varepsilon$, $l \in \mathbf{Z}$, $k = -1, -2, \dots$

6.2.2. Periodic Toda

For the periodic Toda chain one has a simpler formalism with

$$Q(x) = \frac{1}{P(x)P(x+\varepsilon)}, \quad G(x) = \frac{d}{dx} \log \frac{(x-\varepsilon)}{(x+\varepsilon)}$$

$$\mathcal{W}^{\text{pert}} = \frac{1}{2\varepsilon} \log \left(\frac{\Lambda}{\varepsilon} \right) \sum_{n=1}^N a_n^2 + \sum_{l,n=1}^N \varpi_\varepsilon(a_l - a_n) \quad (53)$$

and the contour $\mathcal{C}$ that goes around the points $a_l + k\varepsilon$, $k \geq 0$. It can be deformed to go along the real line, when $a_l \in \mathbf{R} + \frac{\varepsilon}{2}$, and $\text{Im}\varepsilon > 0$.

6.2.3. Ruijsenaars-Schneider model

Let us denote: $q = e^{-\beta\varepsilon}$, $t = e^{-\beta m}$, $z = e^{-\beta x}$, $w_l = e^{\beta a_l}$. The perturbative contribution:

$$\mathcal{W}^{\text{pert}} = \frac{\pi i \tau}{2\varepsilon} \sum_{n=1}^N a_n^2 + \sum_{l,n=1}^N (\Pi_q(w_l/w_n) - \Pi_q(tw_l/w_n))$$

where

$$\Pi_q(w) = \frac{1}{\beta} \sum_{n=1}^{\infty} \frac{w^n}{n^2(1-q^n)} \sim \sum_{n \in \mathbf{Z}} \varpi_\varepsilon \left(\frac{\log(w) + 2\pi i n}{\beta} \right)$$

The P, Q, G functions are given by:

$$P(x) = \prod_{l=1}^N (1 - w_l z), \quad Q(x) = \frac{P(zt^{-1})P(qtz)}{P(z)P(qz)} \quad (54)$$

$$G(x) = z \frac{d}{dz} \log \left(\frac{z^2(z-tq)(1-tz)(1-qz)}{(z-t)(z-q)(1-qtz)} \right)$$

We assume $|q|, |t|, |w_l| < 1$, and the contour $\mathcal{C}$ is the unit circle $|z| = 1$.

6.2.4. The spectrum of observables

The spectrum of the type A quantum system is given by the solutions of the equations (13) with $\mathcal{W}(a, \varepsilon)$ given by (47). Let us denote the extremum of the functional (47) by $\rho_A(x), \varphi_A(x)$. The spectrum of the type B quantum system is given by the solutions of the “dual” equations

$$a_i = \varepsilon n_i, \quad n_i \in \mathbf{Z} \quad (55)$$

With these a_i 's one can again define the functional (47) and study its extremum $\rho_B(x), \varphi_B(x)$.

The eigenvalues of the quantized Hamiltonians

$$H_k = \int dz d\bar{z} \operatorname{tr} \Phi(z)^k = \sum_{l=1}^N p_l^k + \dots :$$

are given respectively, for the type A and type B models:

$$\mathcal{E}_k = \sum_{l=1}^N a_l^k + k \int_{\mathcal{C}} dx ((x + \varepsilon)^{k-1} - x^{k-1}) \rho_{A,B}(x), \quad (56)$$

In particular $\mathcal{E}_2 = \varepsilon q \frac{d}{dq} \mathcal{W}(a; q)$.

6.2.5. On the relation between the type A and the type B models

The $\mathcal{N} = 2^*$ gauge theory becomes the $\mathcal{N} = 4$ super-Yang-Mills theory in the ultraviolet. The latter has the celebrated Montonen-Olive $SL_2(\mathbf{Z})$ symmetry. In particular, the S -duality transformation maps the gauge theory with the gauge group G and the coupling τ to the gauge theory with the gauge group ${}^L G$ and the coupling $-\frac{1}{h\tau}$, for some integer $h = 1, 2, 3$. When the theory is perturbed by the mass term, the S -duality symmetry still acts. We claim it maps the type A model of the elliptic Calogero-Moser system with the modular parameter τ to the type B model with the modular parameter $-\frac{1}{\tau}$. The special coordinates a_i map to τa_i . The modularity of the superpotential is clearly supported by the expansion (59).

6.2.6. More details and the origin of integral equation

Here we give more details in regard to the origin of claims from previous subsections. We do it for the case of the $\mathcal{N} = 2^*$ theory (the relativistic case is studied analogously, along the lines of [45, 46]). One starts with the contour integral representation [1, 6–8] for the instanton partition function in the general Ω -background:

$$\mathcal{Z}^{\text{inst}}(\mathbf{a}; q, m, \varepsilon_1, \varepsilon_2) = \quad (57)$$

$$\sum_{k=0}^{\infty} \frac{q^k}{k!} \int_{\mathbf{R}^k} \prod_{1 \leq I < J \leq k} \mathcal{D}(\phi_{IJ}) \prod_{I=1}^k Q(\phi_I) \frac{\varepsilon(m + \varepsilon_1)(m + \varepsilon_2)}{\varepsilon_1 \varepsilon_2 m(m + \varepsilon)} \frac{d\phi_I}{2\pi i},$$

where $\phi_{IJ} = \phi_I - \phi_J$,

$$\mathcal{D}(x) = \frac{x^2(x^2 - \varepsilon^2)(x^2 - (m + \varepsilon_1)^2)(x^2 - (m + \varepsilon_2)^2)}{(x^2 - \varepsilon_1^2)(x^2 - \varepsilon_2^2)(x^2 - m^2)(x^2 - (m + \varepsilon)^2)},$$

$Q(x)$ was introduced in (51). Next use the observation, reported earlier in [18, 47], that (57) is a partition function of a one-dimensional non-ideal gas of particles $\phi_1, \dots, \phi_k$ subject to the external potential $U_{\text{ext}}(x)$ and pair-wise interaction $V_{\text{int}}(x)$:

$$U_{\text{ext}}(x) = -\log \left(Q(x) \frac{(m + \varepsilon_1)(m + \varepsilon_2)\varepsilon}{\varepsilon_1 \varepsilon_2 m(m + \varepsilon)} \right); \quad V_{\text{int}}(x) = -\log(\mathcal{D}(x)) \quad (58)$$

The free energy of this gaz can be analyzed using Mayer expansion [48–50]. The subtlety with the $\varepsilon_2 \rightarrow 0$ limit is the clustering of the instanton particles, which leads to the multiple vertices of a given valency, labelled by an arbitrary positive integer k (the number of instantons in a given cluster) weighted by the partition function of a simpler one-dimensional gas

$$\frac{1}{k! \varepsilon_2^k} \int_{\mathbf{R}^k} \delta \left(\sum_I \phi_I \right) d^k \phi \prod_{1 \leq I < J \leq k} \frac{\phi_{IJ}}{\phi_{IJ} + \varepsilon_2}$$

which is equal to $\frac{1}{k^2 \varepsilon_2}$, cf. [8]. By summing over k 's one arrives at the dilogarithm function in (48). Finally, the limit in (20) singles out the tree diagrams, which lead to (48), (49) with the propagator function $G(x)$ given by: $G(x) = \text{Limit}_{\varepsilon_2 \rightarrow 0} \frac{\mathcal{D}(x)-1}{\varepsilon_2}$.

6.3. The sum over partitions

The $U(N)$ instanton partition function can also be written as a sum over N -tuples of partitions $\lambda^{(1)}, \dots, \lambda^{(N)}$, which represent various configurations of instantons sitting on top of each other in $\mathbf{R}^4$. This representation can be effectively used to compute the first few terms in the q -expansion of $\mathcal{W}^{\text{inst}}$, or the first terms in the $\frac{1}{a^2}$ expansion, say, in the region $|a_i - a_j| \gg |\varepsilon|$ (here $\varphi(q) = \log \prod_{n=1}^{\infty} \frac{1}{1-q^n}$):

$$\mathcal{W}^{\text{inst}}(a, \varepsilon; m, q) = \frac{Nm(m+\varepsilon)}{\varepsilon} \varphi(q) + \frac{1}{\varepsilon} \sum_{i \neq j} \frac{m^2(m+\varepsilon)^2}{(a_i - a_j)^2 - \varepsilon^2} q \frac{d\varphi(q)}{dq} + \dots \quad (59)$$

7. Discussion

To conclude we briefly review the relation of our results to some recent work.

In [19] the wavefunctions of the periodic Toda chain are constructed in the form of integrals involving the solution $Q(x)$ of the Baxter equation, cf. [32]. In order for the wavefunction to belong to the L^2 Hilbert space the zeroes $\delta_1, \dots, \delta_N$ of Q must obey the *quantization condition*. We claim the *quantization conditions* of [19] are equivalent to our Bethe equations for the pure $\mathcal{N} = 2$ $U(N)$ gauge theory, with the identification $\delta_i = a_i$. We have therefore provided the YY function for the *quantization conditons in the periodic Toda chain*.

In [20] the quantum N -particle elliptic Calogero-Moser system for the A_{N-1} system, for integral ν was studied using the critical level limit of the free field representation of the conformal blocks of the WZW conformal field theory on a torus. The wavefunctions in [20] are written in terms of $m = \frac{N(N-1)}{2}(\nu-1)$ parameters $t_{\alpha_i}^{(i)}$, $\alpha_i = 1, \dots, i(\nu-1)$, $i = 1, \dots, N-1$, obeying the *elliptic Bethe-like* equations, which are derived from the YY function

$$Y(t_{\alpha}^{(i)}; \xi_1, \dots, \xi_N) = 2\pi i \sum_{i=1}^N \left(\frac{1}{2} \tau \xi_i^2 + \xi_i \sum_{\alpha=1}^{i(\nu-1)} t_{\alpha}^{(i)} \right) - S(t_{\alpha}^{(i)}; \tau)$$

with the quasimomenta $\xi_1, \dots, \xi_N$ being the fixed parameters. One can supplement the Bethe equations of [20] by the *quantization conditions* $\xi_1 = \vartheta + n_1, \xi_2 = \vartheta + n_2, \dots, \xi_N = \vartheta + n_N$, with $n_i \in \mathbf{Z}$. We conjecture

$$\text{Crit}_{t_\alpha^{(i)}, \sum_{\alpha=1}^{i(\nu-1)} t_\alpha^{(i)} = a_i} S(t_\alpha^{(i)}; \tau) = \tilde{W}^{\text{eff}}(a_1, \dots, a_N; -\nu\varepsilon, \varepsilon, \tau) \quad (60)$$

Analogously, the critical value at fixed ξ gives, conjecturally, the YY function for the type B model. Note that the potential $S(t; \xi; \tau)$ of [20] corresponds, in the table of dualities sketched in [4, 5] to (a limit of) the four dimensional quiver gauge theory with the gauge group $U(\nu-1) \times U(2(\nu-1)) \times \dots \times U((N-1)(\nu-1))$, compactified on the elliptic curve E_τ . The equivalence (60) suggests an interesting duality between the $\mathcal{N} = 2^*$ theory in the special Ω background with the coupling τ and the quiver gauge theory.

In [51] a partial resummation of the quantum mechanical perturbation theory around the trigonometric Sutherland model is proposed, leading to some equations on the eigenvalue of the $\hat{H}_2$ operator of the elliptic Calogero-Moser system, for arbitrary value of the coupling ν . The relation of this approach to the YY function formalism is not clear to us at the moment.

The last but not least are the works [41], [52] which we feel are relevant to our discussion.

The lack of space and time did not allow us to discuss the relation to the AdS/CFT and the conjectural integrability of the planar $\mathcal{N} = 4$ super-Yang-Mills theory as well as the geometry of vacua of the three dimensional gauge theory.

We would like to return to all these questions in the future.

8. Acknowledgments

The research of NN was supported by the grants ANR-06-BLAN-3-137168 and ANR-05-BLAN-0029-01, RFFI 06-02-17382 and NSH-8065.2006.2, that of SSh was supported by SFI grants 05/RFP/MAT0036 and 08/RFP/MTH1546, by European Science Foundation grant ITGP and the Hamilton Mathematics Institute TCD.

We thank A. Gerasimov, A. Gorsky, A. Okounkov, E. Sklyanin, F. Smirnov, J. Stalker, L. Takhtajan, P. Wiegmann and E. Witten for the discussions.

References

- [1] G. Moore, N. Nekrasov, S. Shatashvili, *Integration over the Higgs branches*, Comm. Math. Phys. 209 (2000) 97-121, arXiv:hep-th/9712241.
- [2] A. Gerasimov, S. Shatashvili, *Higgs Bundles, Gauge Theories and Quantum Groups*, Comm. Math. Phys. 277 (2008) 323-367, arXiv:hep-th/0609024.
- [3] A. Gerasimov, S. Shatashvili, *Two-dimensional Gauge Theories and Quantum Integrable Systems*, In "From Hodge Theory to Integrability and TQFT: tt*-geometry", pp. 239-262, R. Donagi and K. Wendland, Eds., Proc. of Symposia in Pure Mathematics Vol. 78, American Mathematical Society, Providence, Rhode Island, 2008; arXiv:0711.1472.

- [4] N. Nekrasov, S. Shatashvili, *Supersymmetric vacua and Bethe Ansatz*, In “Cargese 2008, Theory and Particle Physics: the LHC perspective and beyond”, arXiv:0901.4744.
- [5] N. Nekrasov, S. Shatashvili, *Quantum integrability and supersymmetric vacua*, Prog. Theor. Phys. Suppl. 177:105-119, 2009, arXiv:0901.4748.
- [6] N. Nekrasov, *Seiberg-Witten prepotential from instanton counting*, Adv. Theor. Math. Phys. 7:831-864, 2004, hep-th/0206161.
- [7] A. Losev, N. Nekrasov, S. Shatashvili, *Testing Seiberg-Witten solution*, In “Cargese 1997, Strings, branes and dualities” 359-372, hep-th/9801061.
- [8] G. Moore, N. Nekrasov, S. Shatashvili, *D particle bound states and generalized instantons*, Commun. Math. Phys. 209 (2000) 77-95, hep-th/9803265.
- [9] N. Seiberg, E. Witten, *Monopole Condensation, And Confinement In $\mathcal{N} = 2$ Supersymmetric Yang-Mills Theory*, arXiv:hep-th/9407087, Nucl. Phys. B426 (1994) 19-52.
- [10] A. Gorsky, I. Krichever, A. Marshakov, A. Mironov, A. Morozov, *Integrability and Seiberg-Witten exact solution*, Phys. Lett. B355:466-474, 1995, hep-th/9505035.
- [11] R. Donagi, E. Witten, *Supersymmetric Yang-Mills theory and integrable systems*, Nucl. Phys. B460:299-334, 1996, hep-th/9510101.
- [12] C. N. Yang, C. P. Yang, *Thermodynamics of a one-dimensional system of bosons with repulsive delta-function interaction*, J. Math. Phys. 10 (1969), 1115.
- [13] N. Nekrasov, S. Shatashvili, *SUSY vacua and quantum integrability*, to appear.
- [14] K. Hori, S. Katz, A. Klemm, R. Pandharipande, R. Thomas, C. Vafa, R. Vakil, E. Zaslow (Eds.) *Mirror symmetry*, AMS, Providence, 2003, 929p.
- [15] A. Losev, N. Nekrasov, S. Shatashvili, *Issues in topological gauge theory*, Commun. Math. Phys. 209:97-121, 2000, hep-th/9712241.
- [16] H. Bethe, *On the theory of metals. 1. Eigenvalues and eigenfunctions for the linear atomic chain*, Z. Phys. 71: 205-226, 1931.
- [17] A. Losev, A. Marshakov, N. Nekrasov, *Small instantons, little strings and free fermions*, In “Shifman, M. (ed.): From fields to strings, vol. 1”, 581, hep-th/0302191.
- [18] A. Marshakov, N. Nekrasov, *Extended Seiberg-Witten Theory and Integrable Hierarchy*, JHEP 0701:104, 2007, hep-th/0612019.
- [19] S. Kharchev, D. Lebedev, *Integral representations for the eigenfunctions of quantum open and periodic Toda chains from QISM formalism*, J.Phys.A34:2247-2258, 2001, hep-th/0007040.
- [20] G. Felder, A. Varchenko, *Integral representation of solutions of the elliptic Knizhnik-Zamolodchikov-Bernard equations*, Int.Math.Res.Not.(1995) 221-233, hep-th/9502165.
- [21] H. Braden, I. Krichever, eds. *The Seiberg-Witten and Whitham Equations*, Gordon and Breach Science Publishers, 2000.
- [22] N. Nekrasov, A. Okounkov, *Seiberg-Witten theory and random partitions*, hep-th/0306238.
- [23] A.I. Zamolodchikov, *Thermodynamic Bethe Ansatz in Relativistic Models. Scaling 3-state Potts and Lee-Yang Models*, Nucl. Phys. B342 (1990) 695.
- [24] C. Destri, H. J. de Vega, *New approach to thermal Bethe Ansatz*, Phys. Rev. Lett. 69 (1992), 2313.
- [25] V.V.Bazhanov, S.L.Lukyanov and A.B.Zamolodchikov, *Integrable quantum field theories in finite volume: excited state energies*, hep-th/9607099.
- [26] P. Dorey and R. Tateo, *Excited states by analytic continuation of TBA equations*, Nucl. Phys. B482 (1996) 639-659, hep-th/9607167.
- [27] J. Teschner, *On the spectrum of the sinhGordon model in finite volume*, hep-th/0702214.

- [28] A. Givental, *Equivariant Gromov-Witten Invariants*, arXiv:alg-geom/9603021
- [29] L.D. Faddeev, E. Sklyanin, L. Takhtajan, Quantum inverse problem method, Theor. Math. Phys. 40:2 (1980) 688-706, Teor. Mat. Fiz. 40:194-220,1979 (in Russian)
- [30] L. A. Takhtajan, L. D. Faddeev, *The Quantum method of the inverse problem and the Heisenberg XYZ model*, Russ. Math. Surveys 34:11-68,1979.
- [31] L. D. Faddeev, *How algebraic Bethe Ansatz works for integrable model*, hep-th/9605187.
- [32] E. K. Sklyanin, *Separation of variables - new trends*. Prog. Theor. Phys. Suppl. 118:35-60, 1995. e-Print: solv-int/9504001
- [33] R. Baxter, *Exactly solved models in statistical mechanics*, AP, London 1982.
- [34] A. Beilinson, V. Drinfeld, *Quantization of Hitchin's integrable system and Hecke eigen-sheaves*, preprint (ca. 1995), <http://www.math.uchicago.edu/~arinkin/langlands/>.
- [35] I. Krichever, *Elliptic solutions of the Kadomtsev-Petviashvili equation and integrable systems of particles*, Funkts. Anal. Prilozh., 14:4 (1980), 45-54
- [36] V. Inozemtsev, *The Finite Toda Lattices*, Comm. Math. Phys.121 (1989) 629-638
- [37] G. Lamé, *Sur les surfaces isothermes dans les corps homogènes en équilibre de température*, J. Math. Pures Appl. 2 (1837), 147-188.
- [38] N. Nekrasov, *Holomorphic bundles and many-body systems*, Commun. Math. Phys. 180 (1996) 587-604, hep-th/9503157.
- [39] A. Gorsky, N. Nekrasov, *Elliptic Calogero-Moser System from Two Dimensional Current Algebra*, hep-th/9401021.
- [40] D. Gaiotto, $\mathcal{N} = 2$ dualities, arXiv:0904.2715v1
- [41] A. Kapustin, E. Witten, *Electric-Magnetic Duality And The Geometric Langlands Program*, arXiv:hep-th/0604151
- [42] S. Ruijsenaars, H. Schneider, *A new class of integrable systems and its relation to solitons*, Ann. Phys. (NY) 170 (1986) 370-405.
- [43] S. Ruijsenaars, *Complete integrability of relativistic Calogero-Moser systems and elliptic function identities*, Comm. Math. Phys. 110 (1987) 191-213.
- [44] I. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford, 1998
- [45] N. Nekrasov, *Five dimensional gauge theories and the relativistic gauge theories*, Nucl. Phys. B531 (1998) 323-344, hep-th/9609219.
- [46] N. Nekrasov, S. Shadchin, *ABCD of instantons*, Comm. Math. Phys. 252 (2004) 359-391, arXiv:hep-th/0404225
- [47] N. Nekrasov, *Random partitions, topological strings and Mayer expansion*, lectures at the University of Amsterdam conference "On random Partitions and Topological Strings", June 2005.
- [48] J. Mayer, M. G. Mayer, *Statistical Mechanics*, New York 1940
- [49] J. Mayer, E. Montroll, *Molecular distributions*, J. Chem. Phys. **9** (1941) 2-16
- [50] A. Polyakov, *Microscopic description of critical phenomena* JETP 55, 1968, 1026.
- [51] E. Langmann, *An explicit solution of the (quantum) elliptic Calogero-Sutherland model*, arXiv:math-ph/0407050
- [52] S. Gukov, E. Witten, *Branes and Quantization*, arXiv:0809.0305

CHANGING VIEWS OF QUANTUM FIELD THEORY

STEVEN WEINBERG

Theory Group, Department of Physics, University of Texas, Austin, TX 78712
E-mail: weinberg@physics.utexas.edu

The first part of this talk reviews changes in our views regarding quantum field theory since its beginnings, leading eventually to the modern view that our most successful field theories may in fact be effective field theories, valid only as low energy approximations to an underlying theory that may not be a field theory at all. In the second part, I reminisce about the early development of effective field theories of the strong interactions, comment briefly on some other applications of effective field theories, then take up the idea that the Standard Model and General Relativity are the leading terms in an effective field theory, and finally cite recent calculations that suggest that the effective field theory of gravitation and matter is asymptotically safe. The second part is substantially the same as a talk given a month earlier at the 6th International Workshop on Chiral Dynamics, at the University of Bern, which is reproduced here.

1

The history of quantum field theory, like the history of much else, shows a continued oscillation between optimism and pessimism. Quantum field theory goes back to the origins of quantum mechanics itself, when Born, Heisenberg, and Jordan in 1926 applied the newly developed quantum mechanics to the electromagnetic field. Originally quantum field theory was just that, the quantum theory of the electromagnetic field. Some theorists such as Dirac held to the vision of a theory of relativistic wave equations for particles like the electron and a quantum field theory for the electromagnetic field – to my mind a rather unattractive dualistic approach to physics. Pauli, on the other hand, in papers with Heisenberg and Weisskopf, was instrumental in bringing into being a thoroughgoing quantum field theory of all matter. Everything, every particle is a quantum of some field. The photon is a quantum of the electromagnetic field, the electron is a quantum of the electron field. In this view, there really is not a fundamental distinction between particles like the electron and fields like the electromagnetic field.

Just as soon as quantum field theory in this sense was being born, around 1930, it ran into a problem. Oppenheimer and Waller discovered that when one tries to push the quantum field theory of electrons and photons beyond the lowest order of approximation in calculating the energy of an electron (either a bound electron in Oppenheimer's work or a free electron in Waller's) one encounters an integral over the momentum of the virtual photon that diverges at high momentum. This was

called an ultraviolet divergence, or more colorfully, an ultraviolet catastrophe. As a result the optimism of physicists regarding quantum field theory began to decline into pessimism. Oppenheimer in particular was influential in arguing that quantum field theories simply could not be trusted beyond the lowest approximation, or trusted at all at energies more than a few MeV.

Then in the late 1940s there was a new wave of optimism. The invention of a relativistic perturbation theory made it possible to see that all the infinities in quantum field theory could be absorbed into a redefinition of parameters like the mass of the electron and the charge of the electron. Some theorists thought that this renormalization theory was just a way of sweeping infinities under the rug, but most were more optimistic. I recall that when I was a graduate student in the late 1950s I thought that renormalization theory was really wonderful, because it wasn't just a way of eliminating the infinities in a theory, it was a way of selecting the right theory. As far as we knew at the time, this renormalization procedure, absorbing all ultraviolet divergences into a redefinition of the constants of the theory, would only work for a very limited class of theories, the so called renormalizable theories. These were theories in which all of the coefficients of terms in the Lagrangian, like the charge of the electron or the mass of the electron, either were dimensionless (in natural units, in which Planck's constant and the speed of light equal one, so that all dimensions are just powers of mass) like the charge e of the electron, for which $e^2/4\pi \simeq 1/137$, or have dimensions that are positive powers of mass, like the electron mass itself, but there are no parameters whose dimensions are negative powers of mass.

The requirement of renormalizability allowed only a very limited class of theories, which you might think would be a drawback, but to me it seemed just what was needed, because this requirement provided a way of understanding why quantum electrodynamics was the way it was. For example, there's nothing in the symmetries of quantum electrodynamics, Lorentz invariance and gauge invariance and so on, which rules out putting a term in the field equations or the Lagrangian that could make the magnetic moment of the electron anything you want. It was a great triumph of quantum field theory that Schwinger was able to calculate the $e^2/8\pi^2$ fractional correction to the magnetic moment of the electron, but in fact one could have gotten any answer by putting a term in the Lagrangian proportional to $F^{\mu\nu}\bar{\psi}\sigma_{\mu\nu}\psi$. However, the coefficient of such a term would have the dimensions of mass^{-1} , and would therefore make the theory non-renormalizable. So for the first time there was a rationale for the simplest version of quantum electrodynamics, in which the gyromagnetic ratio of the electron is two, with small corrections of the sort calculated by Schwinger.

Optimism about quantum field theory rose rapidly, but then again began to decline again. After the initial success of quantum field theory in accounting for the theory of electrons and photons it was natural to try to apply it to the strong interactions and the weak interactions, but this effort failed.

It was not possible at the time to find any renormalizable theory of the weak

interactions. The Fermi coupling constant of beta decay has the dimensions mass^{-2} (it equals about $10^{-5} m_{\text{proton}}^{-2}$), which makes it a nonrenormalizable theory, a theory in which this cancellation of infinities wouldn't work. Although the Fermi theory worked very well in accounting for beta decay, when carried beyond the lowest approximation it gave nonsense; it gave infinite decay rates and cross sections.

For the strong interactions it *was* possible to invent a renormalizable theory, the old pseudoscalar meson theory, but it was very soon realized that the coupling in this theory is so strong that you couldn't justify the use of perturbation theory. Further, if you tried to get by with just the lowest order of perturbation theory, you found a conflict with experiment. You could adjust the coupling constant to give the right cross section for pion nucleon scattering at low energy but then you got the wrong result for pion photoproduction. It just didn't work.

As a result during this period, from about the early 1950s to the late 1960s, there was another decline in our optimism about quantum field theory. During this period a number of ideas were introduced that would dispense with quantum field theory. One of them was S-matrix theory; another was current algebra. I will come back to both of them in the rest of my discussion. But starting in the late 1960s and then accelerating in the 1970s optimism about quantum field theory once again returned with the advent of the Standard Model, first of electromagnetic and weak interactions, and then for the strong interactions. Quantum field theory became the tool of choice for doing calculations in particle physics. I gave a talk at Harvard in the early 1970s whose title reflected the mood of this period; it was "The Renaissance of Quantum Field Theory."

From then on, it was clear that quantum field theory would always be with us. This is the way we do calculations at the energies that are accessible in accelerators. But there began to be doubts whether or not the quantum field theory of the Standard Model was truly a fundamental theory or just the first term in an effective field theory in which there appear every possible interaction allowed by symmetries, the nonrenormalizable as well as the renormalizable ones, perhaps an effective field theory that arises from a deeper underlying theory that might not even be a quantum field theory at all. Of course I'm thinking here about string theory, but that's not the only possibility. It began to be thought that not only the Standard Model but also Einstein's General Theory of Relativity are just low energy approximations to a deeper underlying theory that, as I said, may not be a quantum field theory. Just recently, in the last year or two, there appeared some slight reason for renewed hope that the effective field theory containing General Relativity plus the Standard Model may in fact be all we need. So perhaps we may see renewed optimism for quantum field theory as part of a description of nature at the most fundamental level.

[The remainder of this talk dealt with the development of effective field theory, the origins and applications of the idea that the Standard Model and General Relativity are part of an effective field theory, and the question whether this effective

field theory may be a fundamental, ultraviolet complete, theory. This part of my Prague talk followed pretty much along the lines of a talk I gave in July 2009 at the 6th International Workshop on Chiral Dynamics, at the University of Bern. The proceedings of that workshop will not appear in printed form, so in what follows I present the written version of the talk I gave at Bern.]

2

Effective Field Theory: Past and Future

I have been asked by the organizers of this meeting to “celebrate 30 years” of a paper [1] on effective field theories that I wrote in 1979. I am quoting this request at the outset, because in the first half of this talk I will be reminiscing about my own work on effective field theories, leading up to this 1979 paper. I think it is important to understand how confusing these things seemed back then, and no one knows better than I do how confused I was. But I am sure that many in this audience know more than I do about the applications of effective field theory to the strong interactions since 1979, so I will mention only some early applications to strong interactions and a few applications to other areas of physics. I will then describe how we have come to think that our most fundamental theories, the Standard Model and General Relativity, are the leading terms in an effective field theory. Finally, I will report on recent work of others that lends support to a suggestion that this effective field theory may actually be a fundamental theory, valid at all energies.

It all started with current algebra. As everyone knows, in 1960 Yoichiro Nambu had the idea that the axial vector current of beta decay could be considered to be conserved in the same limit that the pion, the lightest hadron, could be considered massless [2]. This assumption would follow if the axial vector current was associated with a spontaneously broken approximate symmetry, with the pion playing the role of a Goldstone boson [3]. Nambu used this idea to explain the success of the Goldberger-Treiman formula [4] for the pion decay amplitude, and with his collaborators he was able to derive formulas for the rate of emission of a single low energy pion in various collisions [5]. In this work it was not necessary to assume anything about the nature of the broken symmetry – only that there was some approximate symmetry responsible for the approximate conservation of the axial vector current and the approximate masslessness of the pion. But to deal with processes involving more than one pion, it was necessary to use not only the approximate conservation of the current but also the current commutation relations, which of course do depend on the underlying broken symmetry. The technology of using these properties of the currents, in which one does not use any specific Lagrangian for the strong interactions, became known as current algebra.^a It scored a dramatic success in the

^aThe name may be due to Murray Gell-Mann. The current commutation relations were given in [6].

derivation of the Adler-Weisberger sum rule [7] for the axial vector beta decay coupling constant g_A , which showed that the current commutation relations are those of $SU(2) \times SU(2)$.

When I started in the mid-1960s to work on current algebra, I had the feeling that, despite the success of the Goldberger-Treiman relation and the Adler-Weisberger sum rule, there was then rather too much emphasis on the role that the axial vector current plays in weak interactions. After all, even if there were no weak interactions, the fact that the strong interactions have an approximate but spontaneously broken $SU(2) \times SU(2)$ symmetry would be a pretty important piece of information about the strong interactions.^b To demonstrate the point, I was able to use current algebra to derive successful formulas for the pion-pion and pion-nucleon scattering lengths.^c When combined with a well-known dispersion relation [9] and the Goldberger-Treiman relation, these formulas for the pion-nucleon scattering lengths turned out to imply the Adler-Weisberger sum rule.

In 1966 I turned to the problem of calculating the rate of processes in which arbitrary numbers of low energy massless pions are emitted in the collision of other hadrons. This was not a problem that urgently needed to be solved. I was interested in it because a year earlier I had worked out simple formulas for the rate of emission of arbitrary numbers of soft gravitons or photons in any collision [10], and I was curious whether anything equally simple could be said about soft pions. Calculating the amplitude for emission of several soft pions by use of the technique of current algebra turned out to be fearsomely complicated; the non-vanishing commutators of the currents associated with the soft pions prevented the derivation of anything as simple as the results for soft photons or gravitons, except in the special case in which all pions have the same charge [11].

Then some time late in 1966 I was sitting at the counter of a café in Harvard Square, scribbling on a napkin the amplitudes I had found for emitting two or three soft pions in nucleon collisions, and it suddenly occurred to me that these results looked very much like what would be given by lowest order Feynman diagrams in a quantum field theory in which pion lines are emitted from the external nucleon lines, with a Lagrangian in which the nucleon interacts with one, two, and more pion fields. Why should this be? Remember, this was a time when theorists had pretty well given up the idea of applying specific quantum field theories to the strong interactions, because there was no reason to trust the lowest order of perturbation theory, and no way to sum the perturbation series. What was popular was to exploit tools such as current algebra and dispersion relations that did not rely on assumptions about particular Lagrangians.

The best explanation that I could give then for the field-theoretic appearance

^bI emphasized this point in my rapporteur's talk on current algebra at the 1968 "Rochester" conference; see *Proceedings of the 14th International Conference on High-Energy Physics*, p. 253.

^cSee [8]. The pion-nucleon scattering lengths were calculated independently by Y. Tomozawa, *Nuovo Cimento* 46A, 707 (1966).

of the current algebra results was that these results for the emission of n soft pions in nucleon collisions are of the minimum order, G_π^n , in the pion-nucleon coupling constant G_π , except that one had to use the exact values for the collision amplitudes without soft pion emission, and divide by factors of the axial vector coupling constant $g_A \simeq 1.2$ in appropriate places. Therefore any Lagrangian that satisfied the axioms of current algebra would have to give the same answer as current algebra in lowest order perturbation theory, except that it would have to be a field theory in which soft pions were emitted only from external lines of the diagram for the nucleon collisions, for only then would one know how to put in the correct factors of g_A and the correct nucleon collision amplitude.

The time-honored renormalizable theory of nucleons and pions with conserved currents that satisfied the assumptions of current algebra was the “linear σ -model,”^d with Lagrangian (in the limit of exact current conservation):

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2} [\partial_\mu \vec{\pi} \cdot \partial^\mu \vec{\pi} + \partial_\mu \sigma \partial^\mu \sigma] \\ & -\frac{m^2}{2} (\sigma^2 + \vec{\pi}^2) - \frac{\lambda}{4} (\sigma^2 + \vec{\pi}^2)^2 \\ & -\bar{N} \gamma^\mu \partial_\mu N - G_\pi \bar{N} \left(\sigma + 2i\gamma_5 \vec{\pi} \cdot \vec{t} \right) N, \end{aligned} \quad (1)$$

where N , $\vec{\pi}$, and σ are the fields of the nucleon doublet, pion triplet, and a scalar singlet, and $\vec{t}$ is the nucleon isospin matrix (with $\vec{t}^2 = 3/4$). This Lagrangian has an $SU(2) \times SU(2)$ symmetry (equivalent as far as current commutation relations are concerned to an $SO(4)$ symmetry), that is spontaneously broken for $m^2 < 0$ by the expectation value of the σ field, given in lowest order by $\langle \sigma \rangle = F/2 \equiv \sqrt{-m^2/\lambda}$, which also gives the nucleon a lowest order mass $2G_\pi F$. But with a Lagrangian of this form soft pions could be emitted from internal as well as external lines of the graphs for the nucleon collision itself, and there would be no way to evaluate the pion emission amplitude without having to sum over the infinite number of graphs for the nucleon collision amplitude.

To get around this obstacle, I used the chiral $SO(4)$ symmetry to rotate the chiral four-vector into the fourth direction

$$(\vec{\pi}, \sigma) \mapsto (0, \sigma'), \quad \sigma' = \sqrt{\sigma^2 + \vec{\pi}^2}, \quad (2)$$

with a corresponding chiral transformation $N \mapsto N'$ of the nucleon doublet. The chiral symmetry of the Lagrangian would result in the pion disappearing from the Lagrangian, except that the matrix of the rotation (2) necessarily, like the fields, depends on spacetime position, while the theory is only invariant under spacetime-independent chiral rotations. The pion field thus reappears as a parameter in the

^dSee [12]. This theory, with the inclusion of a symmetry-breaking term proportional to the σ field, was intended to provide an illustration of a “partially conserved axial current,” that is, one whose divergence is proportional to the pion field.

$SO(4)$ rotation (2), which could conveniently be taken as

$$\vec{\pi}' \equiv F\vec{\pi}/[\sigma + \sigma'] . \quad (3)$$

But the rotation parameter $\vec{\pi}'$ would not appear in the transformed Lagrangian if it were independent of the spacetime coordinates, so wherever it appears it must be accompanied with at least one derivative. This derivative produces a factor of pion four-momentum in the pion emission amplitude, which would suppress the amplitude for emitting soft pions, if this factor were not compensated by the pole in the nucleon propagator of an external nucleon line to which the pion emission vertex is attached. Thus, with the Lagrangian in this form, pions of small momenta can only be emitted from external lines of a nucleon collision amplitude. This is what I needed.

Since σ' is chiral-invariant, it plays no role in maintaining the chiral invariance of the theory, and could therefore be replaced with its lowest-order expectation value $F/2$. The transformed Lagrangian (now dropping primes) is then

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2} \left[1 + \frac{\vec{\pi}^2}{F^2} \right]^{-2} \partial_\mu \vec{\pi} \cdot \partial^\mu \vec{\pi} \\ & - \bar{N} \left[\gamma^\mu \partial_\mu + G_\pi F/2 \right. \\ & \left. + i\gamma^\mu \left[1 + \frac{\vec{\pi}^2}{F^2} \right]^{-1} \left[\frac{2}{F} \gamma_5 \vec{t} \cdot \partial_\mu \vec{\pi} + \frac{2}{F^2} \vec{t} \cdot (\vec{\pi} \times \partial_\mu \vec{\pi}) \right] \right] N . \end{aligned} \quad (4)$$

In order to reproduce the results of current algebra, it is only necessary to identify F as the pion decay amplitude $F_\pi \simeq 184$ MeV, replace the term $G_\pi F/2$ in the nucleon bilinear with the actual nucleon mass m_N (given by the Goldberger–Treiman relation as $G_\pi F_\pi/2g_A$), and replace the pseudovector pion-nucleon coupling $1/F$ with its actual value $G_\pi/2m_N = g_A/F_\pi$. This gives an effective Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{eff}} = & -\frac{1}{2} \left[1 + \frac{\vec{\pi}^2}{F_\pi^2} \right]^{-2} \partial_\mu \vec{\pi} \cdot \partial^\mu \vec{\pi} \\ & - \bar{N} \left[\gamma^\mu \partial_\mu + m_N \right. \\ & \left. + i\gamma^\mu \left[1 + \frac{\vec{\pi}^2}{F_\pi^2} \right]^{-1} \left[\frac{G_\pi}{m_N} \gamma_5 \vec{t} \cdot \partial_\mu \vec{\pi} + \frac{2}{F_\pi^2} \vec{t} \cdot (\vec{\pi} \times \partial_\mu \vec{\pi}) \right] \right] N . \end{aligned} \quad (5)$$

To take account of the finite pion mass, the linear sigma model also includes a chiral-symmetry breaking perturbation proportional to σ . Making the chiral rotation (2), replacing σ' with the constant $F/2$, and adjusting the coefficient of this term to give the physical pion mass m_π gives a chiral symmetry breaking term

$$\Delta \mathcal{L}_{\text{eff}} = -\frac{1}{2} \left[1 + \frac{\vec{\pi}^2}{F_\pi^2} \right]^{-1} m_\pi^2 \vec{\pi}^2 . \quad (6)$$

Using $\mathcal{L}_{\text{eff}} + \Delta\mathcal{L}_{\text{eff}}$ in lowest order perturbation theory, I found the same results for low-energy pion-pion and pion-nucleon scattering that I had obtained earlier with much greater difficulty by the methods of current algebra [13].

A few months after this work, Julian Schwinger remarked to me that it should be possible to skip this complicated derivation, forget all about the linear σ -model, and instead infer the structure of the Lagrangian directly from the non-linear chiral transformation properties of the pion field appearing in (5).^e It was a good idea. I spent the summer of 1967 working out these transformation properties, and what they imply for the structure of the Lagrangian [15]. It turns out that if we require that the pion field has the usual linear transformation under $SO(3)$ isospin rotations (because isospin symmetry is supposed to be not spontaneously broken), then there is a *unique* $SO(4)$ chiral transformation that takes the pion field into a function of itself — unique, that is, up to possible redefinition of the field. For an infinitesimal $SO(4)$ rotation by an angle ϵ in the $a4$ plane (where $a = 1, 2, 3$), the pion field π_b (labelled with a prime in Eq. (3)) changes by an amount

$$\delta_a \pi_b = -i\epsilon F_\pi \left[\frac{1}{2} \left(1 - \frac{\vec{\pi}^2}{F_\pi^2} \right) \delta_{ab} + \frac{\pi_a \pi_b}{F_\pi^2} \right]. \quad (7)$$

Any other field ψ , on which isospin rotations act with a matrix $\vec{t}$, is changed by an infinitesimal chiral rotation in the $a4$ plane by an amount

$$\delta_a \psi = \frac{\epsilon}{F_\pi} \left(\vec{t} \times \vec{\pi} \right)_a \psi. \quad (8)$$

This is just an ordinary, though position-dependent, isospin rotation, so a non-derivative isospin-invariant term in the Lagrangian that does not involve pions, like the nucleon mass term $-m_N \bar{N}N$, is automatically chiral-invariant. The terms in Eq. (5):

$$-\bar{N} \left[\gamma^\mu \partial_\mu + \frac{2i}{F_\pi^2} \gamma^\mu \left[1 + \frac{\vec{\pi}^2}{F_\pi^2} \right]^{-1} \vec{t} \cdot (\vec{\pi} \times \partial_\mu \vec{\pi}) \right] N, \quad (9)$$

and

$$-i \frac{G_\pi}{m_N} \left[1 + \frac{\vec{\pi}^2}{F_\pi^2} \right]^{-1} \bar{N} \gamma^\mu \gamma_5 \vec{t} \cdot \partial_\mu \vec{\pi} N, \quad (10)$$

are simply proportional to the most general chiral-invariant nucleon-pion interactions with a single spacetime derivative. The coefficient of the term (9) is fixed by the condition that N should be canonically normalized, while the coefficient of (10)

^eFor Schwinger's own development of this idea, see [14]. It is interesting that in deriving the effective field theory of goldstinos in supergravity theories, it is much more transparent to start with a theory with linearly realized supersymmetry and impose constraints analogous to setting $\sigma' = F/2$, than to work from the beginning with supersymmetry realized non-linearly, in analogy to Eq. (7); see Z. Komargodski and N. Seiberg, to be published.

is chosen to agree with the conventional definition of the pion-nucleon coupling G_π , and is not directly constrained by chiral symmetry. The term

$$-\frac{1}{2} \left[1 + \frac{\vec{\pi}^2}{F^2} \right]^{-2} \partial_\mu \vec{\pi} \cdot \partial^\mu \vec{\pi} \quad (11)$$

is proportional to the most general chiral invariant quantity involving the pion field and no more than two spacetime derivatives, with a coefficient fixed by the condition that $\vec{\pi}$ should be canonically normalized. The chiral symmetry breaking term (6) is the most general function of the pion field without derivatives that transforms as the fourth component of a chiral four-vector. None of this relies on the methods of current algebra, though one can use the Lagrangian (5) to calculate the Noether current corresponding to chiral transformations, and recover the Goldberger-Treiman relation in the form $g_A = G_\pi F_\pi / 2m_N$.

This sort of direct analysis was subsequently extended by Callan, Coleman, Wess, and Zumino to the transformation and interactions of the Goldstone boson fields associated with the spontaneous breakdown of any Lie group G to any subgroup H [16]. Here, too, the transformation of the Goldstone boson fields is unique, up to a redefinition of the fields, and the transformation of other fields under G is uniquely determined by their transformation under the unbroken subgroup H . It is straightforward to work out the rules for using these ingredients to construct effective Lagrangians that are invariant under G as well as H .^f Once again, the key point is that the invariance of the Lagrangian under G would eliminate all presence of the Goldstone boson field in the Lagrangian if the field were spacetime-independent, so wherever functions of this field appear in the Lagrangian they are always accompanied with at least one spacetime derivative.

In the following years, effective Lagrangians with spontaneously broken $SU(2) \times SU(2)$ or $SU(3) \times SU(3)$ symmetry were widely used in lowest-order perturbation theory to make predictions about low energy pion and kaon interactions.^g But during this period, from the late 1960s to the late 1970s, like many other particle physicists I was chiefly concerned with developing and testing the Standard Model of elementary particles. As it happened, the Standard Model did much to clarify the basis for chiral symmetry. Quantum chromodynamics with N light quarks is automatically invariant under a $SU(N) \times SU(N)$ chiral symmetry,^h broken in the

^fThere is a complication. In some cases, such as $SU(3) \times SU(3)$ spontaneously broken to $SU(3)$, fermion loops produce G -invariant terms in the action that are not the integrals of G -invariant terms in the Lagrangian density; see [17]. The most general such terms in the action, whether or not produced by fermion loops, have been cataloged by E. D'Hoker and S. Weinberg [18]. It turns out that for $SU(N) \times SU(N)$ spontaneously broken to the diagonal $SU(N)$, there is just one such term for $N \geq 3$, and none for $N = 1$ or 2 . For $N = 3$, this term is the one found by Wess and Zumino.

^gFor reviews, see S. Weinberg, in *Lectures on Elementary Particles and Quantum Field Theory* — 1970 Brandeis University Summer Institute in Theoretical Physics, Vol. 1, ed. S. Deser, M. Grisaru, and H. Pendleton (The M.I.T. Press, Cambridge, MA, 1970); B. W. Lee, *Chiral Dynamics* (Gordon and Breach, New York, 1972).

^hFor a while it was not clear why there was not also a chiral $U(1)$ symmetry, that would also be

Lagrangian only by quark masses, and the electroweak theory tells us that the currents of this symmetry (along with the quark number currents) are just those to which the $W^\pm$, Z^0 , and photon are coupled.

During this whole period, effective field theories appeared as only a device for more easily reproducing the results of current algebra. It was difficult to take them seriously as dynamical theories, because the derivative couplings that made them useful in the lowest order of perturbation theory also made them nonrenormalizable, thus apparently closing off the possibility of using these theories in higher order.

My thinking about this began to change in 1976. I was invited to give a series of lectures at Erice that summer, and took the opportunity to learn the theory of critical phenomena by giving lectures about it [20]. In preparing these lectures, I was struck by Kenneth Wilson's device of "integrating out" short-distance degrees of freedom by introducing a variable ultraviolet cutoff, with the bare couplings given a cutoff dependence that guaranteed that physical quantities are cutoff independent. Even if the underlying theory is renormalizable, once a finite cutoff is introduced it becomes necessary to introduce every possible interaction, renormalizable or not, to keep physics strictly cutoff independent. From this point of view, it doesn't make much difference whether the underlying theory is renormalizable or not. Indeed, I realized that even without a cutoff, as long as every term allowed by symmetries is included in the Lagrangian, there will always be a counterterm available to absorb every possible ultraviolet divergence by renormalization of the corresponding coupling constant. Non-renormalizable theories, I realized, are just as renormalizable as renormalizable theories.

This opened the door to the consideration of a Lagrangian containing terms like (5) as the basis for a legitimate dynamical theory, not limited to the tree approximation, provided one adds every one of the infinite number of other, higher-derivative, terms allowed by chiral symmetry.¹ But for this to be useful, it is necessary that in some sort of perturbative expansion, only a finite number of terms in the Lagrangian can appear in each order of perturbation theory.

In chiral dynamics, this perturbation theory is provided by an expansion in powers of small momenta and pion masses. At momenta of order m_π , the number ν of factors of momenta or m_π contributed by a diagram with L loops, E_N external nucleon lines, and V_i vertices of type i , for any reaction among pions and/or nucleons, is

$$\nu = \sum_i V_i \left(d_i + \frac{n_i}{2} + m_i - 2 \right) + 2L + 2 - \frac{E_N}{2}, \quad (12)$$

where d_i , n_i , and m_i are respectively the numbers of derivatives, factors of nucleon

broken in the Lagrangian only by the quark masses, and would either lead to a parity doubling of observed hadrons, or to a new light pseudoscalar neutral meson, both of which possibilities were experimentally ruled out. It was not until 1976 that 't Hooft pointed out that the effect of triangle anomalies in the presence of instantons produced an intrinsic violation of this unwanted chiral $U(1)$ symmetry; see G. 't Hooft [19].

¹I thought it appropriate to publish this in a festschrift for Julian Schwinger, see [1].

fields, and factors of pion mass (or more precisely, half the number of factors of u and d quark masses) associated with vertices of type i . As a consequence of chiral symmetry, the minimum possible value of $d_i + n_i/2 + m_i$ is 2, so the leading diagrams for small momenta are those with $L = 0$ and any number of interactions with $d_i + n_i/2 + m_i = 2$, which are the ones given in Eqs. (5) and (6). To next order in momenta, we may include tree graphs with any number of vertices with $d_i + n_i/2 + m_i = 2$ and just one vertex with $d_i + n_i/2 + m_i = 3$ (such as the so-called σ -term). To next order, we include any number of vertices with $d_i + n_i/2 + m_i = 2$, plus either a single loop, or a single vertex with $d_i + n_i/2 + m_i = 4$ which provides a counterterm for the infinity in the loop graph, or two vertices with $d_i + n_i/2 + m_i = 3$. And so on. Thus one can generate a power series in momenta and m_π , in which only a few new constants need to be introduced at each new order. As an explicit example of this procedure, I calculated the one-loop corrections to pion-pion scattering in the limit of zero pion mass, and of course I found the sort of corrections required to this order by unitarity.[‡]

But even if this procedure gives well-defined finite results, how do we know they are true? It would be extraordinarily difficult to justify any calculation involving loop graphs using current algebra. For me in 1979, the answer involved a radical reconsideration of the nature of quantum field theory. From its beginning in the late 1920s, quantum field theory had been regarded as the application of quantum mechanics to fields that are among the fundamental constituents of the universe — first the electromagnetic field, and later the electron field and fields for other known “elementary” particles. In fact, this became a working definition of an elementary particle — it is a particle with its own field. But for years in teaching courses on quantum field theory I had emphasized that the description of nature by quantum field theories is inevitable, at least in theories with a finite number of particle types, once one assumes the principles of relativity and quantum mechanics, plus the cluster decomposition principle, which requires that distant experiments have uncorrelated results. So I began to think that although specific quantum field theories may have a content that goes beyond these general principles, quantum field theory itself does not. I offered this in my 1979 paper as what Arthur Wightman would call a folk theorem: “if one writes down the most general possible Lagrangian, including *all* terms consistent with assumed symmetry principles, and then calculates matrix elements with this Lagrangian to any given order of perturbation theory, the result will simply be the most general possible S -matrix consistent with perturbative unitarity, analyticity, cluster decomposition, and the assumed symmetry properties.” So current algebra wasn’t needed.

There was an interesting irony in this. I had been at Berkeley from 1959 to 1966, when Geoffrey Chew and his collaborators were elaborating a program for calculating S -matrix elements for strong interaction processes by the use of unitar-

[‡]Unitarity corrections to soft-pion results of current algebra had been considered earlier by H. Schnitzer, L.-F. Li and H. Pagels, and P. Langacker and H. Pagels, see [21].

ity, analyticity, and Lorentz invariance, without reference to quantum field theory. I found it an attractive philosophy, because it relied only on a minimum of principles, all well established. Unfortunately, the S -matrix theorists were never able to develop a reliable method of calculation, so I worked instead on other things, including current algebra. Now in 1979 I realized that the assumptions of S -matrix theory, supplemented by chiral invariance, were indeed all that are needed at low energy, but the most convenient way of implementing these assumptions in actual calculations was by good old quantum field theory, which the S -matrix theorists had hoped to supplant.

After 1979, effective field theories were applied to strong interactions in work by Gasser, Leutwyler, Meissner, and many other theorists. My own contributions to this work were limited to two areas — isospin violation, and nuclear forces.

At first in the development of chiral dynamics there had been a tacit assumption that isotopic spin symmetry was a better approximate symmetry than chiral $SU(2) \times SU(2)$, and that the Gell-Mann–Ne’eman $SU(3)$ symmetry was a better approximate symmetry than chiral $SU(3) \times SU(3)$. This assumption became untenable with the calculation of quark mass ratios from the measured pseudoscalar meson masses [22]. It turns out that the d quark mass is almost twice the u quark mass, and the s quark mass is very much larger than either. As a consequence of the inequality of d and u quark masses, chiral $SU(2) \times SU(2)$ is broken in the Lagrangian of quantum chromodynamics not only by the fourth component of a chiral four-vector, as in (6), but also by the third component of a different chiral four-vector proportional to $m_u - m_d$ (whose fourth component is a pseudoscalar). There is no function of the pion field alone, without derivatives, with the latter transformation property, which is why pion–pion scattering and the pion masses are described by (6) and the first term in (5) in leading order, with no isospin breaking aside of course from that due to electromagnetism. But there are non-derivative corrections to pion–nucleon interactions,^k which at momenta of order m_π are suppressed relative to the derivative coupling terms in (5) by just one factor of m_π or momenta:

$$\begin{aligned} \Delta' \mathcal{L}_{\text{eff}} = & -\frac{A}{2} \left(\frac{1 - \pi^2/F_\pi^2}{1 + \pi^2/F_\pi^2} \right) \bar{N}N \\ & -B \left[\bar{N}t_3N - \frac{2}{F_\pi^2} \left(\frac{\pi_3}{1 + \pi^2/F_\pi^2} \right) \bar{N}\vec{t} \cdot \vec{\pi}N \right] \\ & -\frac{iC}{1 + \pi^2/F_\pi^2} \bar{N}\gamma_5 \vec{\pi} \cdot \vec{t}N \\ & -\frac{iD\pi_3}{1 + \pi^2/F_\pi^2} \bar{N}\gamma_5 N, \end{aligned} \quad (13)$$

where A and C are proportional to $m_u + m_d$, and B and D are proportional to

^kS. Weinberg, in *Chiral Dynamics: Theory and Experiment — Proceedings of the Workshop Held at MIT, July 1994* (Springer-Verlag, Berlin, 1995). The terms in Eq. (13) that are odd in the pion field are given in Section 19.5 of S. Weinberg, *The Quantum Theory of Fields*, Vol. II (Cambridge University Press, 1996)

$m_u - m_d$, with $B \simeq -2.5$ MeV. The A and B terms contribute isospin conserving and violating terms to the so-called σ -term in pion nucleon scattering.

My work on nuclear forces began one day in 1990 while I was lecturing to a graduate class at Texas. I derived Eq. (12) for the class, and showed how the interactions in the leading tree graphs with $d_i + n_i/2 + m_i = 2$ were just those given here in Eqs. (5) and (6). Then, while I was standing at the blackboard, it suddenly occurred to me that there was one other term with $d_i + n_i/2 + m_i = 2$ that I had never previously considered: an interaction with no factors of pion mass and no derivatives (and hence, according to chiral symmetry, no pions), but *four* nucleon fields — that is, a sum of Fermi interactions $(\bar{N}\Gamma N)(\bar{N}\Gamma' N)$, with any matrices Γ and Γ' allowed by Lorentz invariance, parity conservation, and isospin conservation. This is just the sort of “hard core” nucleon–nucleon interaction that nuclear theorists had long known has to be added to the pion-exchange term in theories of nuclear force. But there is a complication — in graphs for nucleon–nucleon scattering at low energy, two-nucleon intermediate states make a large contribution that invalidates the sort of power-counting that justifies the use of the effective Lagrangian (5), (6) in processes involving only pions, or one low-energy nucleon plus pions. So it is necessary to apply the effective Lagrangian, including the terms $(\bar{N}\Gamma N)(\bar{N}\Gamma' N)$ along with the terms (5) and (6), to the two-nucleon irreducible nucleon–nucleon potential, rather than directly to the scattering amplitude [23]. This program was initially carried further by Ordoñez, van Kolck, Friar, and their collaborators [24], and eventually by several others.

The advent of effective field theories generated changes in point of view and suggested new techniques of calculation that propagated out to numerous areas of physics, some quite far removed from particle physics. Notable here is the use of the power-counting arguments of effective field theory to justify the approximations made in the BCS theory of superconductivity [25]. Instead of counting powers of small momenta, one must count powers of the departures of momenta from the Fermi surface. Also, general features of theories of inflation have been clarified by re-casting these theories as effective field theories of the inflaton and gravitational fields [26].

Perhaps the most important lesson from chiral dynamics was that we should keep an open mind about renormalizability. The renormalizable Standard Model of elementary particles may itself be just the first term in an effective field theory that contains every possible interaction allowed by Lorentz invariance and the $SU(3) \times SU(2) \times U(1)$ gauge symmetry, only with the non-renormalizable terms suppressed by negative powers of some very large mass M , just as the terms in chiral dynamics with more derivatives than in Eq. (5) are suppressed by negative powers of $2\pi F_\pi \approx m_N$. One indication that there is a large mass scale in some theory underlying the Standard Model is the well-known fact that the three (suitably normalized) running gauge couplings of $SU(3) \times SU(2) \times U(1)$ become equal at an energy of the order of 10^{15} GeV (or, if supersymmetry is assumed, 2×10^{16} GeV, with better convergence of the couplings.)

In 1979 both Frank Wilczek and Tony Zee [27] and I [28] independently pointed out that, while the renormalizable terms of the Standard Model cannot violate baryon or lepton conservation,^l this is not true of the higher non-renormalizable terms. In particular, four-fermion terms can generate a proton decay into antileptons, though not into leptons, with an amplitude suppressed on dimensional grounds by a factor M^{-2} . The conservation of baryon and lepton number in observed physical processes thus may be an accident, an artifact of the necessary simplicity of the leading renormalizable $SU(3) \times SU(2) \times U(1)$ -invariant interactions. I also noted at the same time that interactions between a pair of lepton doublets and a pair of scalar doublets can generate a neutrino mass, which is suppressed only by a factor M^{-1} , and that therefore with a reasonable estimate of M could produce observable neutrino oscillations. The subsequent confirmation of neutrino oscillations lends support to the view of the Standard Model as an effective field theory, with M somewhere in the neighborhood of 10^{16} GeV.

Of course, these non-renormalizable terms can be (and in fact, had been) generated in various renormalizable grand-unified theories by integrating out the heavy particles in these theories. Some calculations in the resulting theories can be assisted by treating them as effective field theories.^m But the important point is that the existence of suppressed baryon- and lepton-nonconserving terms, and some of their detailed properties, should be expected on much more general grounds, *even if the underlying theory is not a quantum field theory at all*. Indeed, from the 1980s on, it has been increasingly popular to suppose that the theory underlying the Standard Model as well as general relativity is a string theory.

Which brings me to gravitation. Just as we have learned to live with the fact that there is no renormalizable theory of pion fields that is invariant under the chiral transformation (7), so also we should not despair of applying quantum field theory to gravitation just because there is no renormalizable theory of the metric tensor that is invariant under general coordinate transformations. It increasingly seems apparent that the Einstein–Hilbert Lagrangian $\sqrt{g}R$ is just the least suppressed term in the Lagrangian of an effective field theory containing every possible generally covariant function of the metric and its derivatives. The application of this point of view to long range properties of gravitation has been most thoroughly developed by John Donoghue and his collaborators [32]. One consequence of viewing the Einstein–Hilbert Lagrangian as one term in an effective field theory is that any theorem based on conventional general relativity, which declares that under certain initial

^lThis is not true if the effective theory contains fields for the squarks and sleptons of supersymmetry. However, there are no renormalizable baryon or lepton violating terms in “split supersymmetry” theories, in which the squarks and sleptons are superheavy, and only the gauginos and perhaps higgsinos survive to ordinary energies, see [29].

^mThe effective field theories derived by integrating out heavy particles had been considered by T. Appelquist and J. Carrazone [30]. In 1980, in a paper titled “Effective Gauge Theories,” I used the techniques of effective field theory to evaluate the effects of integrating out the heavy gauge bosons in grand unified theories on the initial conditions for the running of the gauge couplings down to accessible energies [31].

conditions future singularities are inevitable, must be reinterpreted to mean that under these conditions higher terms in the effective action become important.

Of course, there is a problem — the effective theory of gravitation cannot be used at very high energies, say of the order of the Planck mass, no more than chiral dynamics can be used above a momentum of order $2\pi F_\pi \approx 1$ GeV. For purposes of the subsequent discussion, it is useful to express this problem in terms of the Wilsonian renormalization group. The effective action for gravitation takes the form

$$I_{\text{eff}} = - \int d^4x \sqrt{-\text{Det}g} \left[f_0(\Lambda) + f_1(\Lambda)R + f_{2a}(\Lambda)R^2 + f_{2b}(\Lambda)R^{\mu\nu}R_{\mu\nu} + f_{3a}(\Lambda)R^3 + \dots \right], \quad (14)$$

where here Λ is the ultraviolet cutoff, and the $f_n(\Lambda)$ are coupling parameters with a cutoff dependence chosen so that physical quantities are cutoff-independent. We can replace these coupling parameters with dimensionless parameters $g_n(\Lambda)$:

$$g_0 \equiv \Lambda^{-4}f_0; \quad g_1 \equiv \Lambda^{-2}f_1; \quad g_{2a} \equiv f_{2a}; \quad g_{2b} \equiv f_{2b}; \quad g_{3a} \equiv \Lambda^2f_{3a}; \quad \dots \quad (15)$$

Because dimensionless, these parameters must satisfy a renormalization group equation of the form

$$\Lambda \frac{d}{d\Lambda} g_n(\Lambda) = \beta_n(g(\Lambda)). \quad (16)$$

In perturbation theory, all but a finite number of the $g_n(\Lambda)$ go to infinity as $\Lambda \rightarrow \infty$, which if true would rule out the use of this theory to calculate anything at very high energy. There are even examples, like the Landau pole in quantum electrodynamics and the phenomenon of “triviality” in scalar field theories, in which the couplings blow up at a *finite* value of Λ .

It is usually assumed that this explosion of the dimensionless couplings at high energy is irrelevant in the theory of gravitation, just as it is irrelevant in chiral dynamics. In chiral dynamics, it is understood that at energies of order $2\pi F_\pi \approx m_N$, the appropriate degrees of freedom are no longer pion and nucleon fields, but rather quark and gluon fields. In the same way, it is usually assumed that in the quantum theory of gravitation, when Λ reaches some very high energy, of the order of 10^{15} to 10^{18} GeV, the appropriate degrees of freedom are no longer the metric and the Standard Model fields, but something very different, perhaps strings.

But maybe not. It is just possible that the appropriate degrees of freedom at all energies are the metric and matter fields, including those of the Standard Model. The dimensionless couplings can be protected from blowing up if they are attracted to a finite value g_{n*} . This is known as *asymptotic safety*.ⁿ

ⁿThis was first proposed in my 1976 Erice lectures, see [20].

Quantum chromodynamics provides an example of asymptotic safety, but one in which the theory at high energies is not only safe from exploding couplings, but also free. In the more general case of asymptotic safety, the high energy limit g_{n*} is finite, but not commonly zero.

For asymptotic safety to be possible, it is necessary that all the beta functions should vanish at g_{n*} :

$$\beta_n(g_*) = 0 . \quad (17)$$

It is also necessary that the physical couplings should be on a trajectory that is attracted to g_{n*} . The number of independent parameters in such a theory equals the dimensionality of the surface, known as the *ultraviolet critical surface*, formed by all the trajectories that are attracted to the fixed point. This dimensionality had better be finite, if the theory is to have any predictive power at high energy. For an asymptotically safe theory with a finite-dimensional ultraviolet critical surface, the requirement that couplings lie on this surface plays much the same role as the requirement of renormalizability in quantum chromodynamics — it provides a rational basis for limiting the complexity of the theory.

This dimensionality of the ultraviolet critical surface can be expressed in terms of the behavior of $\beta_n(g)$ for g near the fixed point g_* . Barring unexpected singularities, in this case we have

$$\beta_n(g) \rightarrow \sum_m B_{nm}(g_m - g_{*m}) , \quad B_{nm} \equiv \left(\frac{\partial \beta_n(g)}{\partial g_m} \right)_* . \quad (18)$$

The solution of Eq. (16) for g near g_* is then

$$g_n(\Lambda) \rightarrow g_{n*} + \sum_i u_{in} \Lambda^{\lambda_i} , \quad (19)$$

where λ_i and u_{in} are the eigenvalues and suitably normalized eigenvectors of B_{nm} :

$$\sum_m B_{nm} u_{im} = \lambda_i u_{in} . \quad (20)$$

Because B_{nm} is real but not symmetric, the eigenvalues are either real, or come in pairs of complex conjugates. The dimensionality of the ultraviolet critical surface is therefore equal to the number of eigenvalues of B_{nm} with negative real part. The condition that the couplings lie on this surface can be regarded as a generalization of the condition that quantum chromodynamics, if it were a fundamental and not merely an effective field theory, would have to involve only renormalizable couplings.

It may seem unlikely that an infinite matrix like B_{nm} should have only a finite number of eigenvalues with negative real part, but in fact examples of this are quite common. As we learned from the Wilson–Fisher theory of critical phenomena, when a substance undergoes a second-order phase transition, its parameters are subject to a renormalization group equation that has a fixed point, with a single infrared-repulsive direction, so that adjustment of a single parameter such as the temperature or the pressure can put the parameters of the theory on an infrared attractive

surface of co-dimension one, leading to long-range correlations. The single infrared-repulsive direction is at the same time a unique ultraviolet-attractive direction, so the ultraviolet critical surface in such a theory is a one-dimensional curve. Of course, the parameters of the substance on this curve do not really approach a fixed point at very short distances, because at a distance of the order of the interparticle spacing the effective field theory describing the phase transition breaks down.

What about gravitation? There are indications that here too there is a fixed point, with an ultraviolet critical surface of finite dimensionality. Fixed points have been found (of course with $g_{n*} \neq 0$) using dimensional continuation from $2 + \epsilon$ to 4 spacetime dimensions [33], by a $1/N$ approximation (where N is the number of added matter fields) [34], by lattice methods [35], and by use of the truncated exact renormalization group equation [36] initiated in 1998 by Martin Reuter. In the last method, which had earlier been introduced in condensed matter physics [37] and then carried over to particle theory [38] one derives an exact renormalization group equation for the total vacuum amplitude $\Gamma[g, \Lambda]$ in the presence of a background metric $g_{\mu\nu}$ with an *infrared* cutoff Λ . This is the action to be used in calculations of the true vacuum amplitude in calculations of graphs with an *ultraviolet* cutoff Λ . To have equations that can be solved, it is necessary to truncate these renormalization group equations, writing $\Gamma[g, \Lambda]$ as a sum of just a finite number of terms like those shown explicitly in Eq. (14), and ignoring the fact that the beta function inevitably does not vanish for the couplings of other terms in $\Gamma[g, \Lambda]$ that in the given truncation are assumed to vanish.

Initially only two terms were included in the truncation of $\Gamma[g, \Lambda]$ (a cosmological constant and the Einstein–Hilbert term $\sqrt{g}R$), and a fixed point was found with two eigenvalues λ_i , a pair of complex conjugates with negative real part. Then a third operator ($R_{\mu\nu}R^{\mu\nu}$ or the equivalent) was added, and a third eigenvalue was found, with λ_i real and negative. This was not encouraging. If each time that new terms were included in the truncation, new eigenvalues appeared with negative real part, then the ultraviolet critical surface would be infinite dimensional, and the theory, though free of couplings that exploded at high energy, would lose all predictive value at high energy.

In just the last few years calculations have been done that allow more optimism. Codello, Percacci, and Rahmede [39] have considered a Lagrangian containing all terms $\sqrt{g}R^n$ with n running from zero to a maximum value $n_{\max}$, and find that the ultraviolet critical surface has dimensionality 3 even when $n_{\max}$ exceeds 2, up to the highest value $n_{\max} = 6$ that they considered, for which the space of coupling constants is 7-dimensional. Furthermore, the three eigenvalues they find with negative real part seem to converge as $n_{\max}$ increases, as shown in the following table

of ultraviolet-attractive eigenvalues:

$n_{\max} = 2 :$	$-1.38 \pm 2.32i$	-26.8
$n_{\max} = 3 :$	$-2.71 \pm 2.27i$	-2.07
$n_{\max} = 4 :$	$-2.86 \pm 2.45i$	-1.55
$n_{\max} = 5 :$	$-2.53 \pm 2.69i$	-1.78
$n_{\max} = 6 :$	$-2.41 \pm 2.42i$	-1.50

In a subsequent paper [40] they added matter fields, and again found just three ultraviolet-attractive eigenvalues. Further, this year Benedetti, Machado, and Saueressig [41] considered a truncation with a different four terms, terms proportional to $\sqrt{g}R^n$ with $n = 0, 1$ and 2 and also $\sqrt{g}C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma}$ (where $C_{\mu\nu\rho\sigma}$ is the Weyl tensor) and they too find just three ultraviolet-attractive eigenvalues, also when matter is added. If this pattern of eigenvalues continues to hold in future calculations, it will begin to look as if there is a quantum field theory of gravitation that is well-defined at all energies, and that has just three free parameters.

The natural arena for application of these ideas is in the physics of gravitation at small distance scales and high energy — specifically, in the early universe. A start in this direction has been made by Reuter and his collaborators [42], but much remains to be done.

I am grateful for correspondence about recent work on asymptotic safety with D. Benedetti, D. Litim, R. Percacci, and M. Reuter, and to G. Colangelo and J. Gasser for inviting me to give this talk. This material is based in part on work supported by the National Science Foundation under Grant NO. PHY-0455649 and with support from The Robert A. Welch Foundation, Grant No. F-0014.

References

- [1] S. Weinberg, *Physica* A96, 327 (1979).
- [2] Y. Nambu, *Phys. Rev. Lett.* 4, 380 (1960).
- [3] J. Goldstone, *Nuovo Cimento* 9, 154 (1961); Y. Nambu and G. Jona-Lasinio, *Phys. Rev.* 122, 345 (1961); J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* 127, 965 (1962).
- [4] M. L. Goldberger and S. B. Treiman, *Phys. Rev.* 111, 354 (1956).
- [5] Y. Nambu and D. Lurie, *Phys. Rev.* 125, 1429 (1962); Y. Nambu and E. Shrauner, *Phys. Rev.* 128, 862 (1962).
- [6] M. Gell-Mann, *Physics* 1, 63 (1964).
- [7] S. L. Adler, *Phys. Rev. Lett.* 14, 1051 (1965); *Phys. Rev.* 140, B736 (1965); W. I. Weisberger, *Phys. Rev. Lett.* 14, 1047 (1965).
- [8] S. Weinberg, *Phys. Rev. Lett.* 17, 616 (1966).
- [9] M. L. Goldberger, Y. Miyazawa, and R. Oehme, *Phys. Rev.* 99, 986 (1955).
- [10] S. Weinberg, *Phys. Rev.* 140, B516 (1965).
- [11] S. Weinberg, *Phys. Rev. Lett.* 16, 879 (1966).
- [12] J. Bernstein, S. Fubini, M. Gell-Mann, and W. Thirring, *Nuovo Cimento* 17, 757 (1960); M. Gell-Mann and M. Lévy, *Nuovo Cimento* 16, 705 (1960); K. C. Chou, *Soviet Physics JETP* 12, 492 (1961).
- [13] S. Weinberg, *Phys. Rev. Lett.* 18, 188 (1967).
- [14] J. Schwinger, *Phys. Lett.* 24B, 473 (1967).

- [15] S. Weinberg, Phys. Rev. 166, 1568 (1968).
- [16] S. Coleman, J. Wess, and B. Zumino, Phys. Rev. 177, 2239 (1969); C. G. Callan, S. Coleman, J. Wess, and B. Zumino, Phys. Rev. 177, 2247 (1969).
- [17] J. Wess and B. Zumino, Phys. Lett. 37B, 95 (1971); E. Witten, Nucl. Phys. B223, 422 (1983).
- [18] E. D'Hoker and S. Weinberg, Phys. Rev. D50, R6050 (1994).
- [19] G. 't Hooft, Phys. Rev. D14, 3432 (1976).
- [20] S. Weinberg, "Critical Phenomena for Field Theorists," in *Understanding the Fundamental Constituents of Matter*, ed. A. Zichichi (Plenum Press, New York, 1977).
- [21] H. Schnitzer, Phys. Rev. Lett. 24, 1384 (1970); Phys. Rev. D2, 1621 (1970); L.-F. Li and H. Pagels, Phys. Rev. Lett. 26, 1204 (1971); Phys. Rev. D5, 1509 (1972); P. Langacker and H. Pagels, Phys. Rev. D8, 4595 (1973).
- [22] S. Weinberg, contribution to a festschrift for I. I. Rabi, Trans. N. Y. Acad. Sci. 38, 185 (1977).
- [23] S. Weinberg, Phys. Lett. B251, 288 (1990); Nucl. Phys. B363, 3 (1991); Phys. Lett. B295, 114 (1992).
- [24] C. Ordoñez and U. van Kolck, Phys. Lett. B291, 459 (1992); C. Ordoñez, L. Ray, and U. van Kolck, Phys. Rev. Lett. 72, 1982 (1994); U. van Kolck, Phys. Rev. C49, 2932 (1994); U. van Kolck, J. Friar, and T. Goldman, Phys. Lett. B 371, 169 (1996); C. Ordoñez, L. Ray, and U. van Kolck, Phys. Rev. C 53, 2086 (1996); C. J. Friar, Few-Body Systems Suppl. 99, 1 (1996).
- [25] G. Benfatto and G. Gallavotti, J. Stat. Phys. 59, 541 (1990); Phys. Rev. 42, 9967 (1990); J. Feldman and E. Trubowitz, Helv. Phys. Acta 63, 157 (1990); 64, 213 (1991); 65, 679 (1992); R. Shankar, Physica A177, 530 (1991); Rev. Mod. Phys. 66, 129 (1993); J. Polchinski, in *Recent Developments in Particle Theory, Proceedings of the 1992 TASI*, eds. J. Harvey and J. Polchinski (World Scientific, Singapore, 1993); S. Weinberg, Nucl. Phys. B413, 567 (1994).
- [26] C. Cheung, P. Creminilli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, J. High Energy Physics 0803, 014 (2008); S. Weinberg, Phys. Rev. D **73**, 123541 (2008).
- [27] F. Wilczek and A. Zee, Phys. Rev. Lett. 43, 1571 (1979).
- [28] S. Weinberg, Phys. Rev. Lett. 43, 1566 (1979).
- [29] N. Arkani-Hamed and S. Dimopoulos, JHEP **0506**, 073 (2005); G. F. Giudice and A. Romanino, Nucl. Phys. B **699**, 65 (2004); N. Arkani-Hamed, S. Dimopoulos, G. F. Giudice, and A. Romanino, Nucl. Phys. B **709**, 3 (2005); A. Delgado and G. F. Giudice, Phys. Lett. B627, 155 (2005).
- [30] T. Appelquist and J. Carrazone, Phys. Rev. D11, 2856 (1975).
- [31] S. Weinberg, Phys. Lett. 91B, 51 (1980).
- [32] J. F. Donoghue, Phys. Rev. D50, 3874 (1984); Phys. Lett. 72, 2996 (1994); lectures presented at the Advanced School on Effective Field Theories (Almunecar, Spain, June 1995), gr-qc/9512024; J. F. Donoghue, B. R. Holstein, B. Garbrecht, and T. Konstandin, Phys. Lett. B529, 132 (2002); N. E. J. Bjerrum-Bohr, J. F. Donoghue, and B. R. Holstein, Phys. Rev. D68, 084005 (2003).
- [33] S. Weinberg, in *General Relativity*, ed. S. W. Hawking and W. Israel (Cambridge University Press, 1979): 700; H. Kawai, Y. Kitazawa, & M. Ninomiya, Nucl. Phys. B 404, 684 (1993); Nucl. Phys. B 467, 313 (1996); T. Aida & Y. Kitazawa, Nucl. Phys. B 401, 427 (1997); M. Niedermaier, Nucl. Phys. B 673, 131 (2003).
- [34] L. Smolin, Nucl. Phys. B208, 439 (1982); R. Percacci, Phys. Rev. D 73, 041501 (2006).
- [35] J. Ambjørn, J. Jurkiewicz, & R. Loll, Phys. Rev. Lett. 93, 131301 (2004); Phys. Rev. Lett. 95, 171301 (2005); Phys. Rev. D72, 064014 (2005); Phys. Rev. D78, 063544 (2008); and in *Approaches to Quantum Gravity*, ed. D. Oriti (Cambridge University

Press).

- [36] M. Reuter, Phys. Rev. D 57, 971 (1998); D. Dou & R. Percacci, Class. Quant. Grav. 15, 3449 (1998); W. Souma, Prog. Theor. Phys. 102, 181 (1999); O. Lauscher & M. Reuter, Phys. Rev. D 65, 025013 (2001); Class. Quant. Grav. 19, 483 (2002); M. Reuter & F. Saueressig, Phys Rev. D 65, 065016 (2002); O. Lauscher & M. Reuter, Int. J. Mod. Phys. A 17, 993 (2002); Phys. Rev. D 66, 025026 (2002); M. Reuter and F. Saueressig, Phys Rev. D 66, 125001 (2002); R. Percacci & D. Perini, Phys. Rev. D 67, 081503 (2002); Phys. Rev. D 68, 044018 (2003); D. Perini, Nucl. Phys. Proc. Suppl. C 127, 185 (2004); D. F. Litim, Phys. Rev. Lett. **92**, 201301 (2004); A. Codello & R. Percacci, Phys. Rev. Lett. 97, 221301 (2006); A. Codello, R. Percacci, & C. Rahmede, Int. J. Mod. Phys. A23, 143 (2008); M. Reuter & F. Saueressig, 0708.1317; P. F. Machado and F. Saueressig, Phys. Rev. D77, 124045 (2008); A. Codello, R. Percacci, & C. Rahmede, Ann. Phys. 324, 414 (2009); A. Codello & R. Percacci, 0810.0715; D. F. Litim 0810.3675; H. Gies & M. M. Scherer, 0901.2459; D. Benedetti, P. F. Machado, & F. Saueressig, 0901.2984, 0902.4630; M. Reuter & H. Weyer, 0903.2971.
- [37] F. J. Wegner and A. Houghton, Phys. Rev. A8, 401 (1973).
- [38] J. Polchinski, Nucl. Phys. B231, 269 (1984); C. Wetterich, Phys. Lett. B 301, 90 (1993).
- [39] A. Codello, R. Percacci, & C. Rahmede, Int. J. Mod. Phys. A23, 143 (2008).
- [40] A. Codello, R. Percacci, & C. Rahmede, Ann. Phys. 324, 414 (2009).
- [41] D. Benedetti, P. F. Machado, & F. Saueressig, 0901.2984, 0902.4630.
- [42] A. Bonanno and M. Reuter, Phys. Rev. D 65, 043508 (2002); Phys. Lett. B527, 9 (2002); M. Reuter and F. Saueressig, J. Cosm. and Astropart. Phys. 09, 012 (2005).

This page intentionally left blank

PART B

Topical Sessions

This page intentionally left blank

Dynamical Systems

(including integrable systems
and Hamiltonian stability)

Session organizers:

Igor Krichever
Lai-Sang Young

ALMOST DENSE ORBIT ON ENERGY SURFACE

VADIM KALOSHIN^{1,2}, KE ZHANG¹, and YONG ZHENG¹

1 Department of Mathematics University of Maryland at College Park MD, USA, 20740

2 Department of Mathematics, Penn State University, PA, USA, 16801,

E-mails: kaloshin@math.psu.edu, kezhang@math.umd.edu, yongzheng@math.umd.edu

We study a C^r nearly integrable Hamiltonian system $\mathcal{H}_\varepsilon(q, p) = \frac{1}{2}\langle p, p \rangle + \varepsilon \mathcal{H}_1(q, p)$ defined on $\mathbb{T}^3 \times \mathbb{R}^3$. Let $\Sigma = \{(q, p) : \mathcal{H}_\varepsilon(q, p) = \frac{1}{2}\}$ and μ_{Σ_1} be the restriction of Lebesgue measure on $\mathbb{T}^3 \times \mathbb{R}^3$ to Σ . We prove there is a perturbation $\mathcal{H}_1(q, p) \in C^r$, $\|\mathcal{H}_1\|_{C^r} \leq 1$ and an orbit $(q(t), p(t)) : \mathbb{R} \rightarrow \mathbb{T}^3 \times \mathbb{R}^3$ of the Hamiltonian equation $\{\dot{q} = \partial_p \mathcal{H}_\varepsilon, \dot{p} = -\partial_q \mathcal{H}_\varepsilon\}$ such that $\mu_\Sigma(\bigcup_{t \in \mathbb{R}} (q(t), p(t))) \geq \frac{1}{2}$.

1. Introduction

The famous question called the ergodic hypothesis suggested that for a typical Hamiltonian on a typical energy surface all, but a set of zero measure of initial conditions, have trajectories covering densely this energy surface itself. However, KAM theory showed that for nearly integrable systems there is a set of initial conditions of positive measure of quasi periodic trajectories. This disproved the ergodic hypothesis and forced to reconsider the problem.

A quasi ergodic hypothesis asks if a typical Hamiltonian on a typical energy surface has a dense orbit. A definite answer whether this statement is true or not is still far out of reach of modern dynamics. There was an attempt to prove this statement by E. Fermi [5], which failed (see [6] for more detailed account).

To simplify the quasi ergodic hypothesis, M. Herman [7] formulated the following question: *Can one find an example of a C^∞ Hamiltonian $\mathcal{H}$ in a C^r small neighborhood of $\mathcal{H}_0(p) = \frac{\langle p, p \rangle}{2}$ such that on the unit energy surface $\{\mathcal{H}^{-1}(\frac{1}{2})\}$ there is a dense trajectory?* Many people believe that such examples do exist and are C^∞ -generic (see [1], [3], [4]).

In this paper we make a step in the direction of answering Herman's question. For any r we construct a Hamiltonian, which is C^r close to $\mathcal{H}_0(p) = \frac{\langle p, p \rangle}{2}$ and has a trajectory dense in a set of Lebesgue measure $1/2$ on the energy surface. Here is the exact statement. Let $q \in \mathbb{T}^3$, $p \in \mathbb{R}^3$ and $\mathcal{H}_0(p) = \frac{\langle p, p \rangle}{2}$ be the unperturbed Hamiltonian, where $\langle p, p \rangle$ is the dot product in $\mathbb{R}^3$.

Theorem 1.1. *For any $r \geq 2$ there is a C^r small perturbation $\mathcal{H}_\varepsilon(q, p) = \mathcal{H}_0(p) + \varepsilon \mathcal{H}_1(q, p, \varepsilon)$ and an orbit $(q(t), p(t)) : \mathbb{R} \rightarrow \mathbb{T}^3 \times \mathbb{R}^3$ of*

$$\dot{q} = \partial_p \mathcal{H}_\varepsilon, \quad \dot{p} = -\partial_q \mathcal{H}_\varepsilon \tag{1}$$

such that $\mu_\Sigma(\overline{\bigcup_{t \in \mathbb{R}} (q(t), p(t))}) \geq \frac{1}{2}$.^a

Let $\Sigma = \{(q, p) : \mathcal{H}_\epsilon(q, p) = \frac{1}{2}\}$, we fix a subset $\mathcal{F} \subset \Sigma$ with $\mu_\Sigma(\mathcal{F}) \geq \frac{1}{2}$. It suffices to prove that for any $\delta > 0$ there exists T_δ such that the δ neighborhood of $\bigcup_{t \in [0, T_\delta]} (q(t), p(t))$ contains $\mathcal{F}$.

We will construct $\mathcal{H}_\epsilon$ in two steps. In step one we build $\mathcal{H}'_\epsilon = \mathcal{H}_0 + \epsilon \mathcal{H}'_1$ so that has a variety of good local normal forms and nice invariant sets. Then $\mathcal{H}_\epsilon = \mathcal{H}'_\epsilon + \epsilon \mathcal{H}''_1$ is designed to have diffusing orbits shadowing these invariant sets.

2. Choice of $\mathcal{F}$

We will describe the choice of the positive measure set $\mathcal{F}$, as well as an approximate path of diffusion. We begin with a informal discussion of the diffusion path and what kind of perturbation we need. Usually diffusing orbits travel along resonant segments. To be able to saturate a set of positive measure one has to be able to move along infinitely many resonant segments. If size of a perturbation is fixed, the analysis of motions near resonances of larger and larger orders in the original coordinate system becomes increasingly complicated as explained in section 4. To be able to control dynamics along some arbitrary high order resonant we define a convenient symplectic coordinate system $\Phi : (\theta, I) \rightarrow (q, p)$ on a neighborhood of $\{\|p\| = 1, p_1 \geq \frac{1}{2}\}$, such that $\mathcal{H}'_\epsilon \circ \Phi(\theta, I) = H_0(I) + H_1(\theta, I)$, where $\|H_1(\theta, I)\|_{C^r}$ gets the smaller as the order of a corresponding resonance increases.

We consider the following set of Diophantine numbers:

$$\begin{aligned} \mathcal{D}_\gamma &= \{\omega = (\omega_1, \omega_2, \omega_3); \|\omega\| = 1, |k \cdot \omega| \geq \gamma |\omega| |k|^{-2-\tau}, \forall k \in \mathbb{Z}^3; \\ &|k_1 \omega_1 + k_2 \omega_2| \geq \gamma^{\delta(1+\delta)} |(k_1, k_2)|^{-1-\delta}, \forall (k_1, k_2) \in \mathbb{Z}^2; \\ &|k_1 \omega_1 + k_3 \omega_3| \geq \gamma^{\delta(1+\delta)} |(k_1, k_3)|^{-1-\delta}, \forall (k_1, k_3) \in \mathbb{Z}^2\}, \end{aligned} \quad (2)$$

where $\delta > 0$ is a small number. The set $\mathcal{D}_\gamma$ has positive measure on the surface $\{\|\omega\| = 1\}$. Let $B = \{\|\omega\| = 1; \omega_1 \geq \frac{1}{2}\}$ and we will choose a subset $\mathcal{D}_\gamma^\infty \subset \mathcal{D}_\gamma \cap B$ with positive measure. The family of Diophantine number corresponds to a family of KAM tori which has measure on the energy surface $\{\mathcal{H}_\epsilon = \frac{1}{2}\}$. Denote it $\mathcal{F}$.

The construction will be done in infinitely many stages, each stage we will define a set of paths in the set B , such that if the Hamiltonian H satisfies a list of properties, there exists an orbit such that $\dot{\varphi}$ shadows the chosen path. The path gets denser in each stage and in the limit $\dot{\varphi}$ accumulates to a set of positive measure.

For any integer vector $k \in \mathbb{Z}^3 \setminus \{0\}$, we can relate to it a resonant plane $\{\omega \in \mathbb{R}^3 : k \cdot \omega = 0\}$. If the plane intersects B , the intersection is a curve on the unit sphere, which we will refer to as Γ_k .

At stage 1 the construction consists of the following components:

^aIn [8] there is a construction of $\mathcal{H}_\epsilon$ and an orbit of $\mathcal{H}_\epsilon$ whose closure has maximal Hausdorff dimension

- (1) Let $\gamma_1 = \gamma^4$. We will choose a discrete set $\mathcal{DN}^1 \subset \mathcal{D}_\gamma \cap B$, and disjoint neighborhoods $\mathcal{U}(\omega_i)$ of $\omega_i \in \mathcal{DN}^1$, such that each $\mathcal{U}(\omega_i)$ contains a ball of radius γ_1 , and is contained in a ball of radius $3\gamma_1$, both centered at ω_i .
- (2) Let $\mathcal{D}_\gamma^1 = \mathcal{D}_\gamma \cap \bigcup_{\omega_i \in \mathcal{DN}^1} \mathcal{U}(\omega_i)$, we have the sets $\mathcal{U}(\omega_i)$ is chosen in such that a way that the measure of $\mathcal{D}_\gamma \setminus \mathcal{D}_\gamma^1$ is small.
- (3) There exists a collection $\mathbb{F}^1$ of integer vectors, such that for any $\omega_i \in \mathcal{DN}^1$ there exists some $k \in \mathbb{F}^1$ such that Γ_k enters $\gamma_1/2$ neighborhood of ω_i . Furthermore, the union $\mathcal{F}^1 := \bigcup_{k \in \mathbb{F}^1} \Gamma_k$ is connected.

In stage 2, let $\gamma_2 = \gamma_1^{1+\alpha}$ for some $\alpha > 0$. For each neighborhood $\mathcal{U}(\omega_i)$ of stage 1, we similarly define the following:

- (1) A discrete set $\mathcal{DN}_i^2 \subset \mathcal{D}_\gamma^1 \cap \mathcal{U}(\omega_i)$, and for each $\omega_{ij} \in \mathcal{DN}_i^2$, we have neighborhoods $\mathcal{U}(\omega_{ij})$, whose radius is between γ_2 and $3\gamma_2$.
- (2) $\mathcal{D}_i^2 = \mathcal{D}_\gamma^1 \cap \bigcup_{\omega_{ij} \in \mathcal{DN}_i^2} \mathcal{U}(\omega_{ij})$. The measure of $\mathcal{D}_\gamma^1 \cap \mathcal{U}(\omega_i) \setminus \mathcal{D}_i^2$ is small.
- (3) For the neighborhood $\mathcal{U}(\omega_i)$, there exists $k' \in \mathbb{F}^1$, such that the resonant line $\Gamma_{k'}$ enters the neighborhood. We further define a collection $\mathbb{F}_i^2$ of integer vectors, such that for any $\omega_{ij} \in \mathcal{DN}_i^2$, there exists some $k \in \mathbb{F}_i^2$ such that Γ_k enters $\gamma_2/2$ neighborhood of ω_{ij} . Write $\mathcal{F}_i^2 = \bigcup_{k \in \mathbb{F}_i^2} \Gamma_k$, we assume that $\mathcal{F}_i^2 \cup \Gamma_{k'}$ is connected. Denote also $\mathbf{F}^n = \bigcup_{i=1}^n \mathcal{F}^i$.

We do this for every neighborhood $\mathcal{U}(\omega_i)$ and let $\mathcal{DN}^2 = \bigcup \mathcal{DN}_i^2$, $\mathcal{D}_\gamma^2 = \bigcup \mathcal{D}_i^2$, $\mathbb{F}^2 = \bigcup \mathbb{F}_i^2$, $\mathcal{F}^2 = \bigcup \mathcal{F}_i^2$. We then continue this construction inductively: for each multi-index $(i_1 \cdots i_n)$, assume that we have the neighborhood $\mathcal{U}(\omega_{i_1 \cdots i_n})$, we can define $\mathcal{DN}_{i_1 \cdots i_n}^{n+1}$, $\mathcal{D}_{i_1 \cdots i_n}^{n+1}$ and $\mathbb{F}_{i_1 \cdots i_n}^{n+1}$ in a similar fashion. Union over all multi-indices of same order is denoted by $\mathcal{DN}^{n+1}$, $\mathcal{D}_\gamma^{n+1}$ and $\mathbb{F}^{n+1}$. Then $\mathcal{D}_\gamma^\infty$ is the intersection of $\mathcal{D}_\gamma^n$ and has almost full measure in $\mathcal{D}_\gamma \cap B$. Finally, we have the following

Theorem 2.1. *The Hamiltonian $H(\theta, I) = H_0(I) + H_1(\theta, I)$ has the following property: Consider the resonant lines $\mathcal{F}^n$ of stage n , there exists an open cover U_j of $\mathcal{F}^n$, such that for each U_j , there exists a neighborhood $\mathbf{U}_j \times \mathbb{T}^3 \supset (\partial_I H)^{-1}(U_j)$, on which H is in one of the two normal forms:*

- (1) *Single and ghost^b resonances: There exist local coordinates $\Psi : (\hat{\theta}, \hat{I}) \rightarrow (\theta, I)$ such that*

$$H \circ \Psi(\hat{\theta}, \hat{I}) = \hat{H}_0(\hat{I}) + a_k \cos(\pi k \cdot \hat{\theta}) + R, \quad (3)$$

where $k \in \mathbb{F}^n$ and $\|R\| \ll |a_k|$.

- (2) *Double resonance:*

$$H(\theta, I)|_{\mathbf{U}_j \times \mathbb{T}^3} = H_0(I) + a_k \cos(\pi k \cdot \theta) + a_{k'} \cos(\pi k' \cdot \theta) + R, \quad (4)$$

where $k \in \mathbb{F}^n$, k' is in $\mathbb{F}^{n-1}$, $\mathbb{F}^n$ or $\mathbb{F}^{n+1}$, $\|R\|_{C^3} \ll \max\{|a_k|, |a_{k'}|\}$.

^bThere are certain $k'' \notin \mathbb{F}^{n-1} \cup \mathbb{F}^n \cup \mathbb{F}^{n+1}$ such that $\Gamma_{k''}$ intersects $\Gamma_{k'}$ inside U_j . We call such an intersection a *ghost double resonance*

3. A proof of existence of a δ -dense orbit using a variational problem with constraints

In this section we reformulate a problem of existence of an orbit following a Cantor set of lines as a variational problem with constraints (following Mather). Recall that under the convenient coordinate system we have the Hamiltonian $H(\theta, I) = H_0(I) + H_1(\theta, I)$.

Due to the convexity with respect to I , the Hamiltonian system (1) is equivalent to the dynamics of the E-L equation with Lagrangian L as $L(\theta, \dot{\theta}) = l_0(\dot{\theta}) + L_1(\theta, \dot{\theta})$, which is positive definite with respect to $\dot{\theta}$ for any $\theta \in \mathbb{T}^3$.

Select $\{\omega_n^k\}_{k=1}^{N_n}$ be a set of points in $\mathbf{F}^n$ such that $|\omega_n^k - \omega_n^{k+1}|$ is sufficiently small.

Denote by $\mathcal{A}^\omega$ a special invariant set of orbits (to be defined later) with rotation vector ω . In our case velocity of these orbits will stay close to ω . Our goal is to construct a transition chain from these sets $\{\mathcal{A}^{\omega_n^k}\}_{k=1}^{N_n}$ and an orbit shadowing these sets. Such an orbit will stay close to the union of the stable set $W^s(\mathcal{A}^{\omega_n^k})$ and the unstable set $W^u(\mathcal{A}^{\omega_n^k})$ for all time. We find these orbits by constructing a variational problem with constraints. This construction is fairly involved and relies heavily on Mather's ideas. We describe its construction into several steps.

Let $\theta \in \mathbb{T}^3$, denote $\hat{\theta} \in \mathbb{R}^3$ a lift to $\mathbb{R}^3$. Let η be a closed one form, denote $\hat{\eta}$ a lift of it to a periodic close one form on $\mathbb{R}^3$. Fix a lift. One can proof existence of the following set of objects:

Collections of numbers α_i , periodic functions $B_i^\pm$ and closed one forms η_i on the 3-torus $\mathbb{T}^3$, errors (negligibly small numbers) δ_i , smooth manifolds S_i with a boundary diffeomorphic to a 2-disk inside the 3-torus $\mathbb{T}^3$ such that the following variational problem with constraints has an interior solution.

$$\text{Let} \quad h_i(\theta_i, \theta_{i+1}, T) = \min \int_0^T (L - \eta_c)(\gamma(s), \dot{\gamma}(s)) dt, \quad (5)$$

where the minimum is taken over all absolutely continuous curves $\gamma : [0, T] \rightarrow \mathbb{T}^3$ such that $\gamma(0) = \hat{\theta}_i = \theta_i(\text{mod } 1)$, and $\gamma(T) = \hat{\theta}_{i+1} = \theta_{i+1}(\text{mod } 1)$. Given $T^* \gg 1$ and $\mathbf{T} \gg NT^*$, consider

$$M(\theta_0, \dots, \theta_N) = \min_{\substack{\theta_i \in S_i, \\ T_0=0, T_N=\mathbf{T}}} \sum_{i=0}^N h_i(\theta_i, \theta_{i+1}, T_{i+1} - T_i), \quad (6)$$

and we seek an interior minimum such that $T_{i+1} - T_i > T_*$ and that each θ_i lies in the interior of S_i . In addition, we need some constraints on the homology class of the minimizing orbit, which can be achieved by going to a proper covering of $\mathbb{T}^3$. We clarify this later in the section. It turns out that for each $i = 0, \dots, N$ we have

$$|h_i(\theta_i, \theta_{i+1}, \Delta T_i) - \alpha_i(T_{i+1} - T_i) + B_i^-(\theta_i) + B_i^+(\theta_{i+1})| \leq \delta_i.$$

Thus, to have an interior minimum it suffices to have a sufficiently deep interior minimum of $B_i^-(\theta_i) + B_i^+(\theta_{i+1})$. It also turns out that η_i and η_{i+1} can be chosen so that they coincide near the disk S_{i+1} .

Having this as the motivating goal, we shall define related objects from the Mather theory – see, e.g., [9]. Here is the correspondence:

$B_i^\pm(\theta)$ are one-sided barrier functions, defined by Mather [9]. These functions form a 3-parameter family, naturally parametrized by $c \in H^1(\mathbb{T}^3, \mathbb{R})$. It turns out that cohomology class of the one form η_i is given by $[\eta_i]_{H^1(\mathbb{T}^3, \mathbb{R})} = c_i$. To determine position of $S_i \subset \mathbb{T}^3$ we need to determine a location of certain invariant sets, usually called Aubry sets $\mathcal{A}_{c_i}$ also naturally parametrized by c .

Let $I = [a, b]$ be an interval of time and $c \in H^1(\mathbb{T}^3, \mathbb{R}) = \mathbb{R}^3$. A curve $\gamma \in C^1(I, \mathbb{T}^3)$ is called c -minimizer if

$$A_c(\gamma) := \int_a^b (L - \eta_c)(\gamma(s), \dot{\gamma}(s)) dt = \min_{\substack{\xi(a)=\gamma(a), \xi(b)=\gamma(b) \\ \xi \in C^1(I, \mathbb{T}^3)}} \int_a^b (L - \eta_c)(\gamma(s), \dot{\gamma}(s)) dt,$$

where η_c is a closed 1-form on $\mathbb{T}^3$ such that $[\eta_c] = c$. Let $\mathcal{M}_L$ be the set of Borel probability measures on $\mathbb{T}^3 \times \mathbb{R}^3$, invariant for the E-L flow φ_t^L . For any $\nu \in \mathcal{M}_L$, the action $A_c(\nu)$ is defined as $A_c(\nu) = \int (L - \eta_c) d\nu$. A probability measure μ is called c -minimal invariant measure if $A_c(\mu) = \min_{\nu \in \mathcal{M}_L} A_c(\nu)$. Denote $\mathcal{M}(c)$ the supports of c -minimal invariant measures and call it *Mather set*. A function $\alpha(c) := -A_c(\mu) : H^1(\mathbb{T}^3, \mathbb{R}) \rightarrow \mathbb{R}$ is called α -function, where μ is a c -minimal invariant measure. Define

$$h_c(\theta, \theta'; t) = \min_{\substack{\gamma \in C^1([0, t], \mathbb{T}^3) \\ \gamma(0)=\theta, \gamma(t)=\theta'}} \int_0^t (L - \eta_c + \alpha(c))(\gamma(s), \dot{\gamma}(s)) ds,$$

$$F_c(\theta, \theta') = \inf_{t \geq 0} h_c(\theta, \theta'; t), \quad h_c^\infty(\theta, \theta') = \lim_{t \geq 0} \lim_{t \rightarrow +\infty} h_c(\theta, \theta'; t).$$

Let $\gamma : \mathbb{R} \rightarrow \mathbb{T}^3$ be a C^1 curve

- It is called *c-semi-static* if $A_c(\gamma|_{[a, b]}) + \alpha(c)(b - a) = F_c(\gamma(a), \gamma(b), b - a)$ for any $a < b$.
- It is called *c-static* if it is *c-semi-static* and $A_c(\gamma|_{[a, b]}) + \alpha(c)(b - a) = -F_c(\gamma(b), \gamma(a), b - a)$.

Denote the set of c -semi-static and c -static orbits as $\mathcal{N}(c)$ and $\mathcal{A}(c)$ respectively. Usually $\mathcal{N}(c)$ is called a *Mañé set* and $\mathcal{A}(c)$ is called an *Aubry set*.

At the n -th stage of the induction we construct a collection $\{c_n^k\}_{k=1}^{N_n}$ such that orbits in the corresponding Aubry sets $\mathcal{A}_{c_n^k}$ have velocity $\dot{\theta}$ close to ω_n^k . Then we find orbits following the stable set $W^s(\mathcal{A}_{c_n^k})$ and the unstable set $W^u(\mathcal{A}_{c_n^k})$ as local minimizer of Euler-Lagrange equation. There are two drastically different cases in our problem: *single resonance and double resonance*.

3.1. Single resonance case

In Theorem 2.1 near a single resonance $k \in \mathbb{Z}^3 \setminus \{0\}$ we obtain a normal form (3). In order to construct a variational problem whose solutions diffuses along this

resonance. Associate to $k \in \mathbb{Z}^3 \setminus \{0\}$ an integer linear transformation $A \in SL_3(\mathbb{Z})$ such that A induces a new coordinate system on $\mathbb{T}^3$, denote $\mathbb{T}^3 = \mathbb{T}_f^2 \times \mathbb{T}_s \ni \theta = (\theta_1, \theta_2, \theta_s)$ so that θ_s is parallel to k . After an associated linear transformation we can consider the following Lagrangian system

$$L(\theta, \dot{\theta}) = l_0(\dot{\theta}) + a \cos^2 \frac{\pi \theta_s}{2} + \delta L_1(\theta, \dot{\theta}), \quad (7)$$

where $l_0(\dot{\theta})$ is close to $\langle \dot{\theta}, \dot{\theta} \rangle / 2$, $\theta \in \mathbb{T}^3$ and δ is sufficiently small compare to a . The form of the Lagrangian implies that there is a co-dimensional 2 normally hyperbolic cylinder $\Lambda_k = \{\dot{\theta}_s = \theta_s = 0\}$. Rotation vectors associated to the single resonant are of form $\omega = (\omega_1, \omega_2, 0)$ in this new coordinates system. Restricted to an energy surface the normally hyperbolic is a 3 dimensional invariant manifold which is diffeomorphic to $T\mathbb{T} \times \mathbb{T}$. View the second $\mathbb{T}$ component as time, the dynamics of the Poincaré return map on the invariant cylinder is an exact area-preserving twist map. For exact area-preserving twist maps structure of Mather and Aubry sets is well understood (see e.g. [10]) For example, minimal invariant measures have rotation number ω_1/ω_2 . If ω_1/ω_2 is irrational, then there is unique $c' = (c_1, c_2)$ corresponds to (ω_1, ω_2) . After we add the hyperbolic part into the dynamics, then there is an open interval $I_\omega \subset \mathbb{R}$ such that $\mathcal{A}(c) = \mathcal{N}(c)$ for any $c = \{(c_1, c_2, c_3) : c_3 \in I_\omega\}$. Moreover, $\mathcal{A}(c)$ is on the invariant cylinder. If ω_1/ω_2 is rational, the situation is a little bit complicated, because there is an open set of $c = (c_1, c_2, c_3)$'s with the same $\mathcal{A}(c)$. It is still true that $\mathcal{A}(c)$ belongs to the invariant cylinder Λ_k .

According to Bernard's theorem [2] Aubry and Mather sets are invariant under symplectic transformation. Once we establish a structure of Aubry-Mather sets in the normal form (3) we can construct a variational problem in the original coordinate system.

To describe the variational problem for the original coordinate system, we consider the covering space $N_k = \mathbb{T}_f^2 \times \mathbb{R}$ of $\mathbb{T}_f^2 \times \mathbb{T}_s$ by unfolding the θ_s direction. Denote by $\pi_k : N_k \rightarrow \mathbb{T}_f^2 \times \mathbb{T}_s$ the natural projection.

Now we construct a variational problem to diffuse along Γ_k . Consider a sufficiently dense set of c 's whose corresponding Aubry sets are on Λ_k , denoted $\{c_j\}_{0 \leq j \leq N}$. We will construct relative open sets $S_j \subset \{\theta_{ss} = j + \frac{1}{2}\}$, a collection of closed one forms η_i on $\mathbb{T}^3$ such that $[\eta_i]_{H^1(\mathbb{T}^3, \mathbb{R})} = c_i$ and η_i coincides with η_{i+1} near $\pi_k(S_i)$. We would like to show that for this choice of S_j and η_j there is $T^* \gg 1$ and $\mathbf{T} \gg NT$ the variational problem as in the notations (6) attains an interior minimum. This can be done, if we add an additional perturbation to $L(\theta, \dot{\theta})$ required, in particular, to use [11].

3.2. Double resonance

In Theorem 2.1 near double resonances we obtain a normal form (4). We would like to diffuse first along Γ_k , come to the intersection with $\Gamma_{k'}$, and then diffuse along $\Gamma_{k'}$. In order to construct a variational problem, whose solutions diffuse along these

resonances we distinguish three regimes: diffusing along Γ_k , switching from Γ_k to $\Gamma_{k'}$, and diffusing along $\Gamma_{k'}$.

Associate to $k, k' \in \mathbb{Z}^3 \setminus \{0\}$ an integer linear transformation $A \in SL_3(\mathbb{Z})$ such that A induces a new coordinate system on $\mathbb{T}^3$, denote $\mathbb{T}_A^3 = \mathbb{T}_f \times \mathbb{T}_s \times \mathbb{T}_{ss} \ni \theta = (\theta_f, \theta_s, \theta_{ss})$ so that θ_s is parallel to k' and θ_{ss} is parallel to k . After such a transformation we have the following Lagrangian system

$$L(\theta, \dot{\theta}) = l_0(\dot{\theta}) + a_1 \cos^2\left(\frac{\pi}{2}\theta_s\right) + a_2 \cos^2\left(\frac{\pi}{2}\theta_{ss}\right) + \delta L_1(\theta, \dot{\theta}), \quad (8)$$

$a_1, a_2 > 0$ and sufficiently small and $\delta = \min\{a_1^{100}, a_2^{100}\}$. Denote by $L^0(\theta, \dot{\theta}) = L - \delta L_1$.

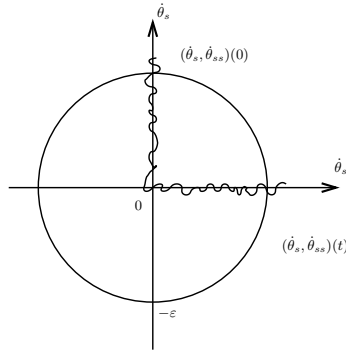


Fig. 1. Velocity diffusing across a double resonance

Let $K = a_1^{2\tau}$. Suppose $\sqrt{a_1}l$ is $(K, 1)$ -Diophantine, i.e. $|p - q\sqrt{a_1}l| > K/|q|^2$ for any $q, p \in \mathbb{Z}$, $q \neq 0$. Consider c such that the corresponding Aubry set $\mathcal{A}_c$ for L^0 has rotation vector $\omega = (1, \sqrt{a_1}l, 0)$ satisfying the above Diophantine condition. Consider a sufficiently dense set of c 's with this property, denoted $\mathbb{R}_k = \{c_j\}_{0 \leq j \leq N}$.

To diffuse along Γ_k , similar to the single resonance case, we can define the manifold N_k , S_j and η_j , and our goal is to prove existence of the interior minimum for the sum as in (6).

One can show that for each c_i 's above the corresponding $\mathcal{A}_{c_i}$ for the Lagrangian system on $\mathbb{T}^3$ can be lifted to a countable collection $\{\mathcal{A}_{c_j}^j\}_{j \in \mathbb{Z}}$ so that projection on to ss -component belongs to $[j - \frac{1}{2}, j + \frac{1}{2}]$. Define $B_{j,c}^+(\theta) = \inf_{\theta' \in \mathcal{A}_c^j} h_c^\infty(\theta, \theta')$ and $B_{j,c}^-(\theta) = \inf_{\theta' \in \mathcal{A}_c^j} h_c^\infty(\theta, \theta')$. Notice that θ and θ' belong to the lift N_k . We show that for $T > T_*$ there is α_j such that we have

$$h_j(\theta, \theta'; T) = \alpha_j T + B_{j,c_j}^+(\theta) + B_{j+1,c_j}(\theta') + \delta_j,$$

where δ_j is sufficiently small. For c satisfying this condition we prove that

Theorem 3.1.

$$\max_{\theta_{ss}=1} B_c^\pm(\theta) - \min_{\theta_{ss}=1} B_c^\pm(\theta) = O(\delta^a \epsilon^{-b}).$$

We add a localized potential perturbation close to $(\theta_f, \theta_s, \theta_{ss}) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ such that the Barrier function $B_{j,c}^{\pm}$ has an isolated local minimum. The diffusion for $N_{k'}$ is similar and we can glue the diffusion orbits together by using a common covering of N_k and $N_{k'}$.

4. Competition between order of resonance and distance to a KAM torus

In this section we show why we need a careful selection of symplectic coordinates near resonant segments. To illustrate the problem consider dynamics of $\mathcal{H}_\varepsilon(q, p) = \mathcal{H}_0(p) + \varepsilon \mathcal{H}_1(q, p)$ near a double resonance given by two resonant segments $\Gamma_k \cap \Gamma_{k'}$. If $k'' := k \times k'$ is sufficiently large, then typically orbits of the unperturbed system at the double resonance are periodic of length $\sim |k''|$. If $|k''| \cdot \varepsilon$ is not small, then standard averaging does not apply.

On the other side, consider a Diophantine number $\omega \in \mathcal{D}_\gamma$. Then for small ε the Hamiltonian $\mathcal{H}_\varepsilon$ has a KAM torus $\mathbb{T}_\omega$. In a certain neighborhood of $\mathbb{T}_\omega$ one can choose a Birkhoff normal form of some order m : $\mathcal{H}_\varepsilon \circ \Phi_\omega(\theta, I) = H^\omega(I) + H_1^\omega(\theta, I)$. Notice that in a ρ -neighborhood of $\mathbb{T}_\omega$ with small ρ perturbation $\|H_1^\omega\|_{C^r}$ is bounded by ρ^m . Notice now that if a double resonance $\Gamma_k \cap \Gamma_{k'}$ belongs to this neighborhood and $|k''|^3 \times \rho^m$ is small then averaging does apply and there is a hope to control dynamics. Selection of resonant segments in (2) is so that on one side resonant segments stay close enough to Diophantine numbers and on the other they fill a set of almost maximal measure.

Acknowledgement. The first author thanks John Mather for numerous useful conversations and many invaluable advises. The first author was partially supported by NSF grants, DMS-0701271.

References

- [1] V. Arnold, *A stability problem and ergodic properties of classical dynamical systems.* (Russian) 1968, Proc of ICM, (Moscow, 1966) 387–392;
- [2] P. Bernard, *Symplectic aspects of Mather theory*, Duke Math. J. 136 (2007), no. 3, 401–420.
- [3] G. D. Birkhoff, *Collected Math Papers*, vol. 2, p. 462–465.
- [4] P. & T. Ehrenfest, *The Conceptual Foundations of the Stat Approach in Mechanics.* Cornell Univ Press, 1959.
- [5] E. Fermi, *Dimostrazione che in generale un sistema meccanico quasi ergodico*, Nuovo Cimento, 25, 267–269, 1923;
- [6] G. Gallavotti, *Fermi and Ergodic problem*, preprint 2001.
- [7] M. Herman, *Some open problems in dynamics*, Proc of ICM, Vol. II (Berlin, 1998), 797–808.
- [8] V. Kaloshin, M. Levi, M. Saprykina, *An example of a nearly integrable Hamiltonian system with a trajectory dense in a set of maximal Hausdorff dimension*, preprint, 2009, 31pp.
- [9] J. Mather, *Variational construction of connecting orbits*, Ann. Inst. Fourier (Grenoble) 43 (1993), no. 5, 1349–1386.

- [10] J. Mather, *Order structure on action minimizing orbits*, preprint, 2009, 90pp.
- [11] J. Mather, *Modulus of continuity for Peierls's barrier*, Periodic solutions of Hamiltonian systems and related topics(II Ciocco, 1986), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci 209 Reidel, Dordrecht (1987) pp.177-202.

DISSIPATIVE PERTURBATIONS OF KdV

SERGEI B. KUKSIN

Centre de Mathématiques Laurent Schwartz, Ecole polytechnique

91128 Palaiseau cedex France

E-mail: kuksin@math.polytechnique.fr

We consider KdV equation under periodic boundary conditions, perturbed by viscosity of size ν and a random force of size $\sqrt{\nu}$. For $\nu \rightarrow 0$ we derive averaged equations which describe behaviour of solutions when $0 \leq t \lesssim \nu^{-1}$ and when $t \rightarrow \infty$.

Keywords: KdV equation, integrals of motion, dissipative perturbation, random perturbation, averaging, stationary measure.

1. KdV equation

Consider KdV equation under periodic boundary conditions with zero mean value:

$$\dot{u} + u_{xxx} - uu_x = 0, \quad x \in \mathbb{T}^1 = \mathbb{R}/2\pi, \quad \int u \, dx \equiv 0. \quad (KdV)$$

It may be written in the Hamiltonian form $\dot{u} = \frac{\partial}{\partial x} \frac{\delta}{\delta u(x)} H_{KdV}$, where $H_{KdV} = \int (\frac{1}{2} u_x^2 + \frac{1}{6} u^3) dx$. So this is a Hamiltonian PDE. It is integrable. This statement has various meanings. In particular, this one: (KdV) defines a dynamical system in the L^2 function space $Z = \{u(x) \in L^2(\mathbb{T}^1) \mid \int u \, dx = 0\}$, given the L_2 -norm $\|\cdot\|$. Due to Novikov-Lax and McKean-Trubowitz, it has infinitely many non-negative analytic integrals of motion $I_1(u), I_2(u), \dots$ which can be chosen in such a way that

1) For a vector $I = (I_1, I_2, \dots)$, $I_j \geq 0$, denote $T_I = \{u \in Z \mid I_j(u) = I_j \ \forall j\}$. Then T_I is an analytic torus in Z of dimension $n := \dim T_I = \#\{j \mid I_j > 0\}$, $n \leq \infty$. It is invariant for the KdV equation.

2) Each T_I carries a cyclic coordinate $q \in \mathbb{T}^n$ ($n \leq \infty$) such that the variables (I, q) are the action-angles for KdV in Z . That is, $\omega_2 = dI \wedge dq$, $H_{KdV} = H_{KdV}(I)$ and in (I, q) -variables KdV takes the integrable form

$$\dot{I} = 0, \quad \dot{q} = W(I) = \nabla_I H_{KdV}. \quad (1)$$

These coordinates are singular when some $I_j = 0$, but the singularities may be removed in the following sense: due to T. Kappeler, for any $j \geq 1$ there exist analytic functions $v_j^+(u)$ and $v_j^-(u)$ such that i) $I_j = \frac{1}{2}(v_j^{+2} + v_j^{-2})$ and $q_j = \text{Arg}(v_j^+ + i v_j^-)$; ii) the mapping $u(\cdot) \mapsto v = (v_1^+, v_1^-, v_2^+, v_2^-, \dots)$ is a global analytic symplectomorphism. The coordinates $v(u)$ are called the *Birkhoff coordinates* for KdV – see, e.g., [2].

KdV never appears in Physics in the pure form, but always is perturbed by extra terms. So it is important to understand what happens to the invariant tori T_I and integrable solutions $I(t) = \text{const}$, $q(t) = q_0 + t W(t)$, under perturbations of the equation. Consider first quasilinear Hamiltonian perturbations of KdV:

$$\dot{u} + u_{xxx} - uu_x = \varepsilon \frac{\partial}{\partial x} f(u, x). \quad (2)$$

KAM-theory applies and implies that (2) has infinitely many time-quasiperiodic solutions. When $\varepsilon \rightarrow 0$, they become asymptotically dense in Z , see [2, 4]. But nothing is known about long time behaviour of a solution for (2) with a given (generic) initial data! To be able to control long time behaviour of all solutions we consider perturbations which contain dissipation.

2. Random dissipative perturbations: averaging

Consider randomly perturbed and damped KdV:

$$\dot{u} + u_{xxx} - uu_x = \varepsilon u_{xx} + \sqrt{\varepsilon} \eta(t, x), \quad u = u^{\varepsilon, \omega}. \quad (3)$$

Here η is a random force, smooth in x and white in t . That is, $\eta = \eta^\omega(t, x) = \frac{\partial}{\partial t} \sum_j b_j \beta_j^\omega(t) e_j(x)$, where ω is a random parameter, $\omega \in (\Omega, \mathcal{F}, \mathbf{P})$, and

- $\{e_s(x), s \in \mathbb{Z} \setminus \{0\}\}$ is the standard trigonometric basis:
 $e_s(x) = \cos sx, s > 0, e_s(x) = \sin sx, s < 0$;
- all $b_j \neq 0$ and for any $N > 0$ there is C_N such that $|b_j| \leq C_N j^{-N} \forall j$;
- $\{\beta_j(t) = \beta_j^\omega(t)\}$ are independent standard Wiener processes.

Remarks. 1) The scaling factor $\sqrt{\varepsilon}$ in front of the noise η is natural since this is the only scaling when a solution $u^{\varepsilon, \omega}(t)$ stays ~ 1 under the limit $\lim_{\varepsilon \rightarrow 0} \lim_{t \rightarrow \infty}$.

2) Our methods apply to study dissipative random perturbations of other integrable PDE. For example, of the Zakharov-Shabat equation. This group of equations describes nonlinear optical systems and turbulence in them.

3) Eq. (3) may be (cautiously) regarded as a model for the 2D NSE with random force, see [5]. Also see [6] for some results on the 2D NSE, related to those, obtained below for (3).

Write $u = u^{\varepsilon, \omega}$ in the action-angles: $u = (I^\varepsilon, q^\varepsilon) = (I, q)$, and write the corresponding equations for I and q :

$$\dot{I} = \varepsilon F(I, q) + \sqrt{\varepsilon} \sigma(I, q) \frac{\partial}{\partial t} \beta(t), \quad \dot{q} = W(I) + \dots$$

Here $\beta = (\beta_1, \beta_2, \dots)^t$, $\sigma(I, q)$ is an infinite matrix and $0 \leq t \leq T\varepsilon^{-1}$. Let us pass to the fast time $\tau = \varepsilon t$. Then

$$\frac{\partial}{\partial \tau} I = F(I, q) + \sigma(I, q) \frac{\partial}{\partial \tau} \beta(\tau), \quad \frac{\partial}{\partial \tau} q = \varepsilon^{-1} W(I) + \dots, \quad 0 \leq \tau \leq T. \quad (4)$$

We are mostly concerned with the I -equation. Let us average it:

$$\frac{\partial}{\partial \tau} I = \langle F \rangle(I) + \langle \langle \sigma \rangle \rangle(I) \frac{\partial}{\partial \tau} \beta(\tau), \quad 0 \leq \tau \leq T. \quad (5)$$

Here

$$\langle F \rangle(I) = \int_{T_I} F(I, q) dq, \quad \langle \langle \sigma \rangle \rangle(I) = \left(\int_{T_I} \sigma(I, q) \sigma^t(I, q) dq \right)^{1/2}$$

(these are the usual rules of finite-dimensional stochastic averaging, e.g. see in [3]).

Initial conditions for the perturbed equation (3) are either Cauchy:

$$I^\varepsilon(0) = I^0, \quad q^\varepsilon(0) = q^0, \quad (C)$$

or ‘stationary initial conditions’: $(I^\varepsilon(\tau), q^\varepsilon(\tau))$ is a stationary solution for (4). That is,

$$\mathcal{D}(I^\varepsilon(\tau), q^\varepsilon(\tau)) \text{ is a } \tau\text{-independent measure.} \quad (St)$$

It is known that the problems (4), (C) and (4), (St) have unique solutions.

Problem (averaging principle): Is $I^\varepsilon(\tau)$ close to a solution of the averaged equation (5) with the corresponding initial condition?

- For *deterministic* finite-dimensional systems ($\eta = 0$), the averaging principle was suggested 200 years ago by Laplace and Lagrange to study numerically problems from Celestial Mechanics. It was rigorously justified in 1960’s - 1970’s by Anosov, Kosuga and Neishtadt. See [1].

- For *stochastic* finite-dimensional systems it was suggested in 1930’s by N. Bogolyubov. First rigorous results are due to R. Khasminskii (late 1960’s). See in [3].

3. Existence of a limit as $\varepsilon \rightarrow 0$

Theorem 1 (cf. [7]). Every sequence $\varepsilon'_j \rightarrow 0$ has a subsequence $\varepsilon_j \rightarrow 0$ such that $I^{\varepsilon_j}(\tau)_{0 \leq \tau \leq T} \rightarrow I^0(\tau)$ in distribution. We have:

- $I^0(\tau)$ satisfies eq. (5) with the same initial conditions as for $\varepsilon > 0$.
- In the case of stationary solutions we have $\mathcal{D}(I^0(\tau)) \equiv \mu_0$, where μ_0 is a stationary measure for (5). Moreover, $\mathcal{D}(I^{\varepsilon_j}(\tau), q^{\varepsilon_j}(\tau)) \rightarrow \mu_0 \times dq$ as $\varepsilon_j \rightarrow 0$, for any τ . Here dq is the Haar measure on $\mathbb{T}^\infty$.

In particular, Cauchy problem for the averaged equation (5) has a solution, and equation (5) has a stationary solution.

4. The uniqueness

See [8]. Does the limit in Theorem 1 depends on the sequence $\varepsilon_j \rightarrow 0$? What are its properties? Let us start with the Cauchy initial conditions. So $I^\varepsilon(0) = I^0$, $q^\varepsilon(0) = q^0$. By Theorem 1 every limiting process $I^0(\tau) = \lim_{\varepsilon_j \rightarrow 0} I^{\varepsilon_j}(\tau)$ solves the Cauchy problem for eq. (5). If we knew that this problem has a unique solution $I^0(\tau)$, then we would have that $I^0(\tau) = \lim_{\varepsilon \rightarrow 0} I^\varepsilon(\tau)$. But the uniqueness is unknown since (5) is a bad equation in the bad phase-space $\mathbb{R}_+^\infty$. It degenerates on the boundary, when $I_j = 0$ for some j . To work with the averaged equation (5) we unfold the singularities at $I_j = 0$, $j = 1, 2, \dots$. Namely, consider again the Birkhoff coordinates v . Denote $\pi : v \mapsto I$, $I_j = \frac{1}{2}(v_j^{+2} + v_j^{-2})$.

Theorem 2 (Lifting Theorem). There exists a stochastic equation

$$\frac{\partial v}{\partial \tau} = F^L(v(\cdot)) + G^L(v(\cdot)) \frac{\partial}{\partial \tau} \beta(\tau) \quad (6)$$

such that:

i) if $v(\tau)$ solves (6), then the vector $I(\tau) = (I_1(\tau), I_2(\tau), \dots) = \pi(v(\tau))$ solves the averaged equation (5). Other way round, every solution $I(\tau)$ of (5) can be lifted to a solution $v(\tau)$ of (6) such that $v_j^-(0) = 0$ for each j .

ii) In (6), $F^L(v(\cdot)) = Av + F_1^L(v(\cdot))$, where A is the Fourier-transform of the Laplace operator,

$$A(v_1^+, v_1^-, v_2^+, v_2^-, \dots) = -(1^2 v_1^+, 1^2 v_1^-, 2^2 v_2^+, 2^2 v_2^-, \dots),$$

while F_1^L is analytic in $v(\cdot)$, $\nabla v(\cdot)$, and G^L is analytic in $v(\cdot)$.

That is, the lifted averaged equation (6) is a Lipschitz perturbation of the heat equation on S^1 , written in Fourier coefficients. Solution of Cauchy problem for eq. (6) is unique. Hence, solution for the Cauchy problem for the averaged equation for I (eq. (5)) also is unique. So $I^\varepsilon(\cdot) \rightarrow I^0(\cdot)$ as $\varepsilon \rightarrow 0$ in distribution.

Now I summarize the results. For a sequence $v = (v_1^+, v_1^-, v_2^+, v_2^-, \dots)$ denote by $\hat{v}(x)$ its Fourier transform $\hat{v} = \mathcal{F}(v)$, $\hat{v}(x) = \sum_{j>0} v_j^+ \cos jx + \sum_{j<0} v_j^- \sin jx$. Write the original perturbed KdV equation, using the fast time $\tau = \varepsilon t$:

$$u'_\tau + \varepsilon^{-1} u_{xxx} - \varepsilon^{-1} uu_x = u_{xx} + \eta(t, x), \quad u(0, x) = u_0(x). \quad (7)$$

Theorem 3. Let $u^\varepsilon(\tau, x)$, $0 \leq \tau \leq T$, be a solution of (7). Consider the vector $I^\varepsilon(\tau) = I(u^\varepsilon(\tau, \cdot))$. Then $I^\varepsilon(\cdot) \rightarrow I^0(\cdot)$ in distribution. Here $I_j^0(\tau) = \frac{1}{2} (v_j^{+2}(\tau) + v_j^{-2}(\tau))$, $j \geq 1$, and the random function $\hat{v}(\tau, x) = \mathcal{F}(v(\tau))$ satisfies

$$\hat{v}'_\tau = \Delta \hat{v} + F'_1(\hat{v}(\cdot)) + G'(\hat{v}(\cdot)) \frac{\partial}{\partial \tau} \beta(\tau), \quad 0 \leq \tau \leq T. \quad (8)$$

The drift $F'_1(\hat{v}(\cdot))$ is a non-local analytic function of $\hat{v}$ and $\nabla \hat{v}$, while the diffusion $G'(\hat{v}(\cdot))$ is a non-local analytic function of $\hat{v}$.

5. The double limit

(S.K, work in progress.) Existing techniques allow to prove that the lifted averaged equation (8) has a unique stationary measure μ . It is supported by smooth functions $\hat{v}(x)$. By the Lifting Theorem eq. (5) has a unique stationary measure $\nu = \hat{\pi}_* \mu$, where $\hat{\pi} : \hat{v} \mapsto I(\mathcal{F}^{-1}(\hat{v}))$.

Consider the original equation (3). Let μ^ε be its unique stationary measure. Then by Theorem 1 $I \circ \mu^\varepsilon \rightarrow \nu$ as $\varepsilon \rightarrow 0$. On the other hand, for any solution $u^\varepsilon(t) = u^\varepsilon(t, x)$ of (3) we have $\lim_{t \rightarrow \infty} \mathcal{D}u^\varepsilon(t) = \mu^\varepsilon$. Accordingly:

Theorem 4 (the Double Limit). For any smooth function $u_0(x)$ let $u^\varepsilon(t, x)$ be a solution of (3), equal u_0 at $t = 0$. Then $\lim_{\varepsilon \rightarrow 0} \lim_{t \rightarrow \infty} \mathcal{D}(I(u^\varepsilon(t))) = \nu = \hat{\pi}_* \mu$.

6. Summary

Consider the randomly perturbed and damped KdV equation (3). For $0 < \varepsilon \ll 1$ its solutions may be well approximated in terms of solutions for the lifted averaged equation (8). That is, for $t = \varepsilon^{-1}\tau$, $\tau \lesssim 1$, we have $\mathcal{D}(I(u(t))) \rightharpoonup \mathcal{D}(I(\hat{v}(\tau)))$, while for $t \gg \varepsilon^{-1}$ we have $\mathcal{D}(I(u(t)), q(u(t))) \rightharpoonup \nu \times dq$ as $\varepsilon \rightarrow 0$, where $\nu = \hat{\pi}_* \mu$ and μ is the unique stationary measure for equation (8).

References

- [1] V. Arnold, V. V. Kozlov, A. I. Neistadt, *Mathematical Aspects of Classical and Celestial Mechanics*, 1989, Springer Verlag.
- [2] T. Kappeler, J. Pöschel, *KAM & KdV*, 2003, Springer Verlag.
- [3] R. Khasminski, *Stochastic Stability of Differential Equations*, 1980, Osijhoff and Noordhoff.
- [4] S. B. Kuksin, *Analysis of Hamiltonian PDEs*, 2003, Oxford University Press.
- [5] S. B. Kuksin, Proc. Steklov Inst. Math. **259** (2007), 128 - 136.
- [6] S. B. Kuksin, Comm. Math. Phys., **284** (2008), 407 - 424.
- [7] S. B. Kuksin, A. L. Piatnitski, J. Math. Pures Appl. **89** (2008), 400 - 428.
- [8] S. B. Kuksin, A. L. Piatnitski, Preprint (2009).

BILLIARD MODELS AND ENERGY TRANSFER

ZSOLT PAJOR-GYULIAI, DOMOKOS SZÁSZ, and IMRE P. TÓTH

Department of Stochastics, Budapest University of Technology

Budapest, H-1111, Hungary

E-mails: pgyzs@math.bme.hu, szasz@math.bme.hu, mogy@math.bme.hu

www.math.bme.hu

Dedicated to the memory of R. L. Dobrushin on the occasion of his 80'th anniversary

For two (or more) interacting classical particles the existing few results (for diffusion or energy transfer, for instance) assume that the mass of one of them - as compared to the other mass - becomes negligible in the limit (cf. [ChD 09]). Here two models are presented for energy transfer in systems with two or more identical hard disks. The first one, a stochastic paradigm for two Lorentz disks, suggests that the joint diffusive limit of two disks is the mixture of independent pairs of Wiener processes (cf. [P-GySz 09]). In the second one it is shown that in a quasi-1D mechanical chain of localized hard disks - in the scaling of [GG 08] - the limit for the energies of the disks is a n. n. interacting Markov process (cf. [SzT 09]). This latter result should open the way toward a rigorous derivation of Fourier law of heat conduction for a deterministic particle system.

Keywords: Billiard models, Lorentz disk, energy transfer, joint diffusion, derivation of master equation.

1. Introduction

In the last decade the theory of hyperbolic billiards got enriched with highly efficient methods, for instance by LS Young (towers, [Y 98]), by N Chernov and D Dolgopyat (standard pairs, growth properties, averaging, Young coupling á la Chernov-Dolgopyat, [ChD 09]), by P Bálint-IP Tóth (multidim towers, [BT 08]), by I Melbourne, M Nicol, A Torok (probability theory of Young towers, citeMN09, MT04), by D Szász, T Varjú (local limit laws, [SzV 04]), etc. An superb reference is the survey at IMC06, [ChD 07] on what had happened until and was happening around 2006. Consequently, it became possible to achieve essential steps toward the main goal of statistical physics: the derivation of macroscopic behavior from microscopic laws, for example, in the study of diffusion or Fourier's law of heat conduction. For two (or more) interacting particles, however, so far there has been no results except when the mass of one of them - as compared to the other mass - becomes negligible in the limit (cf. [ChD 09]). Here we present results in both of the

aforementioned questions of physics which treat or are motivated by the dynamics of Lorentz disks of equal masses: 2D models of elastic disks moving among periodic scatterers. In both of them energy transfer is at the heart of the problem.

2. Diffusive limit of two Lorentz disks

For simplicity consider two moving disks of radii ρ among a periodic configuration of disjoint circular scatterers. Assume the scatterer configuration has finite horizon and ρ is so small that the disks can move unboundedly. For understanding the joint limit of the two disks in the diffusive scaling, we present a stochastic model.

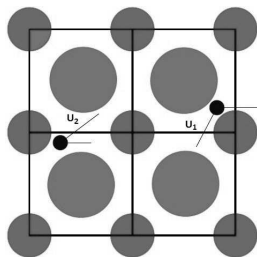


Fig. 1. Four cells of the infinite scatterer configuration (red). Dark circles – moving disks. Grey ones – scatterers.

Definition 2.1. (Continuous time random walk with internal states with general state space.) Assume we are given a rate $\lambda > 0$ and a family $\{P_x(u, \cdot) | x \in \mathbb{Z}^d \setminus \{0\}\}$ of substochastic kernels over $\tilde{H}$ such that $Q = \sum_{x \in \mathbb{Z}^d \setminus \{0\}} P_x$ is a stochastic kernel over $\tilde{H}$. A continuous time pure jump Markov process $\{\xi_t = (x_t, u_t)\}$ – where $(x_t, u_t) \in \mathbb{Z}^d \times \tilde{H}$ – is called a (generalized) Random Walk with Internal States (RWwIS) if

$$\mathbb{P}(\xi_{t+dt} = \xi_t) = 1 - \lambda dt + o(dt)$$

and for every $(x_t, u) \in \mathbb{Z}^d \times \tilde{H}$ and $\forall u \in \tilde{H}, \forall A \subset \tilde{H}$ $x_{t+dt} - x_t \neq 0$

$$\mathbb{P}(\xi_{t+dt} = (x_{t+dt}, u'), u' \in A | \xi_t = (x_t, u)) = \lambda P_{x_{t+dt} - x_t}(u, A) dt + o(dt).$$

In our stochastic model the motions of the two particles are independent RWwIS's unless the two particles occupy the same lattice point. Let $\xi_t^i = (\eta_t^i, \varepsilon_t^i)$, $i = 1, 2$ be two RWwIS with the same kernels but different rates λ_1, λ_2 which only change their values in the moments of collisions. Whenever $\eta_t^1 \neq \eta_t^2$, the joint generator of the two Markov processes is the product of the two individual generators (modeling two independent Lorentz processes).

Next we define the collision interaction. Whenever $\eta_t^1 = \eta_t^2 (= x)$,

$$\begin{aligned} \mathbb{P}(\xi_{t+}^1 = (x + z^1, v_+^1), \xi_{t+}^2 = (x + z^2, v_+^2); v_+^1 \in A^1, v_+^2 \in A^2 \\ |\xi_{t-}^1 = (x, v_-^1), \xi_{t-}^2 = (x, v_-^2)) \\ = C_{z^1, z^2}(v_-^1, v_-^2, A^1, A^2) \end{aligned}$$

is the collision kernel. We assume that C satisfies conservation of energy: $(v_-^1)^2 + (v_-^2)^2 = (v_+^1)^2 + (v_+^2)^2$ (momentum is not conserved since the collision kernel contains averaging over normal of impact). We can and do assume that $(v^1)^2 + (v^2)^2 = 1$. Therefore the state space of the two particle process is isomorphic to $(\mathbb{Z}^2 \times S)^2 \times \tilde{\mathcal{I}}$ where $\tilde{\mathcal{I}} = [0, 1]$. It is worth noting that the concrete form of the collision kernel for the mechanical two disk model is calculated in Appendix A of [GG 08].

Theorem 2.1. (*Joint with Zs. Pajor-Gyulai, [P-GySz 09]*)

For every initial distribution of (ξ_0^1, ξ_0^2) , the density function of the weak limit law of $(\frac{1}{\sqrt{t}}\eta_{t,\lambda_0}^1, \varepsilon_t^1, \frac{1}{\sqrt{t}}\eta_{t,\lambda_0}^2, \varepsilon_t^2)$ exists and is equal to

$$h(x_1, v_1, x_2, v_2) = \frac{\rho(v_1)\rho(v_2)}{(2\pi)^2|\sigma|} \int_0^1 \frac{1}{\lambda\sqrt{1-\lambda^2}} e^{-\frac{1}{2}\left(\frac{x_1^T \sigma^{-1} x_1}{\lambda} + \frac{x_2^T \sigma^{-1} x_2}{\sqrt{1-\lambda^2}}\right)} d\rho_s(\lambda)$$

where $\rho_s(\lambda)$ is the outgoing stationary distribution of the speed of the first particle in the collision Markov chain, ρ is the stationary density of the internal states on S of the component RWwIS's.

We note that in earlier models of joint motion of two identical particles the diffusive limits of the two motions were either independent or got glued together (cf. [Sz 80], [KV 86]).

3. Interacting Markov chain of energies from deterministic dynamics

[GG 08] considered a quasi-1D version of the localized hard ball system of [BLPS 92]. It is a chain of 2D disks each performing a billiard dynamics in its cell and *interacting with the neighboring disks rarely – under their special choice of parameters*. Concretely the parameters are: box size: l (with periodic b. c.'s along y -axis); chain length = N (along the x -axis, with free or periodic boundary conditions); radius of scatterers (shaded circles) = ρ_f ; radius of moving disks (empty circles) = ρ_m ; $\rho_f + \rho_m = \rho$ is being kept fixed; condition of localization: $\rho > l/2$ (and, of course, $\rho < l/\sqrt{2}$); condition of conductivity: $\rho_m > \rho_{crit} = \sqrt{(\rho^2 - (l/2)^2)}$; finally the small parameter in the model is small $\varepsilon = \rho_m - \rho_{crit}$. Under these conditions they, in the limit $\varepsilon \rightarrow 0$, derive a master equation for the time evolution of the energies of the disks from the kinetic equation of the mechanical motion. (Moreover, they also treat the master equation for obtaining the coefficient of heat conductivity: $\kappa = \sqrt{T}$ (T being the temperature)).

The underlying fact is that, in the limit $\varepsilon \rightarrow 0$, the disk chain becomes an uncoupled system of Sinai billiards. In other words, by denoting the wall collision rate by $\nu_{wall,\varepsilon}$ (i. e. for the collisions of the disks with the fixed boundaries of the domain) and the binary collision rate by $\nu_{bin,\varepsilon}$ (i. e. that of the inter-disk-collisions), one has a separation of time scales since as $\varepsilon \rightarrow 0$, $\nu_{wall,\varepsilon} (\sim \nu_{wall,crit} > 0) \gg \nu_{bin,\varepsilon} \rightarrow 0$. Though the original N -disk system is a semi-dispersing billiard in

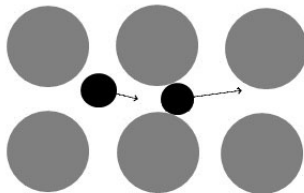


Fig. 2. A chain of length $N = 2$. Dark circles – moving disks. Grey ones – scatterers.

a $2N - 1$ -dimensional configuration space, in the given small coupling limit most of the time the disks evolve according to uncoupled, 2D dispersing billiard dynamics. Thus – on the time scale $\nu_{\text{wall},\varepsilon}^{-1}$ – there is an averaging and the binary collisions only occur on the time scale $\nu_{\text{bin},\varepsilon}$.

Theorem 3.1. (joint work with IP Tóth, [SzT 09])

$N = 2$, free boundary along x -axis. Dynamics: $(M_\varepsilon = \{q_1, v_1; q_2, v_2 | \text{dist}(q_1, q_2) \geq 2\rho_m, v_1^2 + v_2^2 = 1\}, S^\mathbb{R}, \mu_\varepsilon)$. Denote by $0 < \tau_{1,\varepsilon} < \tau_{2,\varepsilon} < \dots$ the successive binary collision times of the two disks. Then, as $\varepsilon \rightarrow 0$

- $(E_1(\nu_{\text{bin},\varepsilon}t), E_2(\nu_{\text{bin},\varepsilon}t))$ converges to a jump Markov process on the state space $E_1 + E_2 = 1$ where $E_j(t) = \frac{1}{2}v_j^2(t); j = 1, 2$
- the transition kernel $k(E_1^+ | E_1^-)$ can be calculated; it is, in fact, a verification of Boltzmann's 'microscopic chaos' property for this model (cf. G-G, '08, and also Theorem 2.1).

In other words the intercollision times for binary collisions are asymptotically exponential where the rate of the exponential clock depends on the energies of the disks. It is also worth mentioning that in the case $N = 2$ $\nu_{\text{bin},\varepsilon} \sim \text{const } \varepsilon^3$.

Idea of proof.

- since binary collisions are rare, most of the time the two disks evolve independently
- between two binary collisions – with an overwhelming probability – there is averaging in each of the in-cell, 2D billiard dynamics
- for these typically long time intervals it is natural to apply the Chernov-Dolgopyat averaging, cf. [ChD 09]
- for that purpose
 - one checks that for an incoming proper family of stable pairs, so is the outgoing family, cf. [ChD 09]
 - one applies martingale approximation for jump processes (à la Ethier-Kurtz, [EK 86]).

Theorem 3.2. (joint work with IP Tóth, [SzT 09])

$N \geq 2$. Dynamics: $(M_\varepsilon = \{q_1, v_1; \dots, q_N, v_N | \text{dist}(q_j, q_{j+1}) \geq 2\rho_m, v_1^2 + \dots + v_N^2 = 1\}, S^\mathbb{R}, \mu_\varepsilon)$.

- $(E_1(\nu_{\text{bin},\varepsilon}t), \dots, E_N(\nu_{\text{bin},\varepsilon}t))$ converges to a jump Markov process on the state space $E_1 + \dots + E_N = 1$ where $E_j(t) = \frac{1}{2}v_j^2(t); j = 1, \dots, N$
- the transition kernel $k(E_1^+, \dots, E_N^+ | E_1^-, \dots, E_N^-)$ can be expressed as the sum of the binary collision kernels of Theorem 3.1.

Acknowledgment

Zs. P.-Gy. and D. Sz. express their sincere thanks to Bálint Tóth, and also to Bálint Vető, for their advice on random walk meanders. The authors are grateful to Hungarian National Foundation for Scientific Research grants No. T 046187, K 71693, NK 63066, TS 049835 (D. Sz.) and PD73609 (I. P. T.).

References

- [BLPS 92] L. Bunimovich, C. Liverani, A. Pellegrinotti, Yu. Sukhov *Special Systems of Hard Balls that Are Ergodic*, Commun. Math. Phys. **146**, 357-396, 1992
- [BT 08] P. Bálint, I. P. Tóth, *Exponential decay of correlations in multi-dimensional dispersing billiards*, Ann. H. Poincaré, 9 (2008) 1309-1369.
- [ChD 07] N. Chernov, D. Dolgopyat. *Hyperbolic billiards and statistical physics* Proc. ICM06. Vol. II, Euro. Math. Soc., Zürich, 2006, pp 1679-1704.
- [ChD 09] N. Chernov, D. Dolgopyat. *Brownian Brownian Motion-1*, Memoirs AMS. **198**, No. 927, pp 193.
- [EK 86] SN Ethier, TG Kurtz, *Markov Processes. Characterization And Convergence*, Wiley, 1986.
- [GG 08] P. Gaspard, T. Gilbert. *Heat conduction and Fourier's law by consecutive local mixing and thermalization* Phys. Rev. Letters. **101** 020601, 2008.
- [KSz 83] A. Krámli, D. Szász. *Random Walks with Internal Degrees of Freedom*, Z. Wahrscheinlichkeitstheorie, **63**, 85-88, 1983.
- [KV 86] C. Kipnis, S. R. S. Varadhan. *Central limit theorem for additive functionals of reversible Markov processes* Commun. Math. Phys. **104**, 1-19, 1986.
- [MN 09] I. Melbourne and M. Nicol. *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*. Annals of Probability 37 (2009) 478-505.
- [MT 04] I. Melbourne and A. Torok. *Statistical limit theorems for suspension flows*. Israel Journal of Math. 144 (2004) 191-209.
- [P-GySz 09] Zs. Pajor-Gyulai, D. Szász, *Energy transfer and joint diffusion*, manuscript, pp. 30.
- [S 81] Ya. G. Sinai. *Random walks and some problems concerning Lorentz gas*. Proceedings of the Kyoto Conference, 6-17, 1981.
- [Sz 80] D. Szász. *Joint diffusion on the line*, J. of Statistical Physics. **23**, 231-240, 1980.
- [SzT 09] D. Szász, I. P. Tóth, work in progress.
- [SzV 04] D. Szász, T. Varjú, *Local Limit Theorem and Recurrence for the Planar Lorentz Process*, Ergodic Theory and Dynamical Systems, 24 (2004), 257-278
- [Y 98] L. S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Math., (1998), 585-650.

QUANTUM CALOGERO-MOSER SYSTEMS: A VIEW FROM INFINITY

ALEXANDER N. SERGEEV and ALEXANDER P. VESELOV*

*Department of Mathematical Sciences, Loughborough University,
Loughborough, Leicestershire LE11 3TU, UK*

Lomonosov Moscow State University, Moscow 119899, Russia

*E-mail: *A.P.Veselov@lboro.ac.uk*

www.lboro.ac.uk/departments/ma/people/veselov.html

Various infinite-dimensional versions of Calogero-Moser operator are discussed in relation with the theory of symmetric functions and representation theory of basic classical Lie superalgebras.

Keywords: Quantum Calogero-Moser systems; Symmetric functions, Lie superalgebras.

1. Introduction

Calogero-Moser systems play truly exceptional role in the modern theory of integrable systems (see e.g. [1] for a variety of both mathematical and theoretical physics problems they are related to). They have natural physical interpretation, describing the interaction of N particles with equal masses on the line with the inverse square potential or, in Sutherland's version, with $\sin^{-2}$ -potential.

They admit natural generalizations related to root systems and simple Lie algebras [2], and, at the quantum level only, also non-symmetric integrable versions called *deformed Calogero-Moser systems* [3], which turned out to be related to basic classical Lie superalgebras [4]. In particular, in the case of Lie superalgebra $\mathfrak{sl}(m, n)$ we have two groups of particles with two different masses with the parameters of interaction inside the groups and between them being “tuned” in a very special way (see [4]).

It turned out that these mysterious deformations can be clearly “seen from infinity”. For this one should first explain what is the analogue of the Calogero-Moser operators when $N = \infty$. A proper framework is given by the theory of symmetric functions and goes back to Stanley [5] and Macdonald [6], who were inspired by the work of H. Jack. It is interesting that Jack did his work on what is now called *Jack polynomials* around 1970 - almost at the same time as the pioneering work by Calogero and Sutherland, but a close relation between these two important developments was not recognized until much later.

2. Calogero-Moser operators in infinite dimension

Consider the following Calogero-Moser-Sutherland operator (*CMS operator*):

$$L_k^{(N)} = \sum_{i=1}^N \frac{\partial^2}{\partial x_i^2} - \sum_{i < j}^N \frac{2k(k+1)}{\sinh^2(x_i - x_j)},$$

where k is a parameter and for convenience we have changed the sign of the operator. Its gauged version $\mathcal{L}_k^{(N)} = \Psi_0^{-1}(L_N - \lambda_0)\Psi_0$ with $\Psi_0 = \prod_{i < j}^N \sinh^{-k}(x_i - x_j)$, $\lambda_0 = k^2 N(N-1)/4$ in the exponential coordinates $z_i = e^{2x_i}$ has the form

$$\mathcal{L}_k^{(N)} = \sum_{i=1}^N \left(z_i \frac{\partial}{\partial z_i} \right)^2 - k \sum_{i < j}^N \frac{z_i + z_j}{z_i - z_j} \left(z_i \frac{\partial}{\partial z_i} - z_j \frac{\partial}{\partial z_j} \right). \quad (1)$$

It preserves the algebra of symmetric polynomials $\Lambda_N = \mathbb{C}[z_1, \dots, z_N]^{S_N}$. This algebra has a natural infinite-dimensional version $\Lambda = \varprojlim \Lambda_N$ defined as the inverse limit in the category of graded algebras [6]; the elements of Λ are called *symmetric functions*.

The power sums $p_l = z_1^l + z_2^l + \dots$, $l = 1, 2, \dots$ give a convenient set of free generators of this algebra, which can be considered also as the coordinates in the corresponding infinite-dimensional space. We have a natural homomorphism $\varphi_N : \Lambda \rightarrow \Lambda_N$, sending p_l to its N -dimensional analogue $p_l = z_1^l + \dots + z_N^l$. There is a problem with $p_0 = 1 + 1 + \dots$, which in finite-dimensional situation is just the dimension N , but in infinite dimension does not make sense. The solution is to consider p_0 as a formal *additional parameter*, which our operators may depend on. Under the homomorphism φ_N it must be specialized to dimension N .

An infinite-dimensional analogue of CMS operator $\mathcal{L}_{k,p_0}^{(\infty)} : \Lambda \rightarrow \Lambda$ is defined as the unique second order differential operator polynomially dependent on p_0 , such that for all $N = 1, 2, \dots$ and $p_0 = N$ the following diagram is commutative

$$\begin{array}{ccc} \Lambda & \xrightarrow{\mathcal{L}_{k,p_0}^{(\infty)}} & \Lambda \\ \downarrow \varphi_N & & \downarrow \varphi_N \\ \Lambda_N & \xrightarrow{\mathcal{L}_k^{(N)}} & \Lambda_N \end{array}$$

The operator $\mathcal{L}_{k,p_0}^{(\infty)}$ has the following explicit form in power sums (see [5, 7, 8]):

$$\mathcal{L}_{k,p_0}^{(\infty)} = \sum_{a,b>0} p_{a+b} \partial_a \partial_b - k \sum_{a,b>0} p_a p_b \partial_{a+b} - k p_0 \sum_{a>0} p_a \partial_a + (1+k) \sum_{a>0} a p_a \partial_a, \quad (2)$$

where as before $\partial_a = a \frac{\partial}{\partial p_a}$. This form reveals an important *duality*

$$\theta^{-1} \circ \mathcal{L}_{k,p_0}^{(\infty)} \circ \theta = k \mathcal{L}_{k^{-1}, k^{-1} p_0}^{(\infty)}, \quad (3)$$

where $\theta : p_a \rightarrow k p_a, k \rightarrow k^{-1}$. There is also a remarkable symmetry between the first and the second terms, so that if following [7] we define $\tilde{\partial}_a = -\frac{a}{k} \frac{\partial}{\partial p_a}$ the operator is invariant under swapping p_a and $\tilde{\partial}_a$ (*Fourier duality*).

Note that θ changes the parameter p_0 , which means that it does not work in the finite dimensions. This fact was known already to Stanley and Macdonald, who probably were the first to discover this duality (see [5, 6]). They prefer to use the *stable version* of the CMS operator by subtracting the momentum operator $P = \sum z_i \frac{\partial}{\partial z_i} = \sum p_a \partial_a$ with dimension dependent coefficient $(N - 1)$. These stabilized CMS operators can be lifted to infinite dimension without introducing extra parameter p_0 , which corresponds to the fact that p_0 appears in (2) only as a coefficient at P , which is just another quantum integral of the system.

However, already for the rational Calogero-Moser operator

$$L_k^{(N)} = \sum_{i=1}^N \frac{\partial^2}{\partial x_i^2} - \sum_{i < j}^N \frac{2k(k+1)}{(x_i - x_j)^2}$$

the stabilization trick is not possible: its infinite-dimensional version has the form [8]

$$\mathbf{L}_{k,p_0}^{(\infty)} = \sum_{a,b \geq 1} p_{a+b-2} \partial_a \partial_b - k \sum_{a,b \geq 0} p_a p_b \partial_{a+b+2} + (1+k) \sum_{a \geq 2} (a-1) p_{a-2} \partial_a. \quad (4)$$

The same is true for the rational B_N Calogero-Moser operator

$$L_{k,l}^N = \Delta_N - \sum_{i < j}^N \left(\frac{2k(k+1)}{(x_i - x_j)^2} + \frac{2k(k+1)}{(x_i + x_j)^2} \right) - \sum_{i=1}^N \frac{l(l+1)}{x_i^2},$$

with 2 parameters k, l . The corresponding infinite-dimensional analogue (after a convenient division by 4) in the coordinates $z_i = x_i^2$ has a very similar form [8]:

$$\begin{aligned} \mathcal{B}_{k,l,p_0}^{(\infty)} &= \sum_{a,b \geq 1} p_{a+b-1} \partial_a \partial_b - k \sum_{a,b \geq 1} p_a p_b \partial_{a+b+1} + (1+k) \sum_{a \geq 1} a p_{a-1} \partial_a \\ &\quad - (2kp_0 + l + 1/2) \sum_{a \geq 1} p_{a-1} \partial_a + kp_0^2 \partial_1. \end{aligned} \quad (5)$$

In the trigonometric BC_N case we have the operator

$$L_{k,p,q}^N = \Delta_N - \sum_{i < j}^N \left(\frac{2k(k+1)}{\sinh^2(x_i - x_j)} + \frac{2k(k+1)}{\sinh^2(x_i + x_j)} \right) - \sum_{i=1}^N \left(\frac{p(p+2q+1)}{\sinh^2 x_i} + \frac{4q(q+1)}{\sinh^2 2x_i} \right),$$

depending on 3 parameters k, p, q . The BC_∞ CMS operator has the form [9]:

$$\begin{aligned} \mathcal{L}_{k,p,q,h}^{(\infty)} &= \sum_{a,b > 0} (p_{a+b} + 2p_{a+b-1}) \partial_a \partial_b - k \sum_{a=2}^{\infty} \left[\sum_{b=0}^{a-2} p_{a-b-1} (2p_b + p_{b+1}) \right] \partial_a \\ &\quad + \sum_{a=1}^{\infty} [(a+k(a+1)+2h)p_a + (2a-1+2ka+2h-p)p_{a-1}] \partial_a, \end{aligned} \quad (6)$$

where the additional parameter h is related to p_0 as $h = -(kp_0 + \frac{1}{2}p + q)$. This form is invariant under three involutions (dualities) [9]. One of them is an extension of the duality $k \rightarrow k^{-1}$ from the previous case. It acts on the other parameters as

$$2\hat{h} - 1 = k^{-1}(2h - 1), \quad \hat{p} = k^{-1}p, \quad (2\hat{q} + 1) = k^{-1}(2q + 1). \quad (7)$$

The same relations first appeared in the formulas for the $BC(m, n)$ deformed CMS operators related to the orthosymplectic Lie superalgebras $\mathfrak{osp}(m, 2n)$ [4]. We are going to see now that this is not a mere coincidence.

3. Deformed CMS operators and classical Lie superalgebras seen from infinity

We restrict ourselves by the BC case (see the A case in [10]). The $BC(m, n)$ deformed CMS operators have the following form [4]:

$$\begin{aligned}
 L^{m,n} = & \Delta_m + k\Delta_n - \sum_{i < j}^m \left(\frac{2k(k+1)}{\sinh^2(x_i - x_j)} + \frac{2k(k+1)}{\sinh^2(x_i + x_j)} \right) \\
 & - \sum_{i < j}^n \left(\frac{2(k^{-1}+1)}{\sinh^2(y_i - y_j)} + \frac{2(k^{-1}+1)}{\sinh^2(y_i + y_j)} \right) \\
 & - \sum_{i=1}^m \sum_{j=1}^n \left(\frac{2(k+1)}{\sinh^2(x_i - y_j)} + \frac{2(k+1)}{\sinh^2(x_i + y_j)} \right) + \sum_{i=1}^m \frac{p(p+2q+1)}{\sinh^2 x_i} \\
 & + \sum_{i=1}^m \frac{4q(q+1)}{\sinh^2 2x_i} - \sum_{j=1}^n \frac{kr(r+2s+1)}{\sinh^2 y_j} - \sum_{j=1}^n \frac{4ks(s+1)}{\sinh^2 2y_j}, \quad (8)
 \end{aligned}$$

where the parameters k, p, q, r, s satisfy the following relations (cf. formula (7)):

$$r = k^{-1}p, \quad 2s + 1 = k^{-1}(2q + 1).$$

Consider the algebra $\Lambda_{m,n,k}$ consisting of polynomials, which are symmetric in $u_1, \dots, u_m$ and $v_1, \dots, v_n$ separately and satisfy the conditions

$$u_i \frac{\partial f}{\partial u_i} - kv_\alpha \frac{\partial f}{\partial u_\alpha} = 0$$

on the hyperplanes $u_i = v_\alpha$ for all $i = 1, \dots, m$ and $\alpha = 1, \dots, n$. For generic values of parameter k this algebra is generated by the *deformed power sums*

$$p_a(u, v, k) = \sum_{i=1}^m u_i^a + k^{-1} \sum_{\alpha=1}^n v_\alpha^a, \quad a = 1, 2, \dots$$

The claim is that the $BC(m, n)$ deformed CMS operators are the restrictions of BC_∞ CMS operator (6) onto the corresponding subvariety $\mathcal{D}(m, n, k) = \text{Spec } \Lambda_{m,n,k}$. More precisely, let $\varphi_{m,n} : \Lambda \longrightarrow \Lambda_{m,n,k}$ be the restriction homomorphism defined by $\varphi_{m,n}(p_a) = p_a(u, v, k)$.

Theorem 3.1 (cf. [9]). *The following diagram is commutative for $h = -km - n - \frac{1}{2}p - q$ and generic values of parameter k :*

$$\begin{array}{ccc}
 \Lambda & \xrightarrow{\mathcal{L}^{(k,p,q,h)}} & \Lambda \\
 \downarrow \varphi_{m,n} & & \downarrow \varphi_{m,n} \\
 \Lambda_{m,n,k} & \xrightarrow{\mathcal{L}^{(m,n)}} & \Lambda_{m,n,k}
 \end{array} \quad (9)$$

where $\mathcal{L}^{(m,n)}$ is a gauged version of the deformed CMS operator (8) rewritten in the coordinates $u_i = 2 \sinh^2 x_i$, $v_j = 2 \sinh^2 y_j$.

A remarkable fact is that these are essentially all possible restrictions of BC_∞ CMS operator (see Theorem 4.6 in [9]), which shows the exceptional role of the deformed CMS operators. The eigenfunctions $\mathcal{J}_\lambda(x; k, p, q, h)$ of the BC_∞ CMS operator are labelled by partitions λ and known as *Jacobi symmetric functions*. Their images

$$\mathcal{S}J_\lambda(u, v; k, p, q) = \varphi_{m,n}(\mathcal{J}_\lambda(x; k, p, q, h)),$$

where $h = -km - n - \frac{1}{2}p - q$, are called *super Jacobi polynomials*. It turns out that their specialized versions have a natural interpretation in the representation theory as the *Euler supercharacters* of the orthosymplectic Lie superalgebras $\mathfrak{osp}(m, 2n)$.

There is a classical construction due to Borel, Weil and Bott of the irreducible representations of the complex semisimple Lie groups G in terms of the cohomology of the holomorphic line bundles over the corresponding flag varieties G/P . In the Lie supergroup case this leads to a virtual representation given by the Euler characteristic

$$\mathcal{E}_\lambda = \sum (-1)^i H^i(G/P, \mathcal{O}_\lambda)$$

for certain sheaf cohomology groups; the corresponding supercharacter E_λ is called Euler supercharacter (see [12]).

Theorem 3.2 ([10]). *For special choices of parabolic subgroup P the Euler supercharacters of $\mathfrak{osp}(2m+1, 2n)$ coincide with specialized super Jacobi polynomials*

$$E_\lambda = \mathcal{S}J_\lambda(u, v; -1, -1, 0).$$

A similar fact holds for Lie superalgebra $\mathfrak{osp}(2m, 2n)$ and $\mathcal{S}J_\lambda(u, v; -1, 0, 0)$, but the analogue of this result for the Lie superalgebra $\mathfrak{sl}(m, n)$ is still to be found (see [8] for further discussion).

References

- [1] J. F. van Dijen and L. Vinet (Editors) *Calogero-Moser-Sutherland models*. CRM Ser. Math. Phys., Springer, New York, 2000.
- [2] M.A. Olshanetsky, A.M. Perelomov, *Phys. Rep.* **94**, 313–404 (1983).
- [3] O. Chalykh, M. Feigin, A.P. Veselov, *J. Math. Phys.* **39** 695–703 (1998).
- [4] A.N. Sergeev, A.P. Veselov, *Comm. Math. Phys.* **245** 249–278 (2004).
- [5] R. Stanley, *Adv. Math.* **77**, 76–115 (1989).
- [6] I.G. Macdonald *Symmetric functions and Hall polynomials*. Oxford Univ. Press, 1995.
- [7] H. Awata, in [1], 23–35 (2000).
- [8] A.N. Sergeev, A.P. Veselov, arXiv:0910.1984 (2009).
- [9] A.N. Sergeev, A.P. Veselov, *Adv. Math.* **222**, 1687–1726 (2009).
- [10] A.N. Sergeev, A.P. Veselov, arXiv:0905.2603 (2009).
- [11] A.N. Sergeev, A.P. Veselov, *Adv. Math.* **192**, 341–375 (2005).
- [12] V. Serganova, in *Proc. ICM (Berlin, 1998)*. *Doc. Math.*, Extra Vol. II, 583–593 (1998).

OTHER TALKS

INVITED TALKS

UNBOUNDED ORBITS IN POSITIVE DEFINITE HAMILTONIAN SYSTEMS

CHONG-QING CHENG, XIA LI

Department of Mathematics, Nanjing University, China

E-mail: chengcq@nju.edu.cn

In this talk, we briefly describe how to construct unbounded orbits exhibited by the systems:

$$H(\xi, t) = T(\xi) + \epsilon U(\pi\xi, t)$$

where $\xi \in T\mathbb{T}^2$, T represents a Riemannian metric and U is a potential, periodically depending on t . T is assumed to have exactly one closed geodesic Γ for some indivisible $h_0 \in H_1(\mathbb{T}^2, \mathbb{Z})$ and there is only one positive Morse geodesic Λ^+ as well as only one negative geodesic Λ^- , both approach to Γ as time goes to positive and negative infinity. We show that for generic potential perturbation U , there exists an orbits whose energy goes to infinity as time goes to infinity. In the previous work of John Mather, he has shown the existence of unbounded orbits, but the initial value needs to be sufficiently large. Here, the initial speed is not assumed sufficiently large, in fact, the component of the initial velocity along Γ can be any value.

PARTIAL HYPERBOLICITY, CHAOS AND SYMMETRY

AMIE WILKINSON

Department of Mathematics, Northwestern University,

2033 Sheridan Road, Evanston, IL 60208, USA

E-mail: wilkinso@math.northwestern.edu

Partial hyperbolicity arises in a variety of classical contexts, in systems of both algebraic and geometric origin. These systems are subject to a range of symmetries and at yet are typically chaotic (ergodic, mixing, etc). In this talk I will survey results about ergodicity of conservative partially hyperbolic systems and discuss the interplay between the mechanisms for ergodicity and symmetry, which lead to various rigidity phenomena.

CONTRIBUTED TALKS

ON THE INTEGRALS OF MOTION FOR A CALOGERO-MOSER SYSTEM WITH SPINS AND ELLIPTIC EXCHANGE INTERACTION

JAROSLAV DITTRICH*, VLADIMIR I. INOZEMTSEV

**Nuclear Physics Institute ASCR, CZ-250 68 Řež, Czech Republic
E-mail: dittrich@ujf.cas.cz*

The problem of finding integrals of motion for a quantum elliptic Calogero-Moser system with arbitrary number of particles extended by spin-exchange interaction with the strength given by Weierstrass elliptic function of the particle distance is considered. Two mutually commuting non-trivial integrals of motion are found using an Ansatz and straightforward calculations based on the elliptic functions properties.

NEKHOROSHEV THEOREM FOR THE PERIODIC TODA LATTICE

ANDREAS HENRICI, THOMAS KAPPELER

*Institut für Mathematik, Universität Zürich
E-mail: andreas.henrici@math.uzh.ch*

We consider the periodic Toda lattice with N particles, which is globally symplectomorphic to a 2-parameter family of $N - 1$ coupled harmonic oscillators, and the action variables fill out the whole positive quadrant of $\mathbb{R}^{N-1}$. We prove that in the interior of the positive quadrant and a neighborhood of the origin, the Hamiltonian is strictly convex in the action variables, and therefore the Nekhoroshev theorem on perturbed integrable systems applies on almost all of the phase space. The main technical tool is a version of a theorem by Krichever concerning the period map of certain Abelian differentials on a Riemann surface associated to the phase space of the system.

A RESONANCE OF THE CYCLOTRON FREQUENCY WITH A TIME-PERIODIC SINGULAR FLUX TUBE

JOACHIM ASCH, TOMÁŠ KALVODA*, PAVEL ŠTŮVÍČEK

**Department of Mathematics, FNSPE, Czech Technical University
E-mail: kalvotom@fjfi.cvut.cz*

We study the dynamics of a classical charged particle moving on a punctured plane under the influence of a homogeneous magnetic field, and driven by a periodically time-dependent singular flux tube through the hole. For this system we exhibit a resonance effect depending on the frequency of the time-dependent flux tube and on the cyclotron frequency: the cyclotron orbits blow up, and the particle oscillates between the hole and

infinity. This observation is supported by an analytic study of two different approximations for small amplitudes of the flux.

THE NONCLASSICAL APPROACH TO ANALYSIS OF HAMILTONIAN AND CONSERVATIVE SYSTEMS

NIKOLAI A. MAGNITSKII

Institute for Systems Analysis of Russian Academy of Sciences, Moscow, Russia

E-mail: mag@su29.ru

The new approach is offered for the analysis of solutions of complex perturbed conservative (in particular, Hamiltonian) systems. This approach implies the construction of an approximating extended two-parametrical dissipative system of the equations, whose stable solutions (attractors) are as much as exact approximations to solutions of the original conservative system. On the basis of the carried out numerical calculations for several three- and four-dimensional conservative systems and Hamiltonian systems with two and three degrees of freedom, it is shown, that in all these systems transition to chaos occurs not through destruction of two-dimensional or three-dimensional tori of unperturbed system, but, conversely, through a generation of complex two-dimensional tori around of cycles of extended dissipative system and through the infinite cascade of bifurcations of generations of new cycles and singular trajectories according to the Feigenbaum-Sharkovskii-Magnitskii (FSM) theory.

[1] Magnitskii N. A., Sidorov S.V. New Methods for Chaotic Dynamics. - Singapore: World Scientific, 2006, 363p.

[2] Magnitskii N. A. Universal theory of dynamical chaos in nonlinear dissipative systems of differential equations. - Comm. Nonlin. Sci. Numer. Simul., 13, 2008, 416-433p.

[3] Magnitskii N. A. New approach to analysis of Hamiltonian and conservative systems. - Differential equations, 2008, 44, 12, 1682-1690.

INTEGRABILITY ANALYSIS OF A NONLOCAL OSTROVSKY-WHITHAM FLOW

ANATOLIY PRYKARPATSKY, DENIS L. BLACKMORE*, JOLANTA GOLENIA

**The New Jersey Institute of Technology, Newark, NJ 07102, USA*

E-mail: denis.l.blackmore@njit.edu

1. The Ostrovsky-Whitham medium model with spatial memory: We consider an Ostrovsky-Whitham type nonlinear medium model described by the evolution equation $du = dt = 2uu_x + \int_{\mathbf{R}} K(u; s)u_s ds$ discussed first in [2]. If the corresponding Whitham kernel K is of the form $K(x; s) := i\mu(x|s)us$ for $x, s \in \mathbf{R}$; naturally modeling [1] the relaxing spatial memory effects, the equation becomes $u = dt = 2uu_x + \int_{\mathbf{R}} i\mu(x|s)us u_s ds$, or $\dot{u} = K[u]$, and is strongly nonlocal and appears to possess very interesting mathematical properties.

2. Regularization and Lax integrability: We prove that the following regularized nonlinear dynamical system $u_t = 2uu_x|v; v_t = 2uv_x$, or $(u_t, v_t) := K[u, v]$, which is of hydrodynamic type and well defined on the extended $2\pi/4$ -periodic function space $M := C_{9/4}^1(\mathbf{R}; \mathbf{R})$, and completely equivalent to that given by expression $\dot{u} = K[u]$, which is a Lax integrable bi-Hamiltonian flow.

- [1] Ostrovsky L.A. Nonlinear internal waves in a rotating ocean, *Oceanology*, 1978, v. 18, p. 119-125.
- [2] Whitham G.B. *Linear and Nonlinear Waves*. Wiley-Interscience, New York, 1974.

LIMIT OF THE INFINITE HORIZON DISCOUNTED HAMILTON JACOBI EQUATION

RENATO ITTURIAGA, HÉCTOR SANCHEZ-MORGADO

Universidad Nacional Autonoma de Mexico

E-mail: hector@math.unam.mx

We consider the infinite horizon discounted optimal control problem and study the convergence of solutions of the Hamilton Jacobi equation when the discount vanishes. If the Aubry set consists in a finite number of hyperbolic periodic orbits or critical points, we give an explicit expression for the limit using the Peierls barrier. Additionally, we give a new characterization of Mañé's critical value as for which the family of viscosity solutions is equibounded.

CLASSICAL MOTION IN RANDOM POTENTIALS

CHRISTOPH SCHUMACHER

Mathematisches Institut, Universität Erlangen

E-mail: schumacher@mi.uni-erlangen.de

The classical motion in two dimensions generated by a Hamiltonian function with random potential with Coulomb singularities is intimately connected with the motion on a negatively curved Riemannian surface of infinite genus and exhibits a lot of parallels with chaotic billiards. In this talk I indicate how to construct a geometric Markov partition in the spirit of C. Series and how this can be used to show diffusive behaviour and ergodicity of the Hamiltonian flow restricted to an energy.

GENERALIZED STÄCKEL TRANSFORM: INTEGRABILITY AND BEYOND

ARTHUR SERGYEYEV, MACIEJ BLASZAK

Silesian University in Opava, Czech Republic

E-mail: Artur.Sergyeyev@math.slu.cz

In this work we introduce multiparameter generalized Stäckel transform – a noncanonical transformation of a special kind relating the sets of (not necessarily commuting) Hamiltonians and leaving the phase space coordinates intact. It is shown that under certain conditions the transformation in question preserves Liouville integrability, noncommutative integrability and superintegrability. For instance, when applied to an n -tuple of commuting Hamiltonians, this transformation yields a (new) n -tuple of commuting Hamiltonians. Thus, one can construct new integrable systems from the old ones and also find new links among known integrable systems using the transformation under study. In particular,

we show that integrable systems associated with the so-called k -hole deformations of the separation curve of Benenti type are related to the ones associated with the original separation curve of Benenti type through an appropriately chosen multiparameter generalized Stäckel transform.

The corresponding transformation for the equations of motion proves to be nothing but a reciprocal transformation of a special form, and we briefly discuss the properties of this class of reciprocal transformations, including applications to hydrodynamic-type systems. For more details see A. Sergeev, M. Blaszak, Generalized Stäckel transform and reciprocal transformations for finite-dimensional integrable systems, *J. Phys. A: Math. Theor.* 41 (2008), 105205.

THE SPECTRAL EDGE OF SOME RANDOM BAND MATRICES

ALEXANDER SODIN

Tel Aviv University

E-mail: sodinale@post.tau.ac.il

We consider random $N \times N$ band matrices of band width W . If the band is wide ($W \ll N^{5/6}$), the extreme eigenvalues behave similarly to those of GUE matrices; in particular, the largest eigenvalue converges to the Tracy-Widom distribution. Otherwise, a different limit appears. These results are consistent with the Thouless criterion, adapted to the band matrix setting by Fyodorov and Mirlin.

LONG-TIME ASYMPTOTICS OF THE TODA LATTICE FOR ASYMPTOTICALLY PERIODIC INITIAL CONDITIONS

SPYRIDON KAMVISSIS, HELGE KRÜGER, GERALD TESCHL*

**Fakultät für Mathematik, Universität Wien*

E-mail: Gerald.Teschl@univie.ac.at

We consider the stability of periodic (and slightly more generally of algebro-geometric finite-gap) solutions of the doubly infinite Toda lattice under a short range perturbation. We prove that the perturbed lattice asymptotically approaches a modulated lattice. More precisely, let g be the genus of the hyperelliptic curve associated with the unperturbed solution. We show that, apart from the phenomenon of the solitons travelling on the quasi-periodic background, the n/t -plane contains $g+2$ areas where the perturbed solution is close to a finite-gap solution in the same isospectral torus. In between there are $g+1$ regions where the perturbed solution is asymptotically close to a modulated lattice which undergoes a continuous phase transition (in the Jacobian variety) and which interpolates between these isospectral solutions. In the special case of the free lattice ($g=0$) the isospectral torus consists of just one point and we recover the known result.

Both the solutions in the isospectral torus and the phase transition are explicitly characterized in terms of Abelian integrals on the underlying hyperelliptic curve.

Our method relies on the equivalence of the inverse spectral problem to a matrix Riemann-Hilbert problem defined on the hyperelliptic curve and generalizes the so-called nonlinear stationary phase/steepest descent method for Riemann-Hilbert problem deformations to Riemann surfaces.

POSTERS

ALENA I. ALEKSENKO (Department of Mathematics, University of Aveiro, Portugal), with ALEXANDER YU. PLAKHOV: **BODIES INVISIBLE IN ONE DIRECTION AND BODIES OF ZERO TOTAL CROSS SECTION**

PETR AMBROŽ (Doppler Institute & Dept of Mathematics, Czech Technical University), with ZUZANA MASÁKOVÁ and EDITA PELANTOVÁ: **ON THE NUMBER OF FACTORS IN CODINGS OF THREE INTERVAL EXCHANGE**

CHRISTOPHER CAMPBELL (University of Queensland): **FUNCTIONAL RELATIONS FOR TRANSFER MATRICES WITH $D(D3)$ SYMMETRY**

VETURIA CHIROIU (Institute of Solid Mechanics, Romanian Academy), with LIGIA MUNTEANU: **ON THE PHENOMENOLOGICAL UNIVERSALITY WITH APPLICATIONS TO POPULATION DYNAMICS**

EMILIO CORTÉS (Universidad Autónoma Metropolitana, Mexico City), with J. FUJIOKA, A. ESPINOZA-CERÓN, R.F. RODRÍGUEZ, and R. PÉREZ-PASUAL: **RESPONSE OF THE ALGEBRAIC SOLITONS OF A GENERALIZED NLS EQUATION TO PERIODIC PERTURBATIONS**

PETER FINCH (University of Queensland): **THE DIHEDRAL GROUP, DESCENDANTS AND THE YANG-BAXTER EQUATION**

PYOTR N. IVANSKIN (Kazan State University), with ELENA A. SHIROKOVA: **SPLINE-INTERPOLATION SOLUTION OF THE PROBLEM OF ELASTICITY FOR THE SOLID OF REVOLUTION**

ALEXANDER I. KOMECH (Mathematics, Vienna University, and IITP RAS), with ANDREW KOMECH: **GLOBAL ATTRACTION TO QUANTUM STATIONARY STATES**

SERGEY KOMECH (IITP, Russian Academy of Sciences, Moscow): **ENTROPY AND THE BOUNDARY DISTORTION RATE FOR SMOOTH DYNAMICAL SYSTEMS**

ELENA A. KOPYLOVA (IITP, Russian Academy of Sciences, Moscow), with ALEXANDER I. KOMECH: **ON ASYMPTOTIC STABILITY OF KINK FOR RELATIVISTIC GINSBURG-LANDAU EQUATION**

OLGA KRUPKOVÁ (Palacký University Olomouc), with JANA MUSILOVÁ: **METHODS OF NONHOLONOMIC MECHANICS IN THE SPECIAL RELATIVITY THEORY**

SERGEY KRYZHEVICH (Sankt Petersburg State University): **BIFURCATIONS AND CHAOS IN S.D.F. VIBRO-IMPACT SYSTEMS**

MICHAL MACEK (Charles University Prague), with PAVEL STRÁNSKÝ and PAVEL CEJNAR: **PERES LATTICES AS INDICATORS OF QUANTUM CHAOS**

EDUARDO MARTÍNEZ (IUMA and Department of Applied Mathematics, University of Zaragoza): **HAMILTON-JACOBI THEORY ON LIE ALGEBROIDS**

CONCEPTION MURIEL (Departamento de Matemáticas, Universidad de Cadiz), with J.L. ROMERO: **ON THE LAMBDA-SYMMETRIES OF SECOND-ORDER NONLINEAR OSCILLATORS**

JUAN F. NAVARRO (Dept. Matemática Aplicada, Universidad de Alicante): **SYMBOLIC COMPUTATION APPLIED TO PERTURBATION METHODS**

SYLVIE OLIFFSON KAMPFHORST (Departamento de Matematica – UFMG – Brasil), with R. MARKARIAN, R.P. DE CARVALHO: **DOMINATED SPLITTING AND LIMIT SETS OF CONVEX PINBALL BILLIARDS**

GIOVANNI RASTELLI (Dept of Mathematics, University of Torino), with CLAUDIA CHANU and LUCA DEGIOVANNI: **SUPERINTEGRABLE THREE BODY SYSTEMS IN ONE DIMENSION AND GENERALIZATIONS**

YUNBO ZENG (Department of Mathematical Sciences, Tsinghua University), with XIAOJUN LIU: **A GENERALIZED DRESSING APPROACH FOR SOLVING THE EXTENDED KP HIERARCHY**

Equilibrium Statistical Mechanics

(classical and quantum)

Session organizers:

Marek Biskup
David Brydges

RENORMALIZATION GROUP TRAJECTORIES BETWEEN TWO FIXED POINTS

ABDELMALEK ABDESSELAM

*Department of Mathematics, University of Virginia,
P. O. Box 400137, Charlottesville, VA 22904-4137, USA
E-mail: malek@virginia.edu*

We report on our recent rigorous construction of complete renormalization group trajectories between two fixed points for the three-dimensional phi-four model with modified propagator considered by Brydges, Mitter and Scoppola (BMS). These are discrete critical trajectories which connect the ultraviolet Gaussian fixed point to the nontrivial BMS infrared fixed point which is an analogue of the Wilson–Fisher fixed point. The renormalization group map is defined rigorously and nonperturbatively, without using the hierarchical approximation. The trajectories are constructed by a fixed point argument in a suitable Banach space of sequences, where one perturbs a nonlinear one-dimensional iteration.

Keywords: renormalization group, crossover, heteroclinic orbit, invariant manifolds, massless flow, Wilson-Fisher fixed point.

1. Global Dynamics of Wilson’s Renormalization Group

A central mathematical problem posed by quantum field theory (QFT) is that of giving a rigorous meaning to functional integrals such as

$$\int_{\mathcal{F}} D\phi \dots e^{-\int_{\mathbb{R}^d} [\frac{1}{2}(\nabla\phi)^2(x) + \mu\phi(x)^2 + g\phi(x)^4]dx}.$$

In this example corresponding to the so-called ϕ^4 theory, $\mathcal{F}$ is the infinite-dimensional space of functions from $\mathbb{R}^d$ to $\mathbb{R}$, $D\phi$ is the Lebesgue measure on $\mathcal{F}$, and the dots stand for some observable or functional of the field ϕ . A natural approach to this problem is to try to construct such a measure on functions or rather distributions in $\mathcal{S}'(\mathbb{R}^d)$ by a scaling limit of lattice theories on $(a\mathbb{Z})^d \subset \mathbb{R}^d$. This corresponds to imposing a cut-off $\frac{1}{a}$ on momenta in Fourier space. One can easily rescale to a unit lattice and see the approximants to the continuum theory as living on the lattice $\mathbb{Z}^d$ instead of $(a\mathbb{Z})^d$. These approximants become points in the space of all possible unit cut-off theories. In essence, Wilson’s renormalization group (RG) is a dynamical system acting on this space. If $d\nu$ is a measure on random fields ϕ with $\hat{\phi}(p) = 0$ if $|p| > 1$, i.e. with sharp unit cut-off on momenta, one introduces a magnification ratio $L > 1$ and splits the field as $\phi = \zeta + \phi_{\text{low}}$. The fluctuation field ζ corresponds to keeping Fourier modes with $L^{-1} < |p| \leq 1$, whereas the low

momentum or background field ϕ_{low} has Fourier modes with $|p| \leq L^{-1}$. One can then integrate over ζ , and obtain the marginal probability distribution on ϕ_{low} coming from the original probability distribution $d\nu$. Finally one rescales back to unit lattice by letting $\psi(x) = L^{[\phi]} \phi_{\text{low}}(Lx)$, and considering the resulting probability distribution $d\nu'$ on ψ . The RG map is the transformation $d\nu \rightarrow d\nu'$. Important features of this dynamical system are the notions of fixed points, eigenvalues of linearized RG around them, local stable and unstable manifolds etc. They play a key role in controlling scaling limits and constructing continuum functional integrals by lattice approximations. The features just listed which have also been the most studied ones are local, i.e., pertaining to the vicinity of single fixed points. Much less is known about global features of the RG dynamical system, e.g., heteroclinic trajectories between fixed points. This is the main object of this article which briefly reports on recent progress we made in this area [1].

2. Rigorous Results (Selection)

There is a long history of rigorous results on the RG dynamical system. We will only mention a selection of such results which gives the best idea of the context into which the results of our article [1] fit.

The RG exponents for the Gaussian fixed point, the existence of a nontrivial infrared fixed point in ‘ $4 - \epsilon$ ’ dimensions, as well as the construction of its local stable manifold, were obtained by Bleher and Sinai [2, 3], in the hierarchical model (HM) approximation. Similar constructions were also provided by Collet and Eckmann [4, 5], Gawędzki and Kupiainen [6, 7]. Furthermore, the two-dimensional local unstable manifold of the Gaussian fixed point for the ϕ_3^4 HM was constructed [8]. The hierarchical ϕ_3^4 infrared fixed point (i.e. for $\epsilon = 1$) was also constructed [9]. The appearance of new fixed points at dimensions $d = 2 + \frac{2}{n-1}$, $n = 3, 4, \dots$ in the local potential approximation was established by Felder [10].

As per the Euclidean model, i.e., the full model without hierarchical or local potential approximations, the existence of a nontrivial infrared fixed point as well as the control of its local stable manifold was provided by Brydges, Dimock and Hurd [11]. Later, a similar result for a nicer ‘ $(4 - \epsilon)$ -dimensional’ model was derived by Brydges, Mitter and Scoppola [12].

The previous results concern local features of the RG. Rigorous results of a global nature are more scarce. One can mention the proof of uniqueness of the infrared fixed point in the local potential approximation in dimension $3 \leq d < 4$ by Lima [13]. Closest to our result is the construction of the massless Gross-Neveu model in ‘ $2 + \epsilon$ ’ dimension by Gawędzki and Kupiainen [14]. Their result amounts to controlling a heteroclinic trajectory joining a nontrivial ultraviolet fixed point to a trivial infrared fixed point, for a Fermionic QFT. In our article [1] which concerns the Bosonic BMS model [12], we obtained the construction of discrete heteroclinic trajectories joining a Gaussian ultraviolet fixed point to a nontrivial infrared fixed point. One should finally also mention related work on massless nontrivial QFT

models such as ϕ^4 in two and three dimensions [15–17] or the Thirring model [18].

3. The BMS Model

The BMS model for a scalar field $\phi : \mathbb{R}^3 \longrightarrow \mathbb{R}$ corresponds to the functional integral

$$Z = \underbrace{\int D\phi \, e^{-\frac{1}{2}\langle \phi, (-\Delta)^{\frac{3+\epsilon}{4}} \phi \rangle_{L^2(\mathbb{R}^3)}}}_{\text{Gaussian measure}} - \overbrace{\int dx (g:\phi^4(x): + \mu:\phi^2(x):)}^{\text{potential } V(\phi)}.$$

The free propagator is $(-\Delta)^{-\frac{3+\epsilon}{4}}(x, y) \sim \frac{1}{|x-y|^{2[\phi]}}$ where $[\phi] = \frac{3-\epsilon}{4}$ denotes the canonical scaling dimension of the field. One also has a multiscale decomposition of this propagator, in the form $\int_0^\infty \frac{dl}{l} l^{-2[\phi]} u\left(\frac{x-y}{l}\right)$ with u a smooth, compactly supported, rotationally symmetric function which is nonnegative in both x and p . Instead of a lattice regularization, it is more convenient to use continuous cut-offs, by taking $C(x-y) = \int_1^\infty \frac{dl}{l} l^{-2[\phi]} u\left(\frac{x-y}{l}\right)$ as a free propagator. The high versus low momentum split corresponds to the decomposition $C(x-y) = \Gamma(x-y) + C_{L^{-1}}(x-y)$ with $C_{L^{-1}}(x-y) = L^{-2[\phi]} C(L^{-1}(x-y))$. The fluctuation covariance therefore is $\Gamma(x-y) = \int_1^L \frac{dl}{l} l^{-2[\phi]} u\left(\frac{x-y}{l}\right)$. One can rewrite accordingly the original Gaussian measure as a convolution $d\mu_C = d\mu_\Gamma \star d\mu_{C_{L^{-1}}}$. The functional integral, with integrand $\mathcal{Z}(\phi)$, becomes

$$Z = \int d\mu_C(\phi) \, \mathcal{Z}(\phi) = \int d\mu_{C_{L^{-1}}}(\psi) d\mu_\Gamma(\zeta) \, \mathcal{Z}(\psi + \zeta) = \int d\mu_C(\phi) \, (\mathcal{R}\mathcal{Z})(\phi)$$

where $(\mathcal{R}\mathcal{Z})(\phi) = \int d\mu_\Gamma(\zeta) \, \mathcal{Z}(\phi_{L^{-1}} + \zeta)$ and $\phi_{L^{-1}}(x) = L^{-[\phi]}\phi(L^{-1}x)$. Now the RG map can be seen as acting on integrands: $\mathcal{Z} \longrightarrow \mathcal{R}\mathcal{Z}$.

4. Good Infinite Volume Coordinates

An important ingredient for the rigorous implementation of the RG is a good set of coordinates for the dynamical variable $\mathcal{Z}$, which behave well in the infinite volume limit. Such a formalism was introduced by Brydges and Yau [19]. One uses the decomposition of $\mathbb{R}^3$ into unit cubes or cells defined by a lattice $\mathbb{Z}^3$. If Λ is the large finite box where the model is defined, one writes the following polymer representation for the integrand:

$$\begin{aligned} \mathcal{Z}(\Lambda, \phi) &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\substack{X_1, \dots, X_n \\ \text{disjoint in } \Lambda}} \exp \left[- \int_{\Lambda \setminus (\cup X_i)} dx \{ g : \phi^4(x) :_C + \mu : \phi^2(x) :_C \} \right] \\ &\quad \times K(X_1, \phi|_{X_1}) \cdots K(X_n, \phi|_{X_n}). \end{aligned} \tag{1}$$

The X_i are polymers, i.e., connected finite unions of closed unit cells. The functionals $K(X, \phi)$ are local in the sense that they only depend on the restriction of the field to the sets X . The coordinates for $\mathcal{Z}$ now are triples (g, μ, K) where g, μ are the running couplings for the ϕ^4 and mass terms, and $K = (K(X, \cdot))_{X \text{ polymer}}$ is a collection of

local functionals. The need to explicitly extract second order perturbation theory, in order to analyse the RG dynamics, imposes an additional rewriting of the form

$$K(X, \phi) = g^2[\text{explicit complicated formula}]e^{-V(X, \phi)} + R(X, \phi) .$$

So the dynamical variable becomes the collection of functionals R which can be thought of as an $O(g^3)$ remainder. As a result, the RG for the BMS model can be written as an explicit (albeit complicated) map $(g, \mu, R) \longrightarrow (g', \mu', R')$. It has the form

$$\begin{aligned} g' &= L^\epsilon g - L^{2\epsilon} a(L, \epsilon) g^2 + \xi_g(g, \mu, R), \\ \mu' &= L^{\frac{3+\epsilon}{2}} \mu + \xi_\mu(g, \mu, R) , \\ R' &= \mathcal{L}^{(g, \mu)}(R) + \xi_R(g, \mu, R) . \end{aligned}$$

The quantity $a(L, \epsilon)$ is $\sim \log L$. The ξ 's are small nonlinear remainder terms. The $\mathcal{L}^{g, \mu}$ stands for a (g, μ) -dependent linear operator in R . Its crucial property is that it is contractive, which follows from the so-called ‘‘extraction step’’ in the definition of the RG map. This is the analogue in the present setting of the traditional BPHZ subtractions in renormalization theory [20]. It removes the dangerous local part of the two and four point dependence in the low momentum field $\phi_{L^{-1}}$, from the functionals R and transfers them to the exponentiated potential. This relies on the nonuniqueness of the polymer representation of Eq. 1 for a given integrand $\mathcal{Z}$.

Ignoring the ξ remainders makes for a simplified RG map with easy dynamics. In addition to the trivial fixed point $(0, 0, 0)$, it has a fixed point at $(\bar{g}_*, 0, 0)$ with $\bar{g}_* = \frac{L^\epsilon - 1}{L^{2\epsilon} a} \sim \epsilon$. It also has heteroclinic trajectories $(\bar{g}_n, 0, 0)_{n \in \mathbb{Z}}$ parametrized by the single number $\bar{g}_0$ in the interval $(0, \bar{g}_*)$, and constructed using the one-dimensional iteration by the function $f(x) = L^\epsilon x - L^{2\epsilon} a x^2$. The BMS fixed point was obtained as a perturbation of the approximate fixed point $(\bar{g}_*, 0, 0)$. Similarly, the true RG connecting orbits are constructed by perturbing the above approximate trajectories. Our main result [1] is as follows.

Theorem 4.1. *If the bifurcation parameter $\epsilon > 0$ is small enough, for any $\omega_0 \in]0, \frac{1}{2}[$, there exists a (locally unique) complete trajectory $(g_n, \mu_n, R_n)_{n \in \mathbb{Z}}$ for the RG map such that $\lim_{n \rightarrow -\infty} (g_n, \mu_n, R_n) = (0, 0, 0)$ the Gaussian ultraviolet fixed point, and $\lim_{n \rightarrow +\infty} (g_n, \mu_n, R_n) = (g_*, \mu_*, R_*)$ the BMS nontrivial infrared fixed point, and determined by the ‘initial condition’ at unit scale $g_0 = \omega_0 \bar{g}_*$.*

5. Idea of The Proof

One writes $(g_n, \mu_n, R_n) = (\bar{g}_n + \delta g_n, \mu_n, R_n)$, so the trajectory equations are rephrased in terms of the deviation variables $(\delta g_n, \mu_n, R_n)$ relating the true trajectory to the approximate one. One has to enforce three boundary conditions: 1) μ_n does not blow up when $n \rightarrow +\infty$ (infrared), 2) R_n does not blow up when $n \rightarrow -\infty$ (ultraviolet), 3) $\delta g_0 = 0$ (unit scale). One iterates the flow equations either

backwards or forwards towards the appropriate boundary condition. For instance, the μ_n term is reexpressed by forward iteration as:

$$\mu_n = - \sum_{p \geq n} L^{-\left(\frac{3+\epsilon}{2}\right)(p-n+1)} \xi_\mu(\bar{g}_p + \delta g_p, \mu_p, R_p) \quad (2)$$

The R is iterated backwards. Finally the g_n is iterated backwards if $n > 0$, and forwards if $n < 0$. The four ‘integral’ equations, including Eq. 2, obtained in this way are interpreted as a fixed point equation in a big Banach space of two-sided sequences. The sought for trajectories are thus obtained using Banach’s contraction mapping theorem.

6. Functional Analysis, Norms

Let Δ denote a (closed) unit cell. There is a Sobolev imbedding $W^{4,2}(\overset{\circ}{\Delta}) \hookrightarrow C^2(\Delta)$. The fields ϕ live in Hilbert spaces $\text{Fld}(X) = \bigoplus_{\Delta \subset X} W^{4,2}(\overset{\circ}{\Delta})$ where extra C^2 gluing conditions are imposed, using the Sobolev imbedding. The norm on a field is given by $\|\phi\|_{\text{Fld}(X)}^2 = \sum_{\Delta \subset X} \sum_{|\nu| \leq 4} \|\partial^\nu \phi_\Delta\|_{L^2(\overset{\circ}{\Delta})}^2$. The fluctuation measure $d\mu_\Gamma$ is realized in the Hilbert spaces $\text{Fld}(X)$. One also needs a second field norm $\|\phi\|_{C^2(X)} = \sup_{x \in X} \max_{|\nu| \leq 2} |\partial^\nu \phi(x)|$. While the ϕ ’s are real-valued, the functionals can be complex valued and are compared using norms of the form

$$\begin{aligned} \|K\| = & \sup_{\Delta_0} \sum_{X \supset \Delta_0} L^{5|X|} \sup_{\phi \in \text{Fld}(X)} \left\{ e^{-\kappa \sum_{\Delta \subset X} \sum_{1 \leq |\nu| \leq 4} \|\partial^\nu \phi\|_{L^2(\overset{\circ}{\Delta})}^2} \right. \\ & \times \left. \sum_{0 \leq n \leq 9} \frac{(cg^{-\frac{1}{4}})^n}{n!} \sup_{\phi_1, \dots, \phi_n \in \text{Fld}(X) \setminus \{0\}} \frac{|D^n K(X, \phi; \phi_1, \dots, \phi_n)|}{\|\phi_1\|_{C^2(\Delta)} \cdots \|\phi_n\|_{C^2(\Delta)}} \right\} \end{aligned}$$

where $D^n K$ denotes the n -th order differential. One of the main technical difficulties, or fibered norm problem, is that these norms depend on the dynamical variable g . The way around it is to use the approximate solution $\bar{g}_n$ to calibrate them [1].

7. Perspectives

Many open problems in the continuation of this work remain. One should study more refined dynamical systems features of the BMS model such as the construction of the full heteroclinic invariant curve. One should study its regularity properties and smoothness at $g = 0$ (asked by K. Gawędzki). There is the problem of constructing the correlation functions for the primary field ϕ and also for composite fields and study possible anomalous dimensions (there are preliminary results on this by P. K. Mitter). A good testing ground for these more ambitious goals is the case of a hierarchical version of the BMS model (ongoing work by the author together with G. Guadagni and Ph. D. student A. Chandra). This could lead to the construction of the massless Euclidean BMS model over the p -adics. If the previous hurdles are

cleared, one should also study the analytic continuation to Minkowski space (in the Archimedean case). Another interesting project would be to provide an alternate construction of the critical BMS theories using recent phase space expansion techniques [21]. Finally the problem of controlling complete RG trajectories (from the UV to the IR end of the scale spectrum), here addressing the BMS model, is a worthy and certainly difficult task for many more QFT models: Gross-Neveu in $2d$, ϕ_3^4 at large N , noncommutative vulcanized ϕ_4^4 , the $2d$ σ -model, Yang-Mills in $4d$. The inspiring talk by S. Weinberg on QFT and the asymptotic safety scenario, at this congress, also leads one to wonder if one could devise a rigorous RG framework in order to study a nontrivial ultraviolet fixed point in quantum gravity...

Acknowledgments

The research presented in this article is supported by the National Science Foundation under grant # DMS-0907198.

References

- [1] A. Abdesselam, *Commun. Math. Phys.* **276**, 727 (2007).
- [2] P. M. Bleher and Ja. G. Sinai, *Commun. Math. Phys.* **33**, 23 (1973).
- [3] P. M. Bleher and Ya. G. Sinai, *Commun. Math. Phys.* **45**, 247 (1975).
- [4] P. Collet and J.-P. Eckmann, *Commun. Math. Phys.* **55**, 67 (1977).
- [5] P. Collet and J.-P. Eckmann, *A Renormalization Group Analysis of The Hierarchical Model in Statistical Mechanics*, Lect. Notes in Phys. **74** (Springer, Berlin–New York, 1978).
- [6] K. Gawędzki and A. Kupiainen, *Commun. Math. Phys.* **89**, 191 (1983).
- [7] K. Gawędzki and A. Kupiainen, *J. Stat. Phys.* **35**, 267 (1984).
- [8] E. A. Pereira, *J. Math. Phys.* **34**, 5770 (1993).
- [9] H. Koch and P. Wittwer, *Commun. Math. Phys.* **106**, 495 (1986).
- [10] G. Felder, *Commun. Math. Phys.* **111**, 101 (1987).
- [11] D. C. Brydges, J. Dimock and T. R. Hurd, *Commun. Math. Phys.* **198**, 111 (1998).
- [12] D. C. Brydges, P. K. Mitter and B. Scoppola, *Commun. Math. Phys.* **240**, 281 (2003).
- [13] P. C. Lima, *Commun. Math. Phys.* **170**, 529 (1995).
- [14] K. Gawędzki and A. Kupiainen, *Nuclear Phys. B* **262**, 33 (1985).
- [15] O. A. McBryan and J. Rosen, *Commun. Math. Phys.* **51**, 97 (1976).
- [16] J. Glimm and A. Jaffe, *Commun. Math. Phys.* **52**, 203 (1977).
- [17] D. C. Brydges, J. Fröhlich and A. D. Sokal, *Commun. Math. Phys.* **91**, 141 (1983).
- [18] G. Benfatto, P. Falco and V. Mastropietro, *Commun. Math. Phys.* **273**, 67 (2007).
- [19] D. C. Brydges and H. T. Yau, *Commun. Math. Phys.* **129**, 351 (1990).
- [20] V. Rivasseau, *From Perturbative to Constructive Renormalization* (Princeton University Press, Princeton NJ, 1991).
- [21] A. Abdesselam, *Renormalisation Constructive Explicite*, Ph. D. thesis (Ecole Polytechnique, Palaiseau, 1997).

GRADIENT MODELS WITH NON-CONVEX INTERACTIONS

STEFAN ADAMS

University of Warwick, Gibbet Hill Road,

Coventry, CV4 7AL, United Kingdom

E-mail: s.adams@warwick.ac.uk

http://www2.warwick.ac.uk/fac/sci/maths/people/staff/stefan_adams

We outline recent results in collaboration with R. Kotecký and S. Müller in [1] and [2] about gradient models on an integer lattice with non-convex interactions. These Gibbsian models (continuous Ising models) emerge in various branches of physics and mathematics, with a particular frequency in quantum field theory. Our attention is however mostly devoted to interfaces, of which a massless field is an *effective* modelisation, however the motivation stems considering vector valued fields as displacements for atoms of crystal structures and the study of the Cauchy-Born rule for these models. For the interface case we prove the strict convexity of the surface tension (free energy) for low enough temperatures and small enough tilts using multi-scale (renormalisation group analysis) techniques. This is the complementary study of the high temperature regime in [3] and it is an extension of Funaki and Spohn's result [4].

Keywords: Gradient models, massless field, Cauchy-Born rule

1. Introduction

We consider an effective model with gradient interaction. The model describes a phase separation in $\mathbb{R}^{d+1}$, e.g. between the liquid and vapor phase. For simplicity we consider a discrete basis $\Lambda \subset \mathbb{Z}^d$, and real-valued height variables

$$x \in \Lambda \mapsto \varphi(x) \in \mathbb{R}.$$

This model ignores overhangs like in Ising models, but gives a good approximation in the vicinity of the phase separation. The distribution of the interface is given in terms of a Gibbs distribution with nearest neighbour interactions of gradient type, that is, the interaction between neighboring sites $x, x + e_i$ depends only on the gradient $\nabla_i \varphi(x) = \varphi(x + e_i) - \varphi(x)$, $i = 1, \dots, d$. More precisely, the Hamiltonian is of the form

$$H_\Lambda(\varphi) = \sum_{x \in \Lambda} \sum_{i=1}^d W(\nabla_i \varphi(x)),$$

where $W: \mathbb{R} \rightarrow \mathbb{R}$ is a perturbation of a quadratic functions, i.e.

$$W(\eta) = \frac{1}{2}\eta^2 + V(\eta) \quad \text{with some perturbation } V: \mathbb{R} \rightarrow \mathbb{R}.$$

The Gibbs distribution for a given boundary condition $\Psi \in \mathbb{R}^{\partial\Lambda}$, where $\partial\Lambda = \{z \in \mathbb{Z}^d : |z - x| = 1 \text{ for some } x \in \Lambda\}$, at inverse temperature $\beta > 0$ is given by

$$\gamma_{\Lambda}^{\beta, \Psi}(\mathrm{d}\varphi) = \frac{1}{Z_{\Lambda}(\beta, \Psi)} \exp(-\beta H_{\Lambda}(\varphi)) \prod_{x \in \Lambda} \mathrm{d}\varphi(x) \prod_{x \in \partial\Lambda} \delta_{\Psi(x)}(\mathrm{d}\varphi(x)),$$

where the normalization constant $Z_{\Lambda}(\beta, \Psi)$ is the integral of the density and is called the partition function. One is particularly interested in tilted boundary conditions

$$\Psi_u(x) = \langle x, u \rangle, \quad \text{for some tilt } u \in \mathbb{R}^d.$$

An object of basic relevance in this context is the surface energy or free energy defined by the limit

$$\sigma(u) = \lim_{\Lambda \uparrow \mathbb{Z}^d} -\frac{1}{\beta|\Lambda|} \log Z_{\Lambda}(\beta, \Psi_u). \quad (1)$$

This surface tension $\sigma(u)$ can also be seen as the price to pay to tilt a totally flat interface. The existence of the above limit follows from a standard sub-additivity argument. In case of a *strictly* convex potential, Funaki and Spohn showed in [4] that σ is convex as a function of the tilt. The simplest strictly convex potential is the quadratic one with $V = 0$, which corresponds to a Gaussian model, also called the gradient free field or harmonic crystal. Models with non-quadratic potentials W are sometimes called anharmonic crystals. The convexity of the surface tension plays a crucial role in the derivation of the hydrodynamical limit of the Landau-Ginsburg model in [4]. Strict convexity of the surface tension for strictly convex W with $0 < c_1 \leq W'' \leq c_2 < \infty$, was proved in [5]. Under the assumption of the bounds of the second derivative of W , a large deviations principle for the rescaled profile with rate function given in terms of the integrated surface tension has been derived in [5]. Both papers [4] and [5] use explicitly the conditions on the second derivative of W in their proof. In particular they rely on the Brascamp-Lieb inequality and on the random walk representation of Helffer and Sjöstrand, which requires a strictly convex potential W .

In [3] Deuschel *et al* showed the strict convexity of the surface tension for non-convex potentials in the small β (high temperature) regime for potentials of the form

$$W(\eta) = W_0(\eta) + g(\eta),$$

where W_0 is strictly convex as above and where $g \in \mathcal{C}^2(\mathbb{R})$ has a negative bounded second derivative such that $\sqrt{\beta}\|g''\|_{L^1(\mathbb{R})}$ is sufficiently small.

In [1] we show the strict convexity of the surface tension for large enough β (low temperatures) and sufficiently small enough tilt using multi-scale techniques based on a finite range decomposition of the underlying background Gaussian measure in [2]. We will outline this in Section 3 below. In Section 2 we briefly describe the motivation for studying vector valued gradient fields for low temperatures and explain the Cauchy-Born rule. The result is stated in Section 3 with the main ideas of our proof techniques.

Finally note that, due to the gradient interaction, the Hamiltonian has a continuous symmetry. In particular this implies that no Gibbs measures on $\mathbb{Z}^d$ exists for dimensions $d = 1, 2$ where the field 'delocalises', cf. [6]. If one considers the corresponding random field of gradients (discrete gradient image of the height field φ) it is clear that its distribution depends on the gradient of the boundary condition of the height field. One can also introduce gradient Gibbs measures in terms of conditional distributions satisfying DLR equations, cf. [4]. Under the conditions of the bounds on the second derivative of W , Funaki and Spohn in [4] proved the existence and uniqueness of an extremal, i.e. ergodic, gradient Gibbs measure for each tilt $u \in \mathbb{R}^d$. In the case of non-convex W , uniqueness of the ergodic gradient component can be violated, for tilt $u = 0$ this has been proved in [7]. However in this phase transition situation in [7], the surface tension is not strictly convex at tilt $u = 0$.

2. Cauchy-Born rule

The Cauchy-Born rule for a piece of material is satisfied if the elementary cells of the crystal do follow any small enough affine displacement at the boundary (shear force). In a simple mass-spring model with nearest neighbour and next-nearest neighbour interactions between the displacement vectors (gradient model) existence and failure of the Cauchy-Born rule depending on the parameters (external force and spring constants) have been proved in [8]. In a gradient model setting it is possible to address this question for positive temperature. It is believed that for small external forces (displacements) and small enough temperature this rule is valid. We shall consider fields $\varphi: \mathbb{Z}^d \rightarrow \mathbb{R}^m$ for some $m > 1$ and study corresponding gradient models with non-convex interactions. The non-convexity of the interaction is crucial here as the frame indifference implies the non-convexity of the elastic energy. The first natural question is if the free energy is strictly convex for small forces and small temperatures. A stronger version of the Cauchy-Born rule for positive temperature is the unicity of the ergodic gradient Gibbs component for small forces and small temperatures. This latter question is supposed to be very hard, see the occurrence of a phase transition in [7]. In [1] the easier question for the free energy is addressed for $m = 1$, i.e. for the interface model case. In the vector case, $m > 1$, one needs to ensure the volume of the unit cells remain positive. This requires an additional treatment and is devoted to the future.

3. Result and outline of the proof

Following an idea in [4] we work on a torus $T_N := (\mathbb{Z}/L^N\mathbb{Z}^d)^d$ for some $L > 0$, i.e. we can represent the torus by the cube $\Lambda_N = \{x \in \mathbb{Z}^d: |x| \leq \frac{1}{2}(L^N - 1)\}$ once equipped with the torus distance. Hence we consider periodic height functions, and to kill the constant for the discrete gradient mapping we add a condition, i.e. we

consider the space of configurations

$$\Omega_N = \left\{ \varphi: \mathbb{Z}^d \rightarrow \mathbb{R}; \varphi(x+k) = \varphi(x) \forall k \in (L^N \mathbb{Z})^d; \sum_{x \in T_N} \varphi(x) = 0 \right\}.$$

We use λ_N to denote the $(L^N - 1)$ -dimensional Hausdorff measure on Ω_N . The partition function for given tilt $u \in \mathbb{R}$ is then

$$Z_N(\beta, u) = \int_{\Omega_N} \exp(-\beta H_N^u(\varphi)) \lambda_N(d\varphi), \quad (2)$$

with Hamiltonian

$$H_N^u(\varphi) = \frac{1}{2} L^{Nd} |u|^2 + \sum_{x \in \Lambda_N} \sum_{i=1}^d \frac{1}{2} (\nabla_i \varphi(x))^2 + V(\nabla_i \varphi(x)). \quad (3)$$

To state our main result, we need a condition on the smallness of the perturbation V . We will state it in terms of the Mayer function $K_{V,\beta,u}: \mathbb{R}^d \rightarrow \mathbb{R}$ associated with the function $V: \mathbb{R} \rightarrow \mathbb{R}$ determining the Hamiltonian H_N^u . Namely, we take

$$K_{V,\beta,u}(z) = \exp \left\{ -\beta \sum_{i=1}^d V\left(\frac{z_i}{\sqrt{\beta}} - u_i\right) \right\} - 1.$$

Given any $h > 0$, we then consider the Banach space E of functions $K: \mathbb{R}^d \rightarrow \mathbb{R}$ with the norm

$$\|K\|_h = \sup_{z \in \mathbb{R}^d} \sum_{|\alpha| \leq 11} h^{|\alpha|} |\partial_z^\alpha K(z)| e^{-h^{-2}|z|^2}.$$

Here, the sum is over nonnegative integer multiindices $\alpha = (\alpha_1, \dots, \alpha_d)$, $\alpha_i \in \mathbb{N}$, $i = 1, \dots, d$ with $|\alpha| = \sum_{i=1}^d \alpha_i \leq 11$, and $\partial^\alpha = \prod_{i=1}^d \partial_i^{\alpha_i}$. We also use $B_\delta(0) \subset \mathbb{R}^d$ to denote the ball $B_\delta(0) = \{u \in \mathbb{R}^d: |u| < \delta\}$.

Our main result is then:

Theorem 3.1 (Strict convexity of the free energy). *Let us assume that there exist $\delta > 0$, $\epsilon > 0$, $h > 0$, and $\beta_0 < \infty$ such that the map $\mathbb{R}^d \supset B_\delta(0) \ni u \mapsto K_{V,\beta,u} \in E$ is $\mathcal{C}^2$ and*

$$\|K_{V,\beta,u}\|_h + \sum_{i=1}^d \left\| \frac{\partial}{\partial u_i} K_{V,\beta,u} \right\|_h + \sum_{i,j=1}^d \left\| \frac{\partial^2}{\partial u_i \partial u_j} K_{V,\beta,u} \right\|_h \leq \epsilon$$

whenever $u \in B_\delta(0)$.

Then the free energy (surface tension) $\sigma(u) := -\lim_{N \rightarrow \infty} \frac{1}{\beta L^{dN}} \log Z_N(\beta, u)$ exists and it is uniformly strictly convex in u for $u \in B_\delta(0)$ and any $\beta \geq \beta_0$.

The proof employs a multi-scale analysis based on ideas going back to the work of Brydges and Yau in [9]. The proof in [1] follows closely the approach outlined by Brydges in [10]. The main ingredient is a finite range decomposition of families of Gaussian measures in [2]. The Hamiltonian in (3) shows that the partition function is a Gaussian integral with measure μ having density given by

$\exp(-\sum_{x \in \Lambda_N} \sum_{i=1}^d (\nabla_i \varphi(x))^2)$ and normalisation Z_N which is independent on the tilt and β (after some rescaling), that is

$$Z_N(\beta, u) = Z_N e^{-\frac{1}{2} L^{dN} |u|^2} \int_{\Omega_N} \exp\left(-\beta \sum_{x \in \Lambda_N} \sum_{i=1}^d V\left(\frac{1}{\sqrt{\beta}} \nabla_i \varphi(x) - u_i\right)\right) \mu(d\varphi). \quad (4)$$

The finite range decomposition in [2] of the Gaussian measure μ means that $\mu(d\varphi) = \mu_1 * \dots * \mu_{N+1}(d\varphi)$ where $\mu_1, \dots, \mu_{N+1}$ are Gaussian measures with a particular finite range property. Namely, the covariances of the measures $\mu_k, k = 1, \dots, N+1$, vanish for $|x| \geq \frac{1}{2} L^k$. Hence it is possible to perform the Gaussian integration in (4) in steps according to length scales given by $L^k, k = 1, \dots, N+1$. This defines roughly the renormalisation group mapping, eg. on scale L^k we get for an integrand F

$$(R_k F)(\varphi) = \int_{\Omega_N} F(\varphi + \xi) \mu_k(d\xi).$$

These mappings generate a dynamical system with an expanding and a contracting direction and trivial fix point. The crucial step is to get the correct initial Gaussian such that after N steps the mapping is close to the fix point. This enables one to control the second derivative with respect to the tilt of the integral in (4). To get the correct initial Gaussian measure we employ a change of Gaussian measure at the beginning with some quadratic form defined by a symmetric matrix $q \in \mathbb{R}^{d \times d}$. An application of the implicit function theorem will ensure the existence of a unique matrix q_0 for which the system reaches its fix point. However, the renormalisation mappings depend on these matrices and we have a loss of regularity once we compute derivatives of Gaussian expectations with respect to this parameter. This can be resolved with some version of the implicit function theorem which allows a loss of regularity, cf. [1].

References

- [1] S. Adams and R. Kotecký and S. Müller, *preprint* (2009).
- [2] S. Adams and R. Kotecký and S. Müller, *preprint* (2009).
- [3] C. Cotar and J.-D. Deuschel and S. Müller, *Communications in Mathematical Physics* **286**, 359-376 (2009).
- [4] T. Funaki and H. Spohn, *Communications in Mathematical Physics* **185**, 1-36 (2009).
- [5] J.-D. Deuschel and G. Giacomini and D. Ioffe, *Probability Theory and Related Fields* **117**, 49-111 (2000).
- [6] J. Fröhlich and C. Pfister, *Communications in Mathematical Physics* **81**, 277-298 (1981).
- [7] M. Biskup and R. Kotecký, *Probability Theory and Related Fields* **139**, 1-39 (2007).
- [8] G. Friesecke and F. Theil, *J. Nonlinear Sci.* **12**, 445-478 (2002).
- [9] D.C. Brydges and H.T. Yau, *Communications in Mathematical Physics* **129**, 351-392 (1990).
- [10] D. Brydges, in *Park City Lecture Notes* (2007).

SUPERSYMMETRIC WARD IDENTITIES IN QUANTUM DIFFUSION

MARGHERITA DISERTORI

*Laboratoire de Mathématiques Raphaël Salem, UMR CNRS 6085, Université de Rouen,
Avenue de l'Université, BP.12, 76801 Saint Etienne du Rouvray, France
E-mail: margherita.disertori@univ-rouen.fr*

This is a joint work with T. Spencer and M. Zirnbauer. We consider a lattice field model which qualitatively reflects the phenomenon of Anderson localization and delocalization for real symmetric band matrices. We prove [1] that in three or more dimensions the model has a ‘diffusive’ phase at low temperatures. Localization is expected at high temperatures. The classical tools to study this kind of problem are multiscale analysis (renormalization) together with cluster+Mayer expansions or convexity bounds. Our analysis uses instead estimates on non-uniformly elliptic Green’s functions and a family of Ward identities coming from internal supersymmetry.

Keywords: Anderson localization, quantum diffusion, Ward identities, renormalization

1. Introduction

It is well known that the study of localization properties in a disordered material can be translated in the study of correlation functions in some statistical mechanical model, with an internal non compact symmetry group [2–4]. In the physics literature one usually assumes the sigma model approximation, which is believed to capture the essential features of the energy correlations and transport properties of the underlying quantum system. These models are difficult to analyze with mathematical rigor in more than one dimension.

In this context Zirnbauer introduced a lattice field model which may be thought of as a simplified version of one of Efetov’s nonlinear sigma models [5, 6]. This model is expected to reflect the qualitative behavior of random band matrices – namely localization and diffusion – in any dimension. It is a vector model (instead of a matrix model). After integrating out the Grassmann variables the effective bosonic measure is **positive** (so probability arguments may be applied). Finally the model is supersymmetric (just like the real matrix model) but the symmetries are easier to exploit.

1.1. The model

Let $\Lambda \subset \mathbb{Z}^d$ be a cube. At each lattice point $j \in \Lambda$ we introduce a vector u_j with two bosonic components and three fermionic components.

$$u_j = (z_j, x_j, y_j, \xi_j, \eta_j) \quad (1)$$

where ξ, η are odd elements and z, x, y are even elements of a real Grassmann algebra. The scalar product is defined by

$$(u, u') = -zz' + xx' + yy' + \xi\eta' - \eta\xi', \quad (u, u) = -z^2 + x^2 + y^2 + 2\xi\eta. \quad (2)$$

The action is defined by

$$\mathcal{S}[u] = \frac{1}{2} \sum_{\langle j, j' \rangle} \beta(u_j - u_{j'}, u_j - u_{j'}) + \sum_j \varepsilon(z_j - 1), \quad (3)$$

where $\langle j, j' \rangle$ are nearest neighbor pairs, $\beta \geq 0$ and $\varepsilon \geq 0$. Since this model should mimic the sigma model for a band matrix problem we add a sigma model constraint

$$(u_j, u_j) = -1 \quad \forall j \in \Lambda \Rightarrow z_j = \pm \sqrt{1 + x_j^2 + y_j^2 + 2\xi_j\eta_j}. \quad (4)$$

We choose the manifold corresponding to the $+$ sign, so now in the action $z_j - 1 \geq 0$ for each j . Note that z is not really a number, but rather an even element of a Grassmann algebra so it may have fermionic components (as in (4) above).

Now, passing to horospherical coordinates and integrating out the two Grassmann variables and one bosonic variable we end up with an effective bosonic field theory defined by

$$d\mu(t) = \left[\prod_j \frac{dt_j}{\sqrt{2\pi}} e^{-t_j} \right] e^{-\mathcal{B}(t)} \det^{1/2}[M(t)], \quad (5)$$

where $t_j \in \mathbb{R}$, $j \in \Lambda$,

$$\begin{aligned} \mathcal{B}(t) &= \sum_{\langle j, j' \rangle} \beta (\cosh(t_j - t_{j'}) - 1) + \sum_{j \in \Lambda} \varepsilon (\cosh t_j - 1) \\ &= (t, (-\beta\Delta + \varepsilon)t) + \text{higher order terms}, \end{aligned} \quad (6)$$

and $M(t) > 0$ is a positive quadratic form

$$(f, M(t) f) = \sum_{\langle j, j' \rangle} \beta e^{t_j + t_{j'}} (f_j - f_{j'})^2 + \sum_j \varepsilon e^{t_j} f_j^2 \quad \forall f_j \in \mathbb{R}.$$

Remark 1. The measure $d\mu(t)$ is real, positive and normalized to one (by supersymmetry), so we can apply probability arguments.

Remark 2. The factor $\mathcal{B}(t)$ (6) in the exponent is a gaussian term plus corrections. This can be treated by standard techniques (renormalization group, multiscale analysis). On the other hand the determinant $\det^{1/2}[M(t)]$ introduces highly non local interactions that make the usual techniques hard to apply. Moreover the action $\mathcal{B}(t) - \ln \det^{1/2}[M(t)]$ is non convex.

2. Main results

For the model defined above localization was established in one dimension [5] in the sense that the conductance was proved to decay exponentially in the system size. In three or more dimensions a transition is expected from a localized phase for β small to a diffusive phase for β large. Numerical simulations in 3d [7] show that the critical parameter is small $\beta_c \simeq 0.038$. In [1] we prove

Theorem 2.1. *In dimension $d \geq 3$, for $\beta \gg 1$, the t field does not fluctuate. More precisely*

$$\langle (\cosh t_x - t_y)^m \rangle \leq 2 \quad \forall x, y \in \Lambda \quad (7)$$

for $0 \leq m \leq \beta^{1/8}$ uniformly in the volume Λ and ε .

Theorem 2.2. *In dimension $d \geq 3$, for $\beta \gg 1$, the t field remains bounded. More precisely*

$$\langle (\cosh t_x)^p \rangle \leq 4 \quad \forall x \in \Lambda \quad (8)$$

for $0 \leq p \leq 10$ uniformly in the volume Λ and ε as long as $\varepsilon \geq \frac{1}{|\Lambda|^{1-\alpha}}$, $\alpha > \frac{1}{\ln \beta}$.

These results imply the existence of a 'diffusive' phase in a weak sense (see [1]). In the next section we sketch the proof of (7). It can be resumed in three steps.

Step 1: bound on nearest neighbor fluctuations. Using a Ward identity from the internal supersymmetry we prove (7) for nearest neighbor pairs $|x - y| = 1$.

Step 2: conditional bound on large scale fluctuations. Using again a Ward identity from the internal supersymmetry we prove (7) for $|x - y| > 1$, but only adding constraints on scales shorter than $|x - y|$.

Step 3: unconditional bound on large scale fluctuations. In order to remove the constraints and prove (7) for any pair x, y , we use the previous two results and an induction on scales.

3. Outline of the proof

3.1. Ward identities

The internal supersymmetries of (3) imply a large number of Ward identities. Here we are interested only in one namely

$$1 \geq \left\langle \cosh^m(t_x - t_y) \left(1 - \frac{m}{\beta} C_{xy} \right) \right\rangle, \quad (9)$$

where the number $C_{xy} > 0$ is defined as

$$C_{xy} = \left(\delta_x - \delta_y, \frac{1}{D(t, x, y)} \delta_x - \delta_y \right), \quad (10)$$

the (positive) matrix $D(t, x, y)$ (that depends on the t configuration and on the points x, y) is defined as

$$(f, D f) = \sum_{\langle j, j' \rangle} (f_j - f_{j'})^2 A_{xy}(jj') , \quad (11)$$

and $A_{xy}(jj')$ can be thought as a “local conductance” through the edge jj'

$$A_{xy}(jj') = e^{t_j + t_{j'} - t_x - t_y} \cosh(t_x - t_y) > 0 .$$

The inverse of D inside C_{xy} is well defined since we are projecting out the zero mode. The local conductance A depends only on fluctuations of the field t (and not on the value of the field itself). This is a key point to prove the theorem.

3.2. NN bounds

For $|x - y| = 1$ it is not difficult to see that $C_{xy} \leq 1$ **for all t configurations**. Inserting this in (9) we have

$$1 = \left\langle \cosh^m(t_x - t_y) \left(1 - \frac{m}{\beta} C_{xy}\right) \right\rangle \geq \langle \cosh^m(t_x - t_y) \rangle \left(1 - \frac{m}{\beta}\right) \quad (12)$$

$$\Rightarrow \langle \cosh^m(t_x - t_y) \rangle \leq \frac{1}{1 - \frac{m}{\beta}}$$

as long as $m/\beta < 1$. Since β is large the power m can be large too.

3.3. Conditional bounds

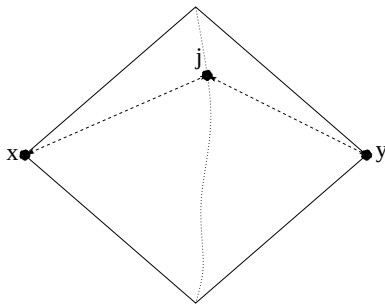
For $|x - y| > 1$, (9) is still true but this time there is no uniform bound on C_{xy} since $A_{xy}(jj')$ can be arbitrarily small. On the other hand, C_{xy} is bounded if we add constraints on the field fluctuations at lower scales. More precisely $C_{xy} \leq \text{const}$ uniformly in x, y if

$$A_{xy}(jj') \geq \frac{1}{|j - x|^\alpha} + \frac{1}{|j - y|^\alpha} \quad \forall j, j' \in R_{xy} , \alpha < 1 ,$$

where the region R_{xy} is a three dimensional diamond (see Figure 1). This result holds **only in dimension three or higher**. Note that we need to impose constraints only inside the region R_{xy} (and not everywhere) and $A_{xy}(j, j')$ may become very small far from x, y . This will be important when we will remove the constraints (step 3 of the proof). Inserting this in (9) we have

$$1 \geq \left\langle \cosh^m(t_x - t_y) \bar{\chi}_{xy} \left(1 - \frac{m}{\beta} C_{xy}\right) \right\rangle \geq \langle \cosh^m(t_x - t_y) \bar{\chi}_{xy} \rangle \left(1 - \frac{m}{\beta}\right) . \quad (13)$$

The second inequality is true because we have constraints on lower scale fluctuations, but the first is false in general since $\bar{\chi}_{xy}$ is not a supersymmetric expression. Actually it is possible to add the constraint respecting supersymmetry so the equation above is correct.

Fig. 1. the region R_{xy} .

3.4. Unconditional bounds

Finally, to remove the constraint $\bar{\chi}_{xy}$ we write

$$\langle \cosh^m(t_x - t_y) \rangle = \langle \cosh^m(t_x - t_y) \bar{\chi}_{xy} \rangle + \text{remainder} , \quad (14)$$

where the remainder is

$$\text{remainder} = \langle \cosh^m(t_x - t_y) \bar{\chi}_{xy}^c \rangle .$$

To complete the proof we show that the remainder is small by an induction on scales.

4. Conclusive remarks

Using SUSY Ward identities plus an induction on scales we obtain bounds without using any multiscale analysis or cluster expansion. Moreover, this method gives information in the extended states region.

The main difficulties to generalize this technique to the real band matrix model are a more complicated fermionic term and an affective measure which is no longer real positive.

References

- [1] M. Disertori, T. Spencer and M. Zirnbauer arXiv:0901.1652.
- [2] F. Wegner and L. Schaefer, *Z. Phys. B* **38**, 113 (1980).
- [3] K. B. Efetov, *Adv. Phys.* **32**, 874 (1983).
- [4] K. B. Efetov, *Supersymmetry in disorder and chaos* (Cambridge University Press, Cambridge, 1997).
- [5] M. Zirnbauer, *Commun. Math. Phys.* **141**, 503 (1991).
- [6] W. Drunk, D. Fuchs and M. Zirnbauer, *Ann. Physik* **1**, 134 (1992).
- [7] T. Dupré, *Physical Review B* **54**, 2763 (1996).

OFF-CRITICAL LATTICE MODELS AND MASSIVE SLEs

NIKOLAI MAKAROV

*Mathematics 253-37, Caltech
Pasadena, CA 91125, USA
E-mail: makarov@caltech.edu*

STANISLAV SMIRNOV

*Section de mathématiques, Université de Genève
2-4, rue du Lièvre, c.p. 64, 1211 Genève 4, Switzerland
E-mail: stanislav.smirnov@math.unige.ch*

We suggest how versions of Schramm's SLE can be used to describe the scaling limit of some off-critical 2D lattice models. Many open questions remain.

Keywords: lattice models, SLE, CFT

1. Introduction

During the last 25 years Conformal Field Theory (CFT) was successful in heuristically describing conformally invariant scaling limits of 2D lattice models at criticality, such as the Ising model, percolation, Self-Avoiding Polymers, Potts models, — see Ref. 1 for a collection of the founding papers of the subject.

Recently there was much progress in the mathematical understanding, in large part due to Oded Schramm's introduction of SLEs, or *Schramm Loewner Evolutions*. $\text{SLE}(\kappa)$ is a one-parameter family of random conformally invariant curves, constructed by running a Loewner Evolution with (real valued) Brownian motion as the driving term. For several models convergence to SLE was established in the scaling limit; moreover, SLE is well-adapted to calculations, which typically boil down to Itô's stochastic calculus. See Ref. 2, 3 for an exposition and references.

The key property of SLE is its conformal invariance, which is expected in 2D lattice models only at criticality, and the question naturally arises:

Can SLE success be replicated for off-critical models?

In most off-critical cases to obtain a non-trivial scaling limit one has to adjust some parameter (like temperature in the Ising model or probability of an open site in percolation), sending it at an appropriate speed to the critical value. Such limits lead to massive field theories, so the question can be reformulated as whether one can use SLEs to describe those. Massive CFTs are no longer conformally invariant, but are still covariant when mass is considered as a variable covariant density.

So far there was limited progress on off-critical SLEs, related work limited to Ref. 4–8. Below we propose an approach based on the combination of SLE *Martingale Observables* (MO) with potential theory and stochastic analysis, and we start by describing the critical case.

1.1. SLE and critical lattice models

Suppose that a critical lattice model defines for every simply connected domain Ω with two marked boundary points a, b a random discrete simple curve joining them inside Ω . One can think e.g. of a domain wall boundary in the Ising model with Dobrushin boundary conditions or of a LERW – the Loop Erasure of the Random Walk from a to $\partial\Omega \setminus \{b\}$ conditioned on ending at b .

The key observation made by Oded Schramm was that if a conformally invariant scaling limit γ of the discrete curves exists and satisfies Markov property (i.e. the curve progressively drawn from a to b does not distinguish its past from the boundary of the domain Ω), then it can be described by SLE(κ) for some $\kappa \in [0, \infty[$. Take an appropriate time parameterization $\gamma(t)$ and denote by Ω_t the component at b of the domain $\Omega \setminus \gamma[0, t]$. Then the random Loewner conformal map

$$Z_t(z) : \Omega_t \rightarrow \mathbb{C}_+, \quad \gamma(t) \mapsto 0, \quad b \mapsto \infty, \quad \text{normalized at } b$$

(in what we call a *Loewner chart*), satisfies the *Loewner equation* with the Brownian Motion as the driving term. We write it as a *stochastic differential equation* (SDE)

$$dZ_t(z) = \frac{2}{Z_t(z)} dt - d\xi_t, \quad d\xi_t := \sqrt{\kappa} dB_t, \quad (1)$$

with the SLE *driving term* ξ_t given by the standard Brownian motion B_t .

Moreover, it turns out that to deduce SLE convergence it is enough to show that just one *observable* $M_t(z) = M(z, \Omega_t)$ (e.g. spin correlation or percolation probability) is conformally covariant and Markov in the limit, see Ref. 3 for a discussion. So far conformal invariance of observables was always established by showing that they are discrete holomorphic (or harmonic) functions of a point satisfying some *Boundary Value Problem* (BVP) – Dirichlet, Neumann or Riemann-Hilbert.

The covariant holomorphic MOs for SLE can be easily classified, leading to a one parameter family for each κ or each *spin* (conformal dimension) σ , namely

$$M_t^{\kappa, \sigma, \beta}(z) = M_t(z) = Z_t(z)^\beta Z'_t(z)^\sigma, \quad \text{with } \sigma = \beta + \frac{\beta(\beta-1)}{4}\kappa, \quad (2)$$

is $(dz)^\sigma$ -covariant. It is characterized (see Ref. 3) by the Riemann-Hilbert BVP

$$\bar{\partial} M_t = 0 \text{ in } \Omega_t, \quad M_t(z) \parallel \tau^{-\sigma} \text{ on } \partial\Omega_t, \quad \text{appropriate singularities at } \gamma(t), b. \quad (3)$$

Note that the MO above can also be rewritten in terms of the Schwarz (complex Poisson) kernel in Ω_t at $\gamma(t)$, $P_t(z) := -1/Z_t(z)$.

Of special interest is also the family of holomorphic MOs for SLE(κ):

$$M_t^\kappa(z) = M_t(z) = \log Z_t(z) + \left(1 - \frac{\kappa}{4}\right) \log Z'_t(z). \quad (4)$$

Those are covariant pre-pre-Schwarzian forms, and so it suffices to study their imaginary parts – harmonic MOs characterized by

$$\Delta M_t(z) = 0 \text{ in } \Omega, \quad M_t(z) = \arg Z_t(z) + \left(1 - \frac{\kappa}{4}\right) \arg Z'_t(z) \text{ on } \partial\Omega_t, \quad (5)$$

the Dirichlet BVP being well posed and independent of time parameterization, i.e. the choice of normalization at b of the map $Z_t(z)$. Such *bosonic* observables (in Coulomb gas formalism of CFT, they are 1-point functions of the bosonic field in the presence of background charge) feature prominently in Ref. 9–11.

The observables above have the martingale property with respect to corresponding SLE(κ) by a simple application of Itô's calculus: vanishing of their drifts under the diffusion (1) is easy to deduce using (1) and its corollary

$$dZ'_t(z) = -\frac{2Z'_t(z)}{Z_t(z)^2}dt. \quad (6)$$

Conversely, if a random curve admits a MO of the mentioned form, similar calculations prove the curve to be a SLE(κ).

1.2. SLE and off-critical perturbations

Most discrete holomorphic observables studied in Ref. 10–16 behave well under some off-critical perturbations, leading to discrete *Massive Martingale Observable* (MMO) satisfying the massive version of the Cauchy-Riemann or Laplace equations:

$$\bar{\partial}^{(m)} M^{(m)} := \bar{\partial} M^{(m)} - im \overline{M^{(m)}} = 0, \quad (7)$$

$$\Delta^{(m)} M^{(m)} := \Delta M^{(m)} - m^2 M^{(m)} = 0, \quad (8)$$

inside Ω and solving the same boundary value problem as the critical MO. Here mass m can be understood as a function of z , changing covariantly under conformal transformations. In discrete setting a power of the lattice mesh enters Eq. (7,8), thus one has to tend the perturbation parameter to its critical value in a coordinated way with the lattice mesh, so that the mass does not blow up in the scaling limit.

Given a discrete random curve with a discrete MMO, we can ask, to what extent the SLE theory can be applied in the massive case, posing the following problems:

Problem A. *Show that discrete MMO has a scaling limit, which is then a MMO for some random curve. Find a SDE for the driving diffusion $d\xi_t$ replacing the Brownian motion in the Loewner Evolution (1).*

Problem B. *Show that this SDE has a unique solution and the corresponding random curve is the scaling limit of the original discrete curves.*

Problem C. *Use massive SLEs to derive properties of massive field theories and massive SLE curves.*

Note that massive models (as well as massive holomorphic functions) usually are even easier to control, so the main problem is the absence of conformal invariance,

or rather presence of conformal covariance with respect to the mass. Consequently the drift terms in the corresponding diffusions depend on (Euclidean) geometry of the domains constructed dynamically by Loewner evolution, leading to SDEs with general previsible path functionals, which are rather complicated.

Outline of the paper. We were able to advance within the suggested framework, and in Section 2 we present some of our results from the forthcoming Ref. 8. In Section 3 we provide two essential elements of the proof, namely we give an example of the drift computation (in the case of bosonic observables) and establish some simple a priori estimates of the drift (in the $\kappa = 4$ bosonic case). Finally we conclude with a list of open questions in Section 4.

Notation. We consider a simply connected domain $\Omega \subset \mathbb{C}$ with two marked points $a, b \in \partial\Omega$, joined by a random Markov curve $\gamma(t)$ with Loewner parameterization. The Loewner map to the half-plane is denoted by $Z_t(z)$. We denote by h the harmonic measure, by G the Green's function, and its boundary differentials by

$$P(u, z) = N_u G(\cdot, z), \quad K(u, b) = N_b P(u, \cdot)$$

(the Poisson kernel and Poisson boundary kernel respectively). Here N_u stands for the normal derivative at u in the Loewner boundary chart. By $G^{(m)}, P^{(m)}, \dots$ we denote the corresponding massive objects. The index t signifies that we work in the domain Ω_t , i.e. the component at b of $\Omega \setminus \gamma[0, t]$, and $\iint$ denotes the area integral.

2. Results

Below we go through all the cases discovered to-date where a lattice model admits a discrete MO with a massive perturbation, and describe results which will appear in Ref. 8. For simplicity we restrict ourselves to the case of constant mass m , but the methods should apply to an appropriate class of variable densities $m(z)$.

2.1. Loop Erased Random Walk with killing, $\kappa = 2$

The LERW converges to SLE(2), as shown in Ref. 12. The law $\gamma(t)$ of the LERW^(m) on a lattice of mesh ϵ is defined by applying the same loop erasing procedure to the random walk with a killing rate (i.e. probability to die out at each step)

$$\delta = m^2 \epsilon^2.$$

The RW is done inside a domain Ω from the boundary point a to the rest of the boundary and conditioned on ending at $b \in \partial\Omega$ – this ensures the Markov property.

Theorem 2.1. *For a bounded domain the scaling limit of LERW^(m) exists and is given by the massive SLE(2) with the driving diffusion*

$$\xi_t = \sqrt{2}dB_t + \lambda_t dt, \quad \lambda_t = 2 \left[\log K^{(m)}(\cdot, b; \Omega_t) \right]' (\gamma_t) \quad (\text{in the chart } Z_t). \quad (9)$$

The law of the scaling limit is absolutely continuous with respect to SLE(2).

The proof begins by constructing a discrete harmonic MMO (on appropriate discretization of the domain Ω):

$$M_t^{(\delta)}(z) \equiv M^{(\delta)}(z; \gamma(t), b, \Omega_t) := \frac{G^{(\delta)}(\gamma(t), z; \partial\Omega \cap \gamma[0, t])}{h^{(\delta)}(\gamma(t), b; \partial\Omega \cap \gamma[0, t])}, \quad z \in \Omega_t^{\text{discrete}},$$

where $G^{(\delta)}$ and $h^{(\delta)}$ denote discrete Green's function and discrete harmonic measure respectively, both with killing rate δ . Then we fix a boundary chart at b and show that after an appropriate normalizations the discrete MMO converges to

$$\frac{P_t^{(m)}(\gamma(t), z)}{K_t^{(m)}(\gamma(t), b)}, \quad (z \in \Omega_t), \quad (10)$$

a continuous MMO corresponding to (2,3) with $\beta = -1$ and $\sigma = 0$. Now let the driving term ξ_t be an Itô process

$$d\xi_t = \lambda(t, \omega) dt + \sigma(t, \omega) dB_t.$$

We claim that if the Loewner chain has MMO (10), then $\sigma \equiv \sqrt{2}$ and λ satisfies (9). We now have the equation

$$d\xi_t = \lambda(t, \xi_\bullet) dt + \sqrt{2} dB_t,$$

with the *previsible* path functional $\lambda(t, \xi_\bullet)$ given by (9). We claim that this equation satisfies the standard finiteness condition (cf. Chapter 5 in Ref. 18): λ is locally integrable on almost all paths. The question of course arises of existence and uniqueness of SDE solutions. In the case under consideration the answer is quite simple: the path functional λ satisfies the Novikov's condition (see Ref. 19)

$$\mathbb{E} \left[\exp \left(\frac{1}{2} \int_0^\infty \lambda(t, \xi_\bullet)^2 dt \right) \right] < \infty, \quad (11)$$

(and in fact, $\int_0^\infty \lambda(t, \xi_\bullet)^2 dt \leq \text{const} < \infty$ holds for all paths). In particular, the SDE has a unique in law (weak) solution ξ_t and its law is absolutely continuous with respect to $\sqrt{2}dB_t$. See Proposition 3.2 below for a similar argument. One can then deduce that the massive SLE is the scaling limit of the massive LERW by using the absolute continuity of latter with respect to the LERW.

2.2. Massive Harmonic Explorer and Gaussian Free Field, $\kappa = 4$

We use the original Harmonic Explorer (HE) construction on a hexagonal lattice, Ref. 10. At each step, the boundary consists of two arcs, and the explorer turns in the direction of one of the arcs with probability equal to its harmonic measure in the current domain Ω_t evaluated at the “growth point.” Introducing the killing rate as in the previous section, we get one distinction from the massless case – the two harmonic measures don't sum up to one, so with complementary probability we toss a fair coin to determine the direction of the turn. The MMO is the massive version of the bosonic MO (4,5) for $\kappa = 4$.

The theory is completely analogous to that of the massive LERW. If the initial domain is bounded, then the scaling limit exists and is absolutely continuous with respect to SLE(4), the scaling limit of massless explorer. We deduce the formula (13) for the drift in more general bosonic case in Proposition 3.1 (for the HE M_t is the difference of the *massless* harmonic measures of the two boundary arcs).

The same framework holds for the massive version of the discrete Gaussian Free Field discussed in Ref. 11, and parts of our construction apply to general bosonic observables.

2.3. Massive Peano curves, $\kappa = 8$

For a lattice approximation of Ω , we choose its cycle-free subgraph Γ with probability proportional to $\alpha^{n(\Gamma)}$, $n(\Gamma)$ being the number of connected components. Conditioning Γ to contain all the edges in the boundary arc (a, b) creates a random Markov interface γ from a to b which traces the component wired on the arc (a, b) . The case $\alpha = 0$ corresponds to the usual Uniform Spanning Tree model, as considered in Ref. 12, whose interface converges to the random Peano curve SLE(8).

The full massive harmonic measure of the boundary with reflection in the unwired part is a discrete MMO, corresponding to (2,3) with $\beta = 1/2$ and $\sigma = 0$. It has a scaling limit, and we show that the drift has to be

$$d\lambda_t = 16 \left(\iint_{\Omega_t} \tilde{P}_t \tilde{P}_t^{(m)} \right) dt, \quad (12)$$

where $\tilde{P}_t$ is the minimal (Martin's) kernel for Dirichlet/Neumann boundary conditions in $(\Omega_t, \gamma(t), b)$ and $\tilde{P}_t^{(m)}$ its massive counterpart in the Loewner chart.

It would be interesting to interpret the formal expression (12) and show that the SDE is well defined. Note that for $\alpha > 0$ the interface is no longer space-filling, thus massive SLE and SLE are mutually singular, while both have scaling dimension 2.

2.4. Fortuin-Kasteleyn Ising model, $\kappa = 16/3$

The *fermionic* MO, considered in Ref. 3, 13–15, 17 for the random cluster representation of the critical Ising model, implies that the interface converges to SLE(16/3). The MO corresponds to (2,3) with $\beta = -1/2$ and $\sigma = 1/2$, and becomes a MMO under the perturbation by p – the weight of an open edge (FK analogue of magnetization). The techniques of Ref. 15 allow to show the existence of a scaling limit MMO. It solves a Riemann-Hilbert boundary value problem, which makes potential theory and hence derivation of drifts more difficult than in the bosonic case.

2.5. Critical Ising model, $\kappa = 3$

A similar fermionic MO appears (see Ref. 3, 16) in the usual spin representation of the Ising model and implies convergence of the interface (domain wall boundary) to SLE(3). This is observable (2,3) with $\beta = -1$ and $\sigma = 1/2$. When perturbed by the

energy field, it becomes a MMO, which again solves a Riemann-Hilbert boundary value problem. While the techniques of Ref. 15 should also be applicable, there are similar difficulties as well.

3. Techniques

As an example, we show how to derive the drift and analyze the corresponding SDE for bosonic MOs (4). We work with their imaginary parts, harmonic MOs (5). The corresponding MMOs are massive harmonic (8), while solving the same BVPs.

3.1. Deriving the drift

Proposition 3.1. *If a random curve is described by a Loewner evolution and has a bosonic MMO satisfying the above BVP, then the driving diffusion is given by*

$$d\xi_t = \sqrt{\kappa}dB_t + d\lambda_t, \quad d\lambda_t = \left(\iint_{\Omega_t} m^2 M_t P_t^{(m)} \right) dt. \quad (13)$$

Remark 3.1. The drift can be rewritten in several ways, e.g. as $\iint_{\Omega_t} m^2 M_t^{(m)} P_t$.

Proof. Let $M_t^{(m)}$ be the massive harmonic MMO solving the said BVP, then

$$\Delta^{(m)} M_t^{(m)} = 0 \Rightarrow \Delta^{(m)} (M_t^{(m)} - M_t) = m^2 M_t \Rightarrow M_t^{(m)} = M_t + m^2 M_t * G_t^{(m)}, \quad (14)$$

Similarly for the massive Poisson kernel we have

$$P_t^{(m)} = P_t + m^2 P_t * G_t^{(m)}. \quad (15)$$

Describe the massive curve by the Loewner evolution with the driving term ξ_t

$$d\xi_t = \sqrt{\kappa}dB_t + d\lambda_t.$$

Denoting drifts with respect to SLE and massive SLE by d^{sle} and d^{msle} , we calculate

$$d^{msle} M_t = d^{sle} M_t + \text{Im}(d\lambda_t/Z_t) = P_t d\lambda_t. \quad (16)$$

Using the massive version of the Hadamard's variational formula

$$d^{msle} G_t^{(m)}(z, w) = -P_t^{(m)}(z) P_t^{(m)}(w) dt, \quad (17)$$

we can write the drift for the MMO:

$$\begin{aligned} 0 &= d^{msle} M_t^{(m)}(z) \stackrel{(14)}{=} d^{msle} \left(M_t + m^2 M_t * G_t^{(m)} \right) \\ &= d^{msle} M_t + m^2 (d^{msle} M_t) * G_t^{(m)} + m^2 M_t(\cdot) * (d^{msle} G_t^{(m)}(z, \cdot)) \\ &\stackrel{(16,17)}{=} \left(P_t + m^2 P_t * G_t^{(m)} \right) d\lambda_t - m^2 M_t(\cdot) * (P_t^{(m)}(z) P_t^{(m)}(\cdot)) dt \\ &\stackrel{(15)}{=} P_t^{(m)}(z) d\lambda_t - P_t^{(m)}(z) \left(\iint_{\Omega_t} m^2 M_t P_t^{(m)} \right) dt, \end{aligned}$$

so we deduce (13). □

3.2. Analysis of diffusion

Proposition 3.2. *In bounded domain Ω the massive SLE(4) driven by the diffusion (13) for $\kappa = 4$ is absolutely continuous with respect to SLE(4).*

Proof. First recall that in this case the harmonic MMO is bounded by the maximum principle since $M_t^{(m)} = \arg Z_t \in [0, \pi]$ on $\partial\Omega_t$ by the BVP (5). Thus by Remark 3.1

$$|\lambda_t| = \iint_{\Omega_t} m^2 M_t^m P_t \lesssim \iint_{\Omega_t} P_t. \quad (18)$$

Now for a bounded initial domain we can write using the Hadamard's variational formula (17):

$$\begin{aligned} \int_0^\infty \lambda_t^2 dt &\stackrel{(18)}{\leq} \int_0^\infty dt \iint_{(z)} P_t(z) \iint_{(w)} P_t(w) \stackrel{(17)}{=} - \int_0^\infty \iint_{(z)} \iint_{(w)} dG_t(z, w) \\ &= \iint_{(z)} \iint_{(w)} [G_0(z, w) - G_\infty(z, w)] \leq \iint_{(w)} \iint_{(z)} G_0(z, w) \leq C < \infty. \end{aligned}$$

The last step follows since Ω_0 is bounded and so

$$G_{\Omega_0}(z, w) \leq -\log |z - w| + \text{const.}$$

Applying the Novikov's criterion (11), we conclude that our diffusion is well defined and its law is absolutely continuous with respect to SLE(4). $\square$

4. Questions

We would like to end with a few questions, which originated in our work.

Question 4.1. How many “physically interesting” (e.g. relevant for the renormalization group) massive perturbations can a CFT have? The arguments above display a one-parameter family of covariant holomorphic MO for each SLE(κ), suggesting that somehow there is a one-parameter family of “canonical” massive perturbations. Are perturbations by non-holomorphic MO also relevant? Are all perturbations generated by MO perturbations?

Question 4.2. Is it always true that diffusion driving a massive version of SLE(κ) is a speed κ Brownian motion plus a drift? In the absence of conformal invariance the drift forcibly depends on the geometry of the domain grown. Are all such drifts locally bounded variation path functionals?

Question 4.3. The drifts so far encountered are either absolutely continuous (e.g. LERW and HE cases above) or monotone (e.g. UST case above or percolation perturbations discussed in Ref. 6, 7). Are all the possible drifts combinations of those?

Question 4.4. Can one show that a SLE and its massive version are always in the same universality class (in the sense of scaling exponents)?

Question 4.5. For $\kappa \leq 4$ in our examples the massive and usual SLE are mutually absolutely continuous. Is it true for all perturbations when $\kappa \leq 4$?

Question 4.6. Is it true that for $\kappa > 4$ the massive and usual SLE are singular, while being in the same universality class? If not, for which perturbations are they mutually absolutely continuous? We expect the bosonic ones to be among those.

Question 4.7. $\text{SLE}(\kappa)$ almost surely produce simple curves for $\kappa \leq 4$ and curves with double points for $\kappa > 4$. “Resampling” the model at a double point alters the curve drastically. Is this the reason for the absolute continuity / singularity dichotomy suggested above?

Question 4.8. Does discrete percolation observable from Ref. 21 have a massive counterpart? If so, which perturbation does it correspond to? This observable requires three marked points, but when two are fused, its continuous counterpart becomes the observable (2,3) with $\beta = 1/3$ and $\sigma = 0$.

Question 4.9. We have MMOs for at least five different values of κ . Guess by analogy the driving diffusions for other values of κ , and show that those are well-defined and lead to random curves.

Question 4.10. In particular, FK Ising MMO from section 2.4 belongs to a family of MMOs arising from (2,3) with $\sigma = -\beta = -1 + 8/\kappa$, cf. Ref. 3. We believe that all those correspond to magnetization perturbations. Is it true, in particular for $\kappa = 3$ (spin Ising)? If it holds for $\kappa = 6$, can one make a connection with a heuristic formula $d\lambda_t = \text{const}|d\gamma(t)|^{3/4}|dt|^{1/2}$ for the drift from Ref. 7?

Question 4.11. Similarly, starting from the spin Ising MMO from section 2.5, we can ask (cf. Ref. 3) whether all MMOs arising from (2,3) with $\sigma = -\beta/2 = 3/\kappa - 1/2$ correspond to fugacity perturbations? Note that while LERW MMO from section 2.1 does not belong to this family, its differential does.

Question 4.12. The $O(N)$ model has conjecturally two critical regimes, corresponding to parameter x (bond weight in the loop representation of the high temperature expansion) belonging to $\{x_c\}$ and $]x_c, \infty[$, see Ref. 20. Interfaces conjecturally converge to $\text{SLE}(\kappa)$ and $\text{SLE}(\tilde{\kappa})$ correspondingly, where κ and $\tilde{\kappa}$ are known functions of N and satisfy duality (though different from the usual duality $\kappa\kappa^* = 16$):

$$\frac{1}{\kappa} + \frac{1}{\tilde{\kappa}} = \frac{1}{2}, \quad \kappa \in [8/3, 4], \quad \tilde{\kappa} \in [4, 8]. \quad (19)$$

As in question 4.11 we expect to have in the first regime a discrete holomorphic MO corresponding to (2,3) with $\sigma = -\beta/2 = \frac{3}{\kappa} - 2$, whose massive counterpart corresponds to perturbation of $x \approx x_c$. Then tending mass to ∞ we should arrive to the second critical regime. Can we observe this effect in massive SLEs? Namely, does such massive $\text{SLE}(\kappa)$ tend to usual $\text{SLE}(\tilde{\kappa})$ as $m \rightarrow \infty$? In particular, does massive $\text{SLE}(3)$ tend to $\text{SLE}(6)$? Do other observables lead to different dualities? Can we observe that tending mass to $-\infty$ leads to the frozen regime $x \in [0, x_c[$?

Question 4.13. Can massive observables (which are in some sense better behaved) help to understand the critical ones? This is the case in Ref. 22, where we establish the conjectured value of the critical temperature for FK model with $q \geq 4$.

Acknowledgments

This research was supported by the N.S.F. Grant No. 0201893, by the Swiss N.S.F. and by the European Research Council AG CONFRA.

References

- [1] C. Itzykson, H. Saleur and J.-B. Zuber (eds.), *Conformal invariance and applications to statistical mechanics* (World Scientific Publishing Co. Inc., Teaneck, NJ, 1988).
- [2] O. Schramm, Conformally invariant scaling limits: an overview and a collection of problems, in *Proceedings of the International Congress of Mathematicians (Madrid, August 22-30, 2006)*, (Eur. Math. Soc., Zürich, 2007) pp. 513–543.
- [3] S. Smirnov, Towards conformal invariance of 2D lattice models, in *Proceedings of the International Congress of Mathematicians (Madrid, August 22-30, 2006)*, (Eur. Math. Soc., Zürich, 2006) pp. 1421–1451.
- [4] M. Bauer, D. Bernard and K. Kytölä, *J. Stat. Phys.* **132**, 721 (2008).
- [5] M. Bauer, D. Bernard, and L. Cantini, *J. Stat. Mech. Theory Exp.*, P07037, (electronic) (2009).
- [6] P. Nolin and W. Werner, *J. Amer. Math. Soc.* **22**, 797 (2009).
- [7] C. Garban, G. Pete and O. Schramm, Work in progress (2008).
- [8] N. Makarov and S. Smirnov, Massive stochastic Loewner evolutions, In preparation (2009).
- [9] J. Dubédat, SLE and the free field: Partition functions and couplings, Preprint (2007).
- [10] O. Schramm and S. Sheffield, *Ann. Probab.* **33**, 2127 (2005).
- [11] O. Schramm and S. Sheffield, *Acta Math.* **202**, 21 (2009).
- [12] G. F. Lawler, O. Schramm and W. Werner, *Ann. Probab.* **32**, 939 (2004).
- [13] S. Smirnov, Conformal invariance in random cluster models. I. Holomorphic spin structures in the Ising model, *Ann. of Math. (2)* (to appear).
- [14] V. Riva and J. Cardy, *J. Stat. Mech. Theory Exp.*, P12001, (electronic) (2006).
- [15] D. Chelkak and S. Smirnov, Universality in the 2D Ising model and conformal invariance of fermionic observables, Preprint arXiv:0910.2045 (2009).
- [16] D. Chelkak and S. Smirnov, Conformal invariance of the 2D Ising model at criticality, Preprint (2009).
- [17] S. Smirnov, Conformal invariance in random cluster models, II. Scaling limit of the interface, Preprint (2009).
- [18] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and martingales. Vol. 2* (Cambridge University Press, Cambridge, 2000).
- [19] B. Øksendal, *Stochastic differential equations*, (Springer-Verlag, Berlin, 2003).
- [20] W. Kager and B. Nienhuis, *J. Statist. Phys.* **115**, 1149 (2004).
- [21] S. Smirnov, *C. R. Math. Acad. Sci. Paris* **333**, 239 (2001).
- [22] V. Beffara, H. Duminil-Copin and S. Smirnov, The self-dual point of the 2D random-cluster model is critical above $q = 4$, Preprint (2009).

DISORDER RELEVANCE FOR PINNING/WETTING MODELS: A REVIEW

FABIO L. TONINELLI

*CNRS and Laboratoire de Physique de l'ENS Lyon,
46 Allée d'Italie, 69364 Lyon, France
E-mail: fabio-lucio.toninelli@ens-lyon.fr
www.ens-lyon.fr/PHYSIQUE*

We give a review of recent rigorous results concerning the Harris criterion (relevance/irrelevance of disorder) for pinning/wetting models with quenched randomness.

Keywords: Pinning/wetting models; Disorder relevance; Harris criterion

1. Random pinning/wetting models and the Harris criterion

The models we are interested in are motivated by (bio)-physical situations (interface pinning in the 2D Ising model, denaturation of biopolymers, pinning of a directed polymer by a defect line, etc; we refer to [1] and to [2, Ch. 1] for more motivations and bibliography) but they can be formulated in a general way starting from a renewal sequence τ . This is a sequence of integer-valued random variables $\tau_0 = 0 < \tau_1 < \tau_2 \dots$ such that the jumps $(\tau_i - \tau_{i-1})$ are i.i.d. random variables. One assumes that the inter-arrival law has heavy tails:

$$\mathbf{P}(\tau_i - \tau_{i-1} = n) = \frac{L(n)}{n^{1+\alpha}} \quad (1)$$

for some $\alpha \in (0, 1)$ (we omit here for brevity the still non-trivial but somewhat less interesting case $\alpha \geq 1$). $L(\cdot)$ is a slowly varying function (say, a power of a logarithm) but, except for a brief discussion at the end of next section, we will suppose for simplicity that it is asymptotically constant for $n \rightarrow \infty$. The parameter α depends on the particular situation one has in mind: for instance, $\alpha = 1/2$ corresponds to 2D wetting and to polymer pinning in dimension $1 + 1$ or $1 + 3$. The law of the renewal τ is modified by the presence of an infinite sequence of i.i.d. (say, Gaussian) random variables $\omega = \{\omega_1, \omega_2, \dots\}$. More precisely, the Gibbs measure $\mathbf{P}_{N,\omega}$ for the system of length N can be written as

$$\frac{d\mathbf{P}_{N,\omega}}{d\mathbf{P}}(\tau) = \frac{e^{\sum_{n=1}^N (\beta\omega_n + h)\mathbf{1}_{n \in \tau}}}{Z_{N,\omega}} \quad (2)$$

where $\beta \geq 0$ represents the disorder strength and h is the homogeneous pinning parameter. For any given $\beta \geq 0$, there exist a *localized* and a *delocalized* phase,

separated by a unique critical point $h_c(\beta)$: for $h > h_c(\beta)$ (localized phase), $\mathbf{E}_{N,\omega}|\tau \cap \{1, \dots, N\}|$ (the average number of points in τ before site N) grows proportionally to N and the free energy $F(\beta, h) = \lim_{N \rightarrow \infty} (1/N) \mathbb{E} \log Z_{N,\omega}$ is positive ($\mathbb{E}$ denotes the average on the disorder ω), while for $h < h_c(\beta)$ (delocalized phase) one has $\mathbf{E}_{N,\omega}|\tau \cap \{1, \dots, N\}| = o(N)$ and $F(\beta, h) = 0$. It is also practical to define the critical exponent $\nu(\beta)$ (it can and actually will depend on β) by $F(\beta, h) \stackrel{h \searrow h_c(\beta)}{\sim} \text{const} \times (h - h_c(\beta))^{\nu(\beta)}$ (the behavior near criticality is strictly speaking not proven in general to be of pure power-law type, but for the purposes of this brief note we will ignore this aspect).

A natural question is whether (quenched) critical point $h_c(\beta)$ and critical exponent $\nu(\beta)$ coincide with those of the annealed model, which is defined by averaging the partition function inside the logarithm: $F^{\text{ann}}(\beta, h) = \lim_{N \rightarrow \infty} (1/N) \log \mathbb{E} Z_{N,\omega}$. The annealed model can be explicitly solved [2, Ch. 2]. In particular, one computes the annealed critical point $h_c^{\text{ann}}(\beta) = -\log \mathbf{P}(\tau_1 < \infty) - \log \mathbb{E}(\exp(\beta\omega_1))$ and $\nu^{\text{ann}} = \nu(0) = 1/\alpha$.

The Harris criterion, which is basically based on a weak-disorder expansion of the free energy around the annealed value [2, Ch. 5.5], predicts that disorder is irrelevant (i.e., $h_c(\beta) = h_c^{\text{ann}}(\beta)$ and $\nu(\beta) = \nu^{\text{ann}}$ for β small) if $\alpha < 1/2$, and relevant if $\alpha > 1/2$ (strict inequalities instead of equalities for every $\beta > 0$). For another way to justify the Harris criterion predictions, cf. [3]. The marginal case $\alpha = 1/2$ is particularly interesting and controversial: some authors (in particular [4, 5]) predicted irrelevance, while B. Derrida *et al.* [3] predicted that $h_c(\beta) - h_c^{\text{ann}}(\beta) \sim \exp(-c/\beta^2)$ for β small.

The last few years have witnessed a remarkable progress in the comprehension of these issues: in particular, as we are going to report below, not only the predictions of the Harris criterion have been mathematically justified, but also the controversy concerning the marginal case $\alpha = 1/2$ has been resolved.

2. Rigorous results

The irrelevant disorder regime is now quite well understood:

Theorem 2.1. *Assume that $\alpha \in (0, 1/2)$. Then, for β smaller than some $\beta_0 > 0$ one has*

$$h_c(\beta) = h_c^{\text{ann}}(\beta) \tag{3}$$

and

$$F(\beta, h) \stackrel{h \searrow h_c(\beta)}{\sim} F^{\text{ann}}(\beta, h) \tag{4}$$

so that in particular $\nu(\beta) = 1/\alpha = \nu^{\text{ann}}$.

This result was proven in [6] via a refined second moment method, and then in [7] via replica-coupling ideas. Actually, the estimate (4) can be refined to show that $F(\beta, h) - F^{\text{ann}}(\beta, h) \sim -\beta^2/2(\partial_h F^{\text{ann}}(\beta, h))^2$ for $\beta \searrow 0$ [8].

An interesting open problem for $\alpha < 1/2$ is that of the transition from weak to strong disorder. Namely, in [9] we proved that for β large enough quenched and annealed critical points differ, in contrast with (3). The precise value of the weak-to-strong transition threshold, as well as the behavior of the system there, are unknown. It is also somewhat natural to expect that $\nu(\beta) \neq 1/\alpha = \nu^{\text{ann}}$ for $\beta \gg 1$, but to our knowledge there are no arguments (even heuristic) which would allow to guess the value of $\nu(\beta)$.

Next, the relevant disorder regime:

Theorem 2.2. *Assume that $\alpha \in (1/2, 1)$. Then, there exists a positive and finite constant c such that, for every $\beta \leq 1$,*

$$c\beta^{2\alpha/(2\alpha-1)} \leq h_c(\beta) - h_c^{\text{ann}}(\beta) \leq c^{-1}\beta^{2\alpha/(2\alpha-1)}. \quad (5)$$

The upper bound is proven in [6, 7] and the lower bound in [10, 11].

While the result (5) on the shift of the critical point says nothing about the critical exponent, we were able to prove that ν is modified by an arbitrarily small disorder: indeed, we proved [12, 13] that as soon as $\beta > 0$ one has $\nu(\beta) \geq 2(> \nu(0) = 1/\alpha = \nu^{\text{ann}})$.

Finally, the controversial case of marginal disorder is now also under control, and we essentially proved the prediction of [3]:

Theorem 2.3. *[14, 15] If $\alpha = 1/2$, there exists a positive constant c and, for every $\epsilon > 0$, a constant $c(\epsilon) > 0$ such that, for $\beta \leq 1$,*

$$\exp(-c(\epsilon)/\beta^{2+\epsilon}) \leq h_c(\beta) - h_c^{\text{ann}}(\beta) \leq \exp(-c/\beta^2).$$

Note that in this case the bound $\nu(\beta) \geq 2$ for $\beta > 0$ [12, 13] is in principle compatible with $\nu(\beta) = \nu^{\text{ann}} = 1/\alpha = 2$, although there is no particular reason to believe that such an equality should hold [3].

We see therefore that, in the marginal case $\alpha = 1/2$, disorder is marginally relevant. This is in contrast with another celebrated marginal case, the Ising model in dimension two with bond disorder, where disorder is believed to be marginally irrelevant [16] and to lead just to logarithmic corrections to the critical behavior of the pure model.

We conclude with one observation concerning the case where the inter-arrival law $\mathbf{P}(\tau_i - \tau_{i-1} = n)$ involves a non-trivial slowly varying function, i.e. $L(\cdot)$ in (1) is not asymptotically constant. While not much new happens for $\alpha \neq 1/2$, in the marginal case $\alpha = 1/2$ the function $L(\cdot)$ allows to zoom more precisely in the critical window. In particular, it is expected that (for $\alpha = 1/2$) disorder is irrelevant if and only if $\sum_{n \geq 1} 1/(nL(n)^2) < \infty$. This condition is equivalent to requiring that the intersection $\tau^{1,2} := \tau^1 \cap \tau^2$ between two independent copies τ^1, τ^2 of the original renewal τ is a *transient* renewal, i.e., that $\tau^{1,2}$ contains only a finite number of points. The “if” part of the conjecture was proven in [6, 7], while the “only if” part is proven in [15] under the assumption that $L(\cdot)$ is asymptotic to a power of $\log(\cdot)$.

3. A coarse graining/change of measure idea

The idea we use to prove that $h_c(\beta) \neq h_c^{\text{ann}}(\beta)$ for β small and $\alpha \geq 1/2$ is roughly the following. One wishes to prove that $Z_{N,\omega} \rightarrow 0$ for $N \rightarrow \infty$, while $\mathbb{E}Z_{N,\omega}$ diverges exponentially. One looks then at fractional moments of the partition function, i.e. $\mathbb{E}(Z_{N,\omega}^\gamma)$ with $\gamma \in (0, 1)$, with the goal to prove that such moment does tend to zero for $N \rightarrow \infty$, while $\mathbb{E}Z_{N,\omega}$ does not. The basic observation is that (via Hölder inequality)

$$\mathbb{E}(Z_{N,\omega}^\gamma) \leq (\tilde{\mathbb{E}}Z_{N,\omega})^\gamma \left(\tilde{\mathbb{E}} \left(\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}} \right)^{\frac{1}{1-\gamma}} \right)^{1-\gamma}, \quad (6)$$

where $\tilde{\mathbb{P}}$ is a new law on the disorder variables, to be chosen suitably. All the trick is to find $\tilde{\mathbb{P}}$ which is sufficiently close to $\mathbb{P}$ (so that the last term in (6) is not too large) but at the same time such that $\tilde{\mathbb{E}}Z_{N,\omega} \ll \mathbb{E}Z_{N,\omega}$. In general, under the new measure the variables ω_i are not independent any more, and the correlations must be carefully chosen to reflect the structure of the Green function of the annealed model [14]. Moreover, such a procedure works only for N up to the size of the correlation length of the annealed model: for larger system sizes, the desired upper bound on $\mathbb{E}(Z_{N,\omega}^\gamma)$ then follows through an iteration argument. The actual proof requires also a coarse-graining step of the type of that introduced in [17].

Extensions

We wish to mention that the fractional moment/change of measure ideas outlined above have been later successfully applied to other models:

- (a) Directed polymers in bulk disorder: H. Lacoïn proved [18] that, for the model in dimension $1 + 2$, quenched and annealed free energies differ for arbitrarily small disorder strength. The same was known to occur in dimension $1 + 1$ [19], and is known *not* to occur in dimension $1 + d$, $d \geq 3$. In a sense, $d = 2$ corresponds here to the marginal case $\alpha = 1/2$ of the pinning model.
- (b) Random Walk Pinning Model: one takes two random walks, X and Y , on $\mathbb{Z}^d$. The Y trajectory is frozen (quenched disorder), and the law of X is modified by the interaction $\lambda \sum_n \mathbf{1}_{X_n=Y_n}$. The critical value λ_c is defined as that for which the two walks start to stick together. M. Birkner and R. Sun proved [20] that λ_c coincides with the annealed critical λ if $d = 1, 2$, while it differs if $d > 3$. In this case, $d = 3$ corresponds in a sense to the marginal situation $\alpha = 1/2$, but the question of the coincidence or not of the two critical points for $d = 3$ is still open for this model.

Acknowledgments

This work has been supported by ANR, grant POLINTBIO.

References

- [1] M. E. Fisher, *J. Statist. Phys.* **34**, 667 (1984).
- [2] G. Giacomin, *Random Polymer Models* (World Scientific, Singapore, 2007).
- [3] B. Derrida, V. Hakim and J. Vannimenus, *J. Statist. Phys.* **66**, 1189 (1992).
- [4] G. Forgacs, J. M. Luck, T. M. Nieuwenhuizen and H. Orland, *Phys. Rev. Lett.* **57**, 2184 (1986).
- [5] D. M. Gangardt and S. K. Nechaev, *J. Statist. Phys.* **130**, 483 (2008).
- [6] K. S. Alexander, *Comm. Math. Phys.* **279**, 117 (2008).
- [7] F. L. Toninelli, *Commun. Math. Phys.* **280**, 389 (2008).
- [8] G. Giacomin and F. L. Toninelli, *Ann. Probab.*, to appear (2007).
- [9] F. L. Toninelli, *Ann. Appl. Probab.* **18**, 1569 (2008).
- [10] B. Derrida, G. Giacomin, H. Lacoïn and F. L. Toninelli, *Comm. Math. Phys.* **287**, 867 (2009).
- [11] K. S. Alexander and N. Zygouras, *preprint arXiv:0805.1708* (2008).
- [12] G. Giacomin and F. L. Toninelli, *Commun. Math. Phys.* **266**, 1 (2006).
- [13] G. Giacomin and F. L. Toninelli, *Phys. Rev. Lett.* **96**, p. 070602 (2006).
- [14] G. Giacomin, H. Lacoïn and F. L. Toninelli, *Comm. Pure Appl. Math.*, to appear (2009).
- [15] G. Giacomin, H. Lacoïn and F. L. Toninelli, *preprint arXiv:0906.1942* (2009).
- [16] V. S. Dotsenko and V. S. Dotsenko, *Adv. Phys.* **32**, p. 129 (1983).
- [17] F. L. Toninelli, *Electron. J. Probab.* **14**, 531 (2009).
- [18] H. Lacoïn, *preprint arXiv:0901.0699* (2009).
- [19] F. Comets and V. Vargas, *ALEA* **2**, 267 (2006).
- [20] M. Birkner and R. Sun, *Ann. Inst. H. Poincaré: Prob. Stat.*, to appear (2008).

OTHER TALKS

INVITED TALK

ISOTROPICITY IMPLIES SUBBALISTICITY

GADY KOZMA

Faculty of Mathematics, Weizmann Institute

E-mail: gady.kozma@weizmann.ac.il

We demonstrate a general method for proving subballisticity. The argument works for the mirror model, the Manhattan model, isotropic RWRE, and partially for reinforced random walk.

CONTRIBUTED TALKS

LAYERING AND WETTING TRANSITIONS FOR AN SOS INTERFACE WITHOUT EXTERNAL FIELD

KENNETH S. ALEXANDER

Department of Mathematics, University of Southern California

E-mail: alexandr@usc.edu

We study the solid-on-solid interface model above a horizontal wall in three dimensional space, with an attractive interaction when the interface is in contact with the wall, at low temperatures. There is no bulk external field. The system presents a sequence of layering transitions, before reaching the wetting transition. The number of layering transitions as the wall attraction is varied becomes arbitrarily large (and may become infinite) as the temperature decreases. This contrasts with existing rigorous work which considers layering transitions as an external field is varied.

POINT GROUPS, GALOIS COVERINGS, AND THEIR APPLICATIONS TO RG ANALYSES IN TWO-DIMENSIONAL LATTICE MODELS

MASAFUMI FUJIMOTO

Department of Physics, Nara Medical University

E-mail: mfujimot@nmu-gw.naramed-u.ac.jp

Recently, for solvable lattice models, it was shown that asymptotic behavior of the correlation function is expressed in terms of differential forms on Riemann surfaces of genus 1 [1-3]. Choosing a suitable parametrization, we found that point group of the system essentially determines the differential forms. We also found a one-to-one correspondence between point groups and Galois coverings of Riemann surfaces [4]. In this presentation we propose a renormalization group (RG) approach by the use of Galois coverings.

We consider the square-lattice Ising model, where the point group is C_{4v} . Using the fact that $C_{4v} \supset C_{2v}$, we construct a Galois extension of an elliptic function field. It follows that the extension corresponds to Landen's transformation, and that Landen's transformation gives a RG approach near the critical point [5]. The RG method derived here is expected to be a quite general one. It is shown that the same method is applicable to a wide class of solvable models possessing four- or six-fold rotational symmetry [2, 6]. With the help of series expansions, we discuss further generalizations to unsolvable systems.

- [1] M. Holzer, Phys. Rev. Lett. 64 (1990) 653; Phys. Rev. B42 (1990) 10570; Y. Akutsu and N. Akutsu, Phys. Rev. Lett. 64 (1990) 1189.
- [2] M. Fujimoto, Physica A 233 (1996) 485; J. Stat. Phys. 90 (1998) 363.
- [3] M. Fujimoto, J. Phys. A:Math. Gen. 35 (2002) 7553.
- [4] M. Namba, Branched Coverings and Algebraic Functions (Longman).
- [5] H.J. Hilhorst, M. Schick, and J.M.J. van Leeuwen, Phys. Rev. B19 (1979) 2749; W. Jezewski, Phys. Rev. B24 (1981) 3954.
- [6] R.J. Baxter, Exactly Solved Models in Statistical Mechanics (Academic Press).

QUENCHED CLT FOR ENERGY IN MF SPIN GLASSES

SHANNON STARR, BRIGITTA VERMESI

Department of Mathematics, University of Rochester

E-mail: kalvotom@fjfi.cvut.cz

Generalizing a result of Chatterjee for the SK model, we prove a quenched CLT for the energy in the annealed region of mean-field spin glasses with non-Gaussian couplings, including alpha-stable couplings. One other example is the Viana-Bray model. We also prove Gaussian and non-Gaussian central limits for the random internal energy, in particular with a phase transition at $\beta = 0$ for the non-Gaussian case.

ON RANDOM PERMUTATIONS WITH NONUNIFORM DISTRIBUTIONS

VOLKER BETZ, DANIEL UELTSCHI*, YVAN VELENIK

**Department of Mathematics, University of Warwick*

E-mail: daniel@ueltschi.org

I will discuss various models of random permutations with nonuniform distributions. One model involves weights that depend on the cycle structure of the permutation. The other models deal with permutations of points in space, and there is an additional weight that involves the length of permutation jumps. The main question is about the possible occurrence of infinite cycles at high enough density. These models of random permutations are motivated by quantum statistical mechanics and by mathematical biology.

POSTERS

SHIGERU FURUICHI (Nihon University): **ON ONE-PARAMETER EXTENDED FISHER INFORMATION AND CRAMER-RAO INEQUALITY**

MILAN KRBÁLEK (FNSPE, Czech Technical University, Prague), with PAVEL HRABÁK: **PROBABILITY DENSITY FOR DISTANCE HEADWAY IN ASYMMETRIC SIMPLE EXCLUSION PROCESS**

MUKADAS D. MISSAROV (Kazan State University, Russia): **FULL RENORMALIZATION GROUP DESCRIPTION OF THE HIERARCHICAL FERMIONIC MODEL**

JAN NAUDTS (Universiteit Antwerpen): **THE GENERALISED EXPONENTIAL FAMILY IN STATISTICAL PHYSICS**

ALEX A. OPOKU (University of Groningen): **GIBBS-NON-GIBBS PROPERTIES OF TRANSFORMS OF LATTICE AND MEAN-FIELD XY MODELS**

VAHAGN S. POGHOSYAN (Joint Institute for Nuclear Research, Dubna): **SPANNING WEBS AND SINGULAR TOEPLITZ MATRICES IN THE DIMER MODEL**

ALEKSEI L. REBENKO (Institute of Mathematics, Kyiv), with MAXYM V. TERTYCHNYI: **QUASI-LATTICE APPROXIMATION OF STATISTICAL SYSTEMS WITH STRONG SUPERSTABLE INTERACTION**

YONCA SEZER (Dept of Mathematics, Yildiz Technical University), with E. ADIGUZELOV: **THE SECOND REGULARIZED TRACE OF A DIFFERENTIAL OPERATOR WITH OPERATOR COEFFICIENT**

HIROHIKO SHIMADA (University of Tokyo): **DISORDERED $O(N)$ LOOP MODEL AND COUPLED CONFORMAL FIELD THEORIES**

ALEXANDER SUKHANOV (Joint Institute for Nuclear Research), with OLGA GOLUBEVA: **A QUANTUM GENERALIZATION OF EQUILIBRIUM STATISTICAL THERMODYNAMICS: MACRO- AND MICRO- VERSIONS OF THE THEORY**

HIROSHI TAMURA (Kanazawa University): **BOSE GASES IN SCALED HARMONIC TRAPS IN TERMS OF RANDOM POINT FIELDS**

JACEK WOJTKIEWICZ (Faculty of Physics, University of Warsaw): **ESTIMATIONS OF THE FREE ENERGY FOR THE HUBBARD MODEL**

JACEK WOJTKIEWICZ (Faculty of Physics, University of Warsaw): **BOUNDS ON SUSCEPTIBILITIES OF HUBBARD MODELS**

Non-Equilibrium Statistical Mechanics

(including stochastic evolutions
except SLE and Boltzmann equation)

Session organizers:

Herbert Spohn
Horng-Tzer Yau

DIFFUSION IN COUPLED MAP LATTICES

ANTTI KUPIAINEN

Department of Mathematics, Helsinki University,

P.O. Box 4, 00014 Helsinki, Finland

E-mail: ajkupiai@cc.helsinki.fi

We consider a dynamical system consisting of subsystems indexed by a lattice. Each subsystem has one conserved degree of freedom (“energy”) the rest being uniformly hyperbolic. The subsystems are weakly coupled together so that the sum of the subsystem energies remains conserved. We prove that the long time dynamics of the subsystem energies is diffusive.

Keywords: diffusion; deterministic dynamics; renormalization group

1. Coupled Maps with a conservation law

One of the fundamental problems in deterministic dynamics is to understand the microscopic origin of diffusion. On a microscopic level a physical system such as a fluid or a crystal can be modeled by Schrödinger or Hamiltonian dynamics with a macroscopic number of degrees of freedom. Although the microscopic dynamics is not dissipative dissipation should emerge in large spatial and temporal scales e.g. in the form of diffusion of heat or concentration of particles.

Dynamically diffusion is related to the existence in the system of conserved quantities such as the energy which are *extensive* i.e. sums (or integrals) of local contributions that are “almost conserved”. Thus, if the system has a microscopic energy density $E(t, x)$, $x \in \mathbb{R}^d$ the total energy $E_{\text{tot}} = \int E(t, x) dx$ is a constant of motion but the energy density is in general not since the dynamics in general redistributes it:

$$\dot{E}(t, x) = \nabla \cdot \mathbf{J}(t, \mathbf{x}). \quad (1)$$

The divergence acting on the energy current J guarantees conservation of the total energy. One would like to show that the conservative dynamics (1) turns in a suitable scaling limit to a diffusive one. In the Hamiltonian context such a limit involves diffusive scaling of space and time and taking typical initial conditions with respect to the Liouville measure with prescribed initial energy profile. The rescaled energy density $E_L(t, x) := L^d E(L^2 t, Lx)$ should then satisfy, as $L \rightarrow \infty$ the diffusion equation

$$\dot{E}(t, x) = \kappa \Delta E(t, x). \quad (2)$$

There has been a lot of work in recent years around these questions in the context of **coupled dynamics** i.e. dynamical systems consisting of elementary systems indexed by a d -dimensional lattice $\mathbb{Z}^d$. The total energy E of the system is a sum $\sum_x E_x$ of energies E_x which involve the dynamical variables of the system at lattice site x and nearby sites.

In particular, two classes of models have been discussed. The first consists of Hamiltonian dynamics of coupled **weakly anharmonic** oscillators. In the weak anharmonicity scaling limit one may use kinetic theory to compute the conductivity κ . Rigorous justification of the kinetic limit let alone the case of fixed (small) coefficient of the anharmonic term is still lacking. For progress in the former case see [1] and an attempt to study the latter one [2].

A second class of models deals with a complementary situation of **weakly coupled chaotic** systems [3], [4], [5]. Hamiltonian systems of this type are obtained by putting at each lattice site a chaotic system, e.g. a billiard and coupling them weakly to each other. One might hope that the strong mixing properties of the billiard dynamics could help in proving diffusion. Rigorous results on such Hamiltonian systems are rare: in [3] ergodicity is proved for the one dimensional case.

In this talk I will consider these issues in the framework of Coupled Map Lattices (CML) which are discrete time versions of the systems discussed above.

A CML consists of subsystems, indexed by a lattice $\mathbb{Z}^d$. The dynamics is a perturbation of an uncoupled dynamics where each subsystem has its proper dynamics. Letting M be the phase space of the uncoupled systems and $f : M \rightarrow M$ their local dynamics the phase space of the full system is $\mathcal{M} = M^{\mathbb{Z}^d}$ and the CML dynamics $F : \mathcal{M} \rightarrow \mathcal{M}$ is a perturbation of the uncoupled dynamics $f^{\mathbb{Z}^d}$.

To motivate our choice of M and F consider again the coupled billiard case. The single billiard dynamics has very good chaotic properties. A discrete time version (say given by a Poincaré map) has one vanishing Lyapunov exponent corresponding to the conserved energy and the remaining ones nonzero. Thus, to model such a situation we let the subsystems have $M = \mathbb{R}_+ \times N$ and so at the lattice site $x \in \mathbb{Z}^d$ we have variables $(E(x), \theta(x))$. We call the non-negative variables E energy. These are the conserved ones: under the uncoupled dynamics

$$(E(x), \theta(x)) \rightarrow (E(x), g(\theta(x), E(x))) \quad (3)$$

for each $x \in \mathbb{Z}^d$.

θ are the fast, chaotic variables. In the billiard case the dynamical system $\theta \rightarrow g(\theta, E)$ is uniformly hyperbolic for any fixed E . We will model this situation by taking $g(\theta, E) = g(\theta)$ a fixed chaotic map, independent of E . For simplicity we fix $N = S^1 = \mathbb{R}/\mathbb{Z}$ and g an expansive circle map, e.g. $g(\theta) = 2\theta$.

The E independence is a serious simplification: in a realistic Hamiltonian system, such as the billiards the E dependence of g can not be ignored. Indeed, it is obvious that as $E \rightarrow 0$ the Lyapunov exponents of $g(\cdot, E)$ also tend to zero since E sets the time scale. We will consider only a perturbative E dependence by taking the full

CML dynamics to be

$$F(x, E, \theta) = (E(x) + f(x, E, \theta), g(\theta(x)) + h(x, E, \theta)). \quad (4)$$

Here f and h are small local functions of (E, θ) i.e. they depend weakly on $(E(y), \theta(y))$ for $|x - y|$ large.

We will consider perturbations such that the *total energy* $\sum_x E(x)$ is conserved. This follows if $\sum_x f(x, E, \theta) = 0$ for all E, θ . A natural way to guarantee this is to consider a “vector field” $\mathbf{J}(x) = \{J^\mu(x)\}_{\mu=1, \dots, d}$ and take

$$f(x, E, \theta) = (\nabla \cdot \mathbf{J})(x, E, \theta) := \sum_{\mu} (J^\mu(x + e_\mu, E, \theta) - J^\mu(x, E, \theta)) \quad (5)$$

With these definitions we arrive at the time evolution

$$E(t+1, x) = E(t, x) + \nabla \cdot \mathbf{J}(x, E(t), \theta(t)) \quad (6)$$

$$\theta(t+1, x) = g(\theta(t, x)) + h(x, E(t), \theta(t)). \quad (7)$$

Note that (6) is a natural discrete space time version of (1).

2. Quenched diffusion

We would like to inquire under what conditions the dynamics (6) of the energies is diffusive **for almost all initial values of θ** . “Almost all” means the following. Suppose first $h = 0$. The θ dynamics is then local and g has an invariant measure μ , absolutely continuous w.r.t. the Lebesgue measure on S^1 . We would thus like to prove diffusion of E a.s. in $\theta(0, \cdot)$ with respect to the measure $\nu = \mu^{\mathbb{Z}^d}$.

The θ dynamics is exponentially mixing in time i.e. $\theta(t)$ are random variables with respect to ν which are exponentially decorrelated in time. Let $\mathbf{J}_t(x, E) := \mathbf{J}(x, E, \theta(t))$. $\mathbf{J}_t(x, E)$ become then random variables and with suitable smoothness assumptions on $\mathbf{J}$ as a function of θ , also $\mathbf{J}_t(x, E)$ have short memory.

Suppose next $h \neq 0$ but it depends only on θ . If the dependence is local and smooth and h small in a suitable sense [6] the θ dynamics has an invariant measure which is a Gibbs measure on $(S^1)^{\mathbb{Z}^d}$ and the $\theta(t, x)$ are exponentially decorrelated both in space and time. Then the random variables $\mathbf{J}_t(x, E)$ will be exponentially weakly correlated in space and time. We will restrict our attention here to this case. E dependence in h will make $\mathbf{J}_t$ depend on $E(s)$ for $s < t$ too. The dependence is exponentially weak as $|t - s| \rightarrow \infty$, but makes the analysis more messy.

We may now rephrase the problem of deriving diffusion in deterministic dynamics as that of **quenched diffusion in random dynamics**:

$$E(t+1, x) = E(t, x) + \nabla \cdot \mathbf{J}_t(x, E(t)) := \Phi_t(x, E(t)) \quad (8)$$

i.e. we want to prove that the $E(t)$ diffuses **almost surely** in the realization of the random field $\mathbf{J}_t$.

We will next list the assumptions on $\mathbf{J}$ that allow us to prove quenched diffusion. The following three assumptions follow by making natural symmetry, locality and analyticity assumptions for the original Coupled Map Lattice (6), (7).

(i) *Symmetry.* Let R be a translation by $y \in \mathbb{Z}^d$ or a rotation preserving $\mathbb{Z}^d$ or the reflection $x \rightarrow -x$. We require that $\nabla \cdot \mathbf{J}_t$ and $\nabla \cdot \mathbf{J}_t(\mathbf{R} \cdot, \mathbf{E}(\mathbf{R}^{-1} \cdot))$ are equal in law.

(ii) *Localization.* We suppose $\mathbf{J}$ is localized as follows:

$$\mathbf{J}_t(x, E) = \sum_{A, B \subset \mathbb{Z}^d} \mathbf{J}_{tAB}(x, E) \quad (9)$$

where $\mathbf{J}_{tAB}$ is a function of $E|_A$. B refers to noise dependence, see (iii).

(iii) *Bounds* $\mathbf{J}_{tAB}$ is analytic in the domain

$$\Sigma_A = \{E \mid |E(x)| < R, \quad x \in A\} \quad (10)$$

with a bound

$$\sup_{E \in \Sigma_A} |\mathbf{J}_{tAB}(x, E)| \leq C e^{-md(x, A, B)} \quad (11)$$

for some $m > 0$ where $d(x, A, B)$ is the length of the shortest graph on $\mathbb{Z}^d$ whose set of vertices contains x and $A \cup B$. Let

$$\mathbf{j}_{tAB} = \mathbf{J}_{tAB} - \mathbb{E} \mathbf{J}_{tAB}.$$

Then $\mathbf{j}_{tAB}$ are weakly correlated in space and time

$$\mathbb{E}(\mathbf{j}_{tAB}(x, E) - \mathbf{j}_{t'A'B'}(x', E))^2 \leq C e^{-2m(\text{dist}(B, B') + |t-t'| + |x-x'|)}$$

The following three assumptions are special and amount to say that our system is a small nonlinear random perturbation of a linear diffusion. In particular in the billiard system we expect rare configurations of the energies for which these assumptions are surely violated.

(iv) *Positivity.* $\Phi_t(x, E) \geq 0$ for a non-negative E .

(v) *Ellipticity.* Let $k_t(x, y) = \frac{\partial}{\partial E(y)} \nabla \cdot \mathbf{J}_t(x, 0)$. Assumptions (i)-(iv) imply that $\mathbb{E} k_t(x, y) = k(x - y)$ where k decays exponentially and has a real valued Fourier transform $\hat{k}(p)$. We make the ellipticity assumption

$$\hat{k}(p) < 0. \quad (12)$$

Then the symmetries imply $\hat{k}(p) = -Dp^2 + \mathcal{O}(|p|^4)$ with $D > 0$.

(vi) *Small perturbation.* We assume that the fluctuations of k are small:

$$\mathbb{E} ((k_t(x, y) - k(x - y)))^2 \leq \epsilon e^{-2m|x-y|} \quad (13)$$

where ϵ will be taken small.

To state our result let $p > d + 1$ and introduce the norm

$$\|E\| = \sup_x |E(x)|(1 + |x|)^p. \quad (14)$$

Theorem. *Let $\mathbf{J}$ satisfy the assumptions (i)-(vi). Then for $\|E\|$ small enough and almost surely in the randomness the rescaled field $\varepsilon_t(x) = t^{d/2}E(t, \sqrt{t}x)$ satisfies*

$$\|\varepsilon_t - Ae^{-x^2/2Dd}\| \rightarrow 0 \quad (15)$$

for some $A \in \mathbb{R}$ as $t \rightarrow \infty$.

The proof [7] is based on a renormalization group introduced in [8] and [9]. The dynamics of the rescaled field ε_t is governed by the **renormalized map**

$$\Phi^t := S_{\sqrt{t}} \circ \Phi_t \circ \cdots \circ \Phi_1 \circ S_{\sqrt{t}}^{-1}$$

where $(S_L f)(x) = L^d f(Lx)$ is a scaling transformation. We show Φ^t can be written

$$\Phi^t(E) = e^{\kappa\Delta}E + \nabla \cdot J^t(E)$$

where Δ is the Laplacean and the current J^t has similar representation as above but tends a.s. to zero as $t \rightarrow \infty$. This means that both the randomness and the nonlinearity are irrelevant variables and the scaling limit is the linear heat flow.

Acknowledgments

This work is partially funded by the Academy of Finland and the European Research Council.

References

- [1] J. Lukkarinen, H. Spohn, Weakly nonlinear Schrödinger equation with random initial data, arXiv:0901.3283v1 [math-ph]
- [2] J.Bricmont, A. Kupiainen, On the derivation of Fourier's law for coupled anharmonic oscillators, Commun.Math. Phys. **274**, 555-626 (2007)
- [3] L. Bunimovich, C. Liverani, A. Pellegrinotti and Yu. Suhov, Ergodic systems of n balls in a billiard table, Commun.Math. Phys. **146**, 357 (1992)
- [4] C.Mejia-Monasterio, H. Larralde, and F. Leyvraz. Coupled normal heat and matter transport in a simple model system. Phys. Rev. Lett. **86** (2001), 5417-5420.
- [5] J.-P. Eckmann and L.-S. Young. Temperature profiles in Hamiltonian heat conduction. Europhysics Letters **68** (2004), 790-796.
- [6] J.Bricmont, A. Kupiainen, Coupled Analytic Maps, Nonlinearity **8**,379-393 (1995)
- [7] J.Bricmont, A. Kupiainen, In preparation
- [8] J.Bricmont, A. Kupiainen, Random Walks in Asymmetric Random Environments, Comm.Math.Phys. **142**, 345-420 (1991)
- [9] J.Bricmont, A. Kupiainen, Renormalization Group and Asymptotics of Solutions of Nonlinear Parabolic Equations, Commun. Pure.Appl.Math. **47**, 893-922 (1994)

DECAY OF EQUILIBRIUM TIME CORRELATIONS IN A WEAKLY NONLINEAR SCHRÖDINGER EQUATION

JANI LUKKARINEN

*Department of Mathematics and Statistics, University of Helsinki,
P.O. Box 68, 00014 Helsingin yliopisto, Finland
E-mail: jani.lukkarinen@helsinki.fi*

HERBERT SPOHN

*Zentrum Mathematik, Technische Universität München,
Boltzmannstr. 3, 85747 Garching, Germany
E-mail: spohn@ma.tum.de*

We report on a first progress in rigorous control of the kinetic scaling limit of a weakly nonlinear perturbation of wave-type evolution, here a discrete Schrödinger equation. Since we consider a Hamiltonian system, a natural choice of random initial data is distributing them according to a Gibbs measure with a chemical potential chosen so that the Gibbs field has exponential mixing. The solution $\psi_t(x)$ of the nonlinear Schrödinger equation yields then a stochastic process stationary in $x \in \mathbb{Z}^d$ and $t \in \mathbb{R}$. If λ denotes the strength of the nonlinearity, we prove that the space-time covariance of $\psi_t(x)$ has a limit as $\lambda \rightarrow 0$ for $t = \lambda^{-2}\tau$, with τ fixed and $|\tau|$ sufficiently small. The limit agrees with the prediction from kinetic theory.

Keywords: Time-dependent perturbation theory; Boltzmann-Peierls equation.

1. Introduction

We outline here the results proven in [1], aiming at a proper presentation of the main theorem. For the proof, and further discussion, we refer to the original manuscript.

The nonlinear Schrödinger equation (NLS) governs the evolution of a complex valued wave field $\psi : \mathbb{R} \times \mathbb{Z}^d \rightarrow \mathbb{C}$ and reads

$$i \frac{d}{dt} \psi_t(x) = \sum_{y \in \mathbb{Z}^d} \alpha(x-y) \psi_t(y) + \lambda |\psi_t(x)|^2 \psi_t(x). \quad (1)$$

Here $\alpha(x)$ are the “hopping amplitudes” and we assume that $\alpha : \mathbb{Z}^d \rightarrow \mathbb{R}$ is exponentially decreasing and symmetric, $\alpha(x) = \alpha(-x)$. We consider only the dispersive case $\lambda \geq 0$. In addition, to avoid technical problems in the definition of the time-evolution, we begin from the evolution equations defined on the finite lattice $\Lambda = \{0, 1, \dots, L-1\}^d$, $L \geq 2$, endowed with periodic boundary conditions. Then standard methods guarantee the existence and uniqueness of a global solution for any initial data $\psi_0 : \Lambda \rightarrow \mathbb{C}$.

Fourier transform of $f : \Lambda \rightarrow \mathbb{C}$ is denoted by $\hat{f} : \Lambda^* \rightarrow \mathbb{C}$, with the dual lattice $\Lambda^* = \{0, \frac{1}{L}, \dots, \frac{L-1}{L}\}^d$ and with $\hat{f}(k) = \sum_{x \in \Lambda} f(x) e^{-i2\pi k \cdot x}$, $k \in \Lambda^*$. Its inverse transform is given by $\tilde{g}(x) = \int_{\Lambda^*} dk g(k) e^{i2\pi k \cdot x}$; here and in the following we employ the shorthand notation $\int_{\Lambda^*} dk \dots = |\Lambda|^{-1} \sum_{k \in \Lambda^*} \dots$. The scalar product in $\ell_2(\Lambda)$ is denoted by $\langle f, \psi \rangle = \sum_{x \in \Lambda} f(x)^* \psi(x)$, and we use a similar but unrelated notation for “regularized” absolute values, $\langle x \rangle = \sqrt{1 + x^2}$. For notational simplicity, we will assume in the following that α has finite support, and we begin with large enough lattices so that $\alpha(x) = 0$ for all $|x| \geq L/2$. We use the notation $\mathbb{1}$ for a generic characteristic function: $\mathbb{1}(P) = 1$, if the condition P is true, and $\mathbb{1}(P) = 0$ otherwise. This allows defining a discrete δ -function on Λ^* by $\delta_\Lambda(k) = |\Lambda| \mathbb{1}(k \bmod 1 = 0)$.

Let $\omega : \mathbb{T}^d \rightarrow \mathbb{R}$ denote the dispersion relation, defined by $\omega = \hat{\alpha}$. We define the finite volume Hamiltonian by $H_\Lambda(\psi) = \sum_{x, y \in \Lambda} \alpha(x-y) \psi(x)^* \psi(y) + \frac{1}{2} \lambda \sum_{x \in \Lambda} |\psi(x)|^4$. Introducing the canonical conjugate pair $q_x, p_x \in \mathbb{R}$ by $\psi(x) = (q_x + ip_x)/\sqrt{2}$, we find that the evolution equations associated to H_Λ are equivalent to the finite volume discrete NLS. The energy $H_\Lambda(\psi)$ is naturally conserved by the time-evolution. In addition, by differentiation we find that the norm $N_\Lambda(\psi) = \|\psi\|_2^2$ is also a constant of motion. From a statistical physics point of view a very natural choice of initial data is to take the initial ψ -field to be distributed according to a Gibbs measure for H and N , which physically means that the wave field is in thermal equilibrium. Explicitly, the initial field $\psi_0 = \psi$ is distributed as

$$\int_{\mathbb{C}^\Lambda} \mathbb{P}_\Lambda^\lambda(d\psi) f(\psi) = \frac{1}{Z_\Lambda^\lambda} \int_{(\mathbb{R}^2)^\Lambda} \prod_{x \in \Lambda} [d(\operatorname{Re} \psi(x)) d(\operatorname{Im} \psi(x))] e^{-\beta(H_\Lambda(\psi) - \mu \|\psi\|^2)} f(\psi).$$

Here $\beta > 0$ is the inverse temperature and $\mu \in \mathbb{R}$ the chemical potential. We also require $\mu < \inf_k \omega(k)$, to ensure the existence of the Gibbs measure also if we drop the nonlinearity, *i.e.*, if we set $\lambda = 0$.

For any realization of ψ_0 , let $\psi_t(x)$ be defined by the corresponding unique solution to the NLS, thus making it a random variable. Expectation values with respect to the finite volume, perturbed measure $\mathbb{P}_\Lambda^\lambda$ are denoted by $\mathbb{E}_\Lambda^\lambda$. $\mathbb{P}_\Lambda^\lambda$ has a well defined infinite volume limit. The zero coupling limit of these measures will be Gaussian and is denoted by $\mathbb{E}^0$. The covariance of the Gaussian measure has a Fourier transform $W(k) = \beta^{-1}(\omega(k) - \mu)^{-1}$.

Since the energy and norm are conserved, the Gibbs measure is time stationary. In addition, the dynamics and the Gibbs measure are invariant under periodic translations of Λ , and thus the stochastic process $(x, t) \mapsto \psi_t(x)$ is stationary jointly in space and time. The process also is also *gauge invariant*: ψ_0 and $e^{i\varphi} \psi_0$ have the same distribution for any $\varphi \in \mathbb{R}$, and the phase rotation commutes with the time-evolution. An immediate consequence of these invariance properties is the existence of a function $W_\Lambda^\lambda : \Lambda^* \times \mathbb{R} \rightarrow \mathbb{C}$ such that for all $k, k' \in \Lambda^*$, and $t', t \in \mathbb{R}$,

$$\mathbb{E}[\hat{\psi}_{t'}(k)^* \hat{\psi}_t(k')] = \delta_\Lambda(k - k') W_\Lambda^\lambda(k, t - t'). \quad (2)$$

In addition, $\mathbb{E}[\hat{\psi}_{t'}(k) \hat{\psi}_t(k')] = 0$ and all odd moments of the field vanish.

In the present work, we are concerned with the decay of $W_\Lambda^\lambda(k, t)$ for large L and small λ . For deterministic infinitely extended Hamiltonian systems, such as the NLS, establishing the qualitative behavior of equilibrium time correlations is known to be an extremely difficult problem with very few results available. For linear systems ($\lambda = 0$) one has an explicit solution, given by $W_\Lambda^\lambda(k, t) = W(k)e^{-it\omega(k)}$. If $\lambda > 0$, as general heuristics the nonlinearity should induce an exponential damping of $W_\Lambda^\lambda(t)$. The physical picture is based on excitations of wave modes which interact weakly and are damped through collisions. However, to mathematically establish such a time-decay is completely out of reach, at present. To make some progress we investigate the regime of small nonlinearity, $\lambda \ll 1$. The idea is not to aim for results which are valid globally in time, but rather to consider the first time scale on which the decay effect of the nonlinearity becomes visible. For small λ the rate of collision for two resonant waves is of order λ^2 . Therefore, the nonlinearity is expected to show up on a time scale λ^{-2} . This suggests to study the limit $W_\Lambda^\lambda(k, \lambda^{-2}t)$, as $\lambda \rightarrow 0$.

The limit $\lambda \rightarrow 0$, $t = \lambda^{-2}\tau$ with τ fixed, together with a possible rescaling of space by a factor λ^{-2} , is called *kinetic limit*, because the limit object is governed by a kinetic type transport equation. Erdős and Yau [2] have studied in great detail the kinetic limit of linear Schrödinger equation with a random potential. For nonlinear wave equations the only related study is by Benedetto, *et al.*, [3] on the dynamics of weakly interacting quantum particles. They use multipoint Wigner functions, which leads to an expansion somewhat different from the one used here. A comparison to their method can be found in [4].

2. Main results

We have to impose two types of assumptions. The first one requires a specific form of the clustering of the Gibbs measure, while the second collects conditions on the dispersion relation ω .

Assumption 2.1 (Equilibrium correlations). *Let $\beta > 0$ and $\mu < \inf_k \omega(k)$ be given. We assume that the corresponding Gibbs measure is ℓ_1 -clustering in the following sense: there exists $\lambda_0 > 0$ and $c_0 > 0$, independent of n , such that for $0 < \lambda \leq \lambda_0$ and all $n \geq 4$ one has the following bound for the fully truncated correlation functions (i.e., cumulants)*

$$\sup_{\Lambda, \sigma \in \{\pm 1\}^n} \sum_{x \in \Lambda^n} \mathbb{1}(x_1 = 0) \left| \mathbb{E}_\Lambda^\lambda \left[\prod_{i=1}^n \psi(x_i, \sigma_i) \right]^{\text{trunc}} \right| \leq \lambda(c_0)^n n!, \quad (3)$$

where $\psi(x, 1) = \psi(x)$, $\psi(x, -1) = \psi(x)^*$. We also assume a similar convergence of the two-point correlation function,

$$\limsup_{\Lambda \rightarrow \infty} \sum_{\|x\|_\infty \leq L/2} \left| \mathbb{E}_\Lambda^\lambda [\psi(0)^* \psi(x)] - \mathbb{E}^0 [\psi(0)^* \psi(x)] \right| \leq \lambda 2(c_0)^2. \quad (4)$$

In the recent preprint [5], Abdesselam, Procacci, and Scoppola prove the above bounds for a large class of coupling functions α .

For the main theorem we will need properties of the linear dynamics, $\lambda = 0$, which can be thought of as implicit conditions on ω , stated in the following.

Assumption 2.2 (Dispersion relation). Suppose $d \geq 4$, and $\omega : \mathbb{T}^d \rightarrow \mathbb{R}$ satisfies all of the following:

- (1) The periodic extension of ω is real-analytic and $\omega(-k) = \omega(k)$.
 (2) (ℓ_3 -dispersivity). Let $p_t(x) = \int_{\mathbb{T}^d} dk e^{i2\pi x \cdot k} e^{-it\omega(k)}$ denote the free propagator. We assume that there are $C, \delta > 0$ such that for all $t \in \mathbb{R}$,

$$\|p_t\|_3^3 = \sum_{x \in \mathbb{Z}^d} |p_t(x)|^3 \leq C \langle t \rangle^{-1-\delta}. \quad (5)$$

- (3) (constructive interference). There exists a set $M^{\text{sing}} \subset \mathbb{T}^d$ consisting of a union of a finite number of closed, one-dimensional, smooth submanifolds, and a constant C such that for all $t \in \mathbb{R}$, $k_0 \in \mathbb{T}^d$, and $\sigma \in \{\pm 1\}$,

$$\left| \int_{\mathbb{T}^d} dk e^{-it(\omega(k) + \sigma\omega(k-k_0))} \right| \leq \frac{C \langle t \rangle^{-1}}{d(k_0, M^{\text{sing}})}, \quad (6)$$

where $d(k_0, M^{\text{sing}})$ is the distance (on $\mathbb{T}^d$) of k_0 from M^{sing} .

- (4) (crossing bounds). Define for $t_0, t_1, t_2 \in \mathbb{R}$, $u_1, u_2 \in \mathbb{T}^d$, and $x \in \mathbb{Z}^d$,

$$K(x; t_0, t_1, t_2, u_1, u_2) = \int_{\mathbb{T}^d} dk e^{i2\pi x \cdot k} e^{-i(t_0\omega(k) + t_1\omega(k+u_1) + t_2\omega(k+u_2))}. \quad (7)$$

We assume that there is a measurable function $F^{\text{cr}} : \mathbb{T}^d \times \mathbb{R}_+ \rightarrow [0, \infty]$ so that constants $0 < \gamma \leq 1$, c_1, c_2 , for the following bounds can be found.

- (a) For any $u_i \in \mathbb{T}^d$, $\sigma_i \in \{\pm 1\}$, $i = 1, 2, 3$, and $0 < \beta \leq 1$, $n \in \{1, 2, 3\}$,

$$\begin{aligned} \int_{-\infty}^{\infty} dt \|p_t\|_3^2 \int_{-\infty}^{\infty} ds e^{-\beta|s|} \|K(t, \sigma_1 s, \sigma_2 s, u_1, u_2)\|_3 &\leq \beta^{\gamma-1} F^{\text{cr}}(u_2 - u_1; \beta), \\ \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} ds e^{-\beta|s|} \prod_{i=1}^3 \|K(t, \sigma_i s, 0, u_i, 0)\|_3 &\leq \beta^{\gamma-1} F^{\text{cr}}(u_n; \beta). \end{aligned} \quad (8)$$

- (b) For all $0 < \beta \leq 1$ we require $\int_{\mathbb{T}^d} dk F^{\text{cr}}(k; \beta) \leq c_1 \langle \ln \beta \rangle^{c_2}$, and if also $u, k_0 \in \mathbb{T}^d$, $\alpha \in \mathbb{R}$, $\sigma \in \{\pm 1\}$, and $n \in \{1, 2, 3\}$, and we denote $k = (k_1, k_2, k_0 - k_1 - k_2)$, then

$$\int_{(\mathbb{T}^d)^2} dk_1 dk_2 F^{\text{cr}}(k_n + u; \beta) \frac{1}{|\alpha - \Omega(k, \sigma) + i\beta|} \leq c_1 \langle \ln \beta \rangle^{1+c_2}, \quad (9)$$

where $\Omega : (\mathbb{T}^d)^3 \times \{\pm 1\} \rightarrow \mathbb{R}$ is defined by

$$\Omega(k, \sigma) = \omega(k_3) - \omega(k_1) + \sigma(\omega(k_2) - \omega(k_1 + k_2 + k_3)). \quad (10)$$

We have proven in Appendix A of [1] that the nearest neighbor interactions satisfy all of the above assumptions for $d \geq 4$.

We study the decay of W_Λ^λ via the quadratic form

$$Q_\Lambda^\lambda[g, f](t) = \mathbb{E}_\Lambda^\lambda \left[\langle \hat{f}, \hat{\psi}_0 \rangle^* \langle e^{-i\omega^\lambda t/\varepsilon} \hat{g}, \hat{\psi}_{t/\varepsilon} \rangle \right] = \int_{\Lambda^*} dk \hat{g}(k)^* \hat{f}(k) e^{i\omega^\lambda(k)t/\varepsilon} W_\Lambda^\lambda(k, t/\varepsilon),$$

where $\varepsilon = \lambda^2$, and $f, g \in \ell_2(\mathbb{Z}^d)$ are arbitrary test-functions with a finite support. In order to get a finite limit, it is necessary to cancel rapidly oscillating factors, encoded in the function ω^λ above. The oscillations are produced by the free evolution and first order corrections, and for us a suitable choice is given by

$$\omega^\lambda(k) = \omega(k) + \lambda R_0, \quad R_0 = R_0(\lambda, \Lambda) = 2\mathbb{E}_\Lambda^\lambda[|\psi_0(0)|^2]. \quad (11)$$

Theorem 2.1. *Consider the system described in Section 1, and let Assumption 2.1 and Assumption 2.2 hold. Then there is $t_0 > 0$ such that for all $|t| < t_0$, and for any $f, g \in \ell_2(\mathbb{Z}^d)$ with finite support,*

$$\lim_{\lambda \rightarrow 0} \limsup_{\Lambda \rightarrow \infty} \left| Q_\Lambda^\lambda[g, f](t) - \int_{\mathbb{T}^d} dk \hat{g}(k)^* \hat{f}(k) W(k) e^{-\Gamma_1(k)|t| - i\Gamma_2(k)t} \right| = 0, \quad (12)$$

where $\Gamma_j(k)$ are real, and $\Gamma(k) = \Gamma_1(k) + i\Gamma_2(k)$ is given by

$$\begin{aligned} \Gamma(k_1) = & -2 \int_0^\infty dt \int_{(\mathbb{T}^d)^3} dk_2 dk_3 dk_4 \delta(k_1 + k_2 - k_3 - k_4) \\ & \times e^{it(\omega_1 + \omega_2 - \omega_3 - \omega_4)} (W_3 W_4 - W_2 W_4 - W_2 W_3) \end{aligned} \quad (13)$$

with $\omega_i = \omega(k_i)$, $W_i = W(k_i)$.

By the earlier remarks, we know that the theorem holds at least for nearest neighbor interactions. (For the proof, and further discussion, we refer to our original paper [1].) Heuristically, the theorem states that for all not too large $t = \mathcal{O}(\lambda^{-2})$, we have $W_\Lambda^\lambda(k, t) \approx W(k) e^{-i\omega_{\text{ren}}^\lambda(k)t} e^{-|\lambda^2 t| \Gamma_1(k)}$, where $\omega_{\text{ren}}^\lambda(k) = \omega(k) + \lambda R_0 + \lambda^2 \Gamma_2(k)$. We point out that $\Gamma_1(k) \geq 0$, as by explicit computation

$$\Gamma_1(k_1) = 2\pi W(k_1)^{-2} \int_{(\mathbb{T}^d)^3} dk_2 dk_3 \delta(\omega_1 + \omega_2 - \omega_3 - \omega_4) \prod_{i=1}^4 W(k_i) \Big|_{k_4 = k_1 + k_2 - k_3}.$$

If $\Gamma_1(k) > 0$, then the term $\exp[-\Gamma_1(k)|t|]$ yields the exponential damping in $|t|$, both forward and backwards in time, and if $\Gamma_1(k) \geq \gamma > 0$ for all $k \in \mathbb{T}^d$, then on the kinetic scale the covariance has an exponential bound $e^{-\gamma|t|}$.

Acknowledgments

J. Lukkarinen was supported by the Academy of Finland.

References

- [1] J. Lukkarinen and H. Spohn, *Weakly nonlinear Schrödinger equation with random initial data*, preprint (2009), <http://arxiv.org/abs/0901.3283>.
- [2] L. Erdős and H.-T. Yau, *Commun. Pure Appl. Math.* **53**, 667 (2000).
- [3] D. Benedetto, F. Castella, R. Esposito and M. Pulvirenti, *Commun. Math. Phys.* **277**, 1 (2008).
- [4] J. Lukkarinen and H. Spohn, *J. Stat. Phys.* **134**, 1133 (2009).
- [5] A. Abdesselam, A. Procacci, and B. Scoppola, *Clustering bounds on n-point correlations for unbounded spin systems*, preprint (2009), <http://arxiv.org/abs/0901.4756>.

LIEB–ROBINSON BOUNDS AND THE EXISTENCE OF INFINITE SYSTEM DYNAMICS

BRUNO NACHTERGAELE

Department of Mathematics, University California, Davis

Davis, CA 95616, USA

E-mail: bxn@math.ucdavis.edu

www.math.ucdavis.edu/~bxn

We present a recent result on the existence of the dynamics in the thermodynamic limit of a class of anharmonic quantum oscillator lattices, which was obtained using Lieb–Robinson bounds.

Keywords: Lieb–Robinson bounds, propagation estimates, thermodynamic limit, quantum dynamics, anharmonic lattice

1. Introduction

In condensed matter physics, three common types of degrees of freedom are often found together: atoms positioned in a lattice, spin magnetic moments, and itinerant electrons. Many phenomena primarily involve only one of these and it has therefore proved useful to study them with separate models. Here, we are concerned with the spatial degrees of freedom of atoms in a lattice. The fact that the atoms oscillate about their equilibrium positions is essential for many important phenomena in condensed matter physics. The question we address is whether one can define the time evolution, consistent with the Schrödinger equation, for an infinite assembly of quantum oscillators such as a crystal lattice.

It is natural to start by considering the standard harmonic interaction, which should be an accurate description when the displacements from the equilibrium positions are small. It is well known, however, that some basic phenomena require that we consider anharmonic perturbations. Our approach applies to multi-body interactions that fall off sufficiently fast but for clarity and space limitations we will restrict ourselves to local and nearest neighbor anharmonicities here and refer to our forthcoming paper for the general case [11]. We will also limit the discussion to oscillators organized in a translation-invariant fashion on a lattice. To keep the notation simple we will work with one-dimensional oscillators at each lattice site, but this is not essential. With some straightforward modifications of the analysis our main result can be generalized to non-translation invariant models defined on a graph satisfying a few natural assumptions.

2. Harmonic lattice systems

Let Λ denote a finite subset of the ν -dimensional hypercubic lattice $\mathbb{Z}^\nu$ and define Hamiltonians H_Λ as self-adjoint operators on $\mathcal{H}_\Lambda = \bigotimes_{x \in \Lambda} L^2(\mathbb{R})$ by the following expression

$$H_\Lambda = \sum_{x \in \Lambda} \frac{1}{2m} p_x^2 + \frac{m}{2} \omega^2 q_x^2 + V(q_x) + \sum_{x, y \in \Lambda, |x-y|=1} \lambda (q_x - q_y)^2 + \Phi(q_x - q_y)$$

where $\omega, \lambda, \geq 0$, and p_x and q_x are the canonical momentum and position operators for the oscillator at $x \in \Lambda$. The Heisenberg dynamics, $\{\tau_t^\Lambda\}_{t \in \mathbb{R}}$, is defined by

$$\tau_t^\Lambda(A) = e^{itH_\Lambda} A e^{-itH_\Lambda}, \quad A \in \mathcal{B}(\mathcal{H}_\Lambda).$$

The algebra of observables is a tensor product

$$\mathcal{B}(\mathcal{H}_\Lambda) = \bigotimes_{x \in \Lambda} \mathcal{B}(\mathcal{H}_x) \equiv \mathcal{A}_\Lambda$$

such that for $X \subset \Lambda$, we have $\mathcal{A}_X \subset \mathcal{A}_\Lambda$, by identifying $A \in \mathcal{A}_X$ with $A \otimes \mathbb{I}_{\Lambda \setminus X} \in \mathcal{A}_\Lambda$. Therefore, for $A \in \mathcal{A}_X$, $\tau_t^\Lambda(A)$ is well-defined for all Λ containing X .

Our main concern is the existence of a limiting dynamics: do we have τ_t such that

$$\tau_t^\Lambda(A) \longrightarrow \tau_t(A), \text{ as } \Lambda \uparrow \mathbb{Z}^\nu$$

in a suitable sense? For an anharmonic lattice of classical oscillators a positive answer to this question was obtained by Lanford, Lebowitz and Lieb [4].

It is well-known that in the case of quantum spin systems (i.e., bounded local Hamiltonians) one can use Lieb-Robinson bounds to establish the existence of the thermodynamic limit [2]. The essential observation is as follows. Let Λ_n be an increasing exhausting sequence of finite volumes with Hamiltonians of the form

$$H_{\Lambda_n} = \sum_{X \subset \Lambda_n} \Phi(X)$$

where $\Phi(X) = \Phi(X)^* \in \mathcal{A}_X$. Then, for $n > m$, one easily derives the bound:

$$\|\tau_t^{\Lambda_n}(A) - \tau_t^{\Lambda_m}(A)\| \leq \sum_{x \in \Lambda_n \setminus \Lambda_m} \sum_{X \ni x} \int_0^{|t|} \|[\Phi(X), \tau_s^{\Lambda_m}(A)]\| ds. \quad (1)$$

Therefore, if we can show that the commutators in the integrand have sufficiently small norms, it will follow that the finite-volume dynamics form a Cauchy sequence. Estimates for such commutators were first derived by Lieb and Robinson [5]. For $A \in \mathcal{A}_X$ and $B \in \mathcal{A}_Y$, they proved a bound of the form

$$\|[\tau_t(A), B]\| \leq C e^{-a(d(X,Y) - v|t|)},$$

where C, a , and v are positive constants and $d(X, Y)$ denotes the distance between X and Y . Estimates of this type are now commonly referred to as Lieb-Robinson bounds [3, 7–9]. For anharmonic lattice systems Lieb-Robinson bounds were recently proved in Ref. 10, and this work builds on the results obtained there.

For the oscillator lattices with $\lambda \neq 0$, the approach suggested by Eq. 1 does not work due to the unboundedness of $(q_x - q_y)^2$. But we note that for the *harmonic* lattice, τ_t^Λ can be calculated explicitly and can also be defined for the infinite lattice. One then sees, however, that $\tau_t^\Lambda(A)$ cannot converge in norm. The idea is to consider the infinite anharmonic system as a limit of finite-volume perturbations of the infinite harmonic system. We mention that, with a different approach, Amour, Levy-Bruhl, and Nourrigat have recently obtained convergence results for certain models by introducing suitable Sobolev-like norms for the observables [1].

Rigorous perturbation theory of infinite systems is available if the unperturbed infinite system dynamics has a suitable continuity property [2]. In our case this will be continuity for the weak operator topology, which follows from known properties and the exact solution of the harmonic lattice.

Up to a redefinition of the parameters, the harmonic lattice model Hamiltonian on a finite subset $\Lambda \subset \mathbb{Z}^\nu$, is

$$H_\Lambda = \sum_{x \in \Lambda} p_x^2 + \omega^2 q_x^2 + \sum_{|x-y|=1} \lambda (q_x - q_y)^2$$

acting on $\mathcal{H}_\Lambda$. The creation and annihilation operators are defined by

$$a_x = \frac{1}{\sqrt{2}} (q_x + ip_x) \quad \text{and} \quad a_x^* = \frac{1}{\sqrt{2}} (q_x - ip_x),$$

which satisfy the Canonical Commutation Relations (CCR):

$$[a_x, a_y] = [a_x^*, a_y^*] = 0 \quad \text{and} \quad [a_x, a_y^*] = \delta_{x,y} \quad \text{for all } x, y \in \Lambda_L$$

H_Λ is a quadratic expression in a_x, a_x^* , which can be diagonalized by a Bogoliubov transformation [6]. This is usually done in Fourier space, resulting in:

$$H_\Lambda = \sum_{k \in \Lambda^*} \gamma(k) (2b_k^* b_k + 1)$$

where $\gamma(k) = (\omega^2 + 4\lambda \sum_{j=1}^\nu \sin^2(k_j/2))^{1/2}$. The b_k^* are creation operators for the eigenmodes of the system, which also satisfy the CCR, and the ground state of H_Λ is the corresponding vacuum. The time evolution is simply given by $\tau_t(b_k^*) = e^{-2i\gamma(k)t} b_k^*$. It is useful for us to express the Bogoliubov transformations in real space: for each $f : \Lambda_L \rightarrow \mathbb{C}$, set

$$a(f) = \sum_{x \in \Lambda_L} \overline{f(x)} a_x, \quad a^*(f) = \sum_{x \in \Lambda_L} f(x) a_x^*.$$

and similarly for the b 's. Then, there are real-linear operators U and V such that

$$a(f) = b(U^* f) - b^*(V^* f), \quad a^*(f) = b^*(U^* f) - b^*(V^* f).$$

Using these relations, one finds a one-parameter group, T_t , such that

$$\tau_t(a(f)) = \tau_t(b(U^* f) - b^*(V^* f)) = a(T_t f), \quad \tau_t(a^*(f)) = a^*(T_t f).$$

Since the canonical operators (p 's, q 's, a 's, a^* 's, ...) are all unbounded, we will derive Lieb-Robinson bounds for the Weyl operators instead:

$$W(f) = \exp \left[\frac{i}{\sqrt{2}} (a(f) + a^*(f)) \right],$$

This can be done very explicitly by using the relations

$$\tau_t(W(f)) = W(T_t f), \text{ and } W(f)W(g) = e^{i\sigma(f,g)} W(f+g)$$

where $\sigma(f, g) = \text{Im} \langle f, g \rangle$. Observe that

$$\begin{aligned} [\tau_t(W(f)), W(g)] &= \{W(T_t f) - W(g)W(T_t f)W(-g)\} W(g) \\ &= \left\{1 - e^{-2i\sigma(T_t f, g)}\right\} W(T_t f)W(g). \end{aligned}$$

Since the Weyl operators are unitary, we therefore have $\|[\tau_t(W(f)), W(g)]\| \leq 2|\sigma(T_t f, g)|$. So, all we have to do is to estimate $\sigma(T_t f, g)$. $T_t f$ is explicitly given by $T_t f = f * \overline{h_{1,t}^{(L)}} + \bar{f} * h_{2,t}^{(L)}$, where the functions $h_{1,t}$ and $h_{2,t}$ are given by

$$\begin{aligned} h_{1,t}^{(L)}(x) &= \frac{i}{2} \text{Im} \left[\frac{1}{|\Lambda_L|} \sum_{k \in \Lambda_L^*} (\gamma(k) + \gamma(k)^{-1}) e^{ik \cdot x - 2i\gamma(k)t} \right] + \text{Re} \left[\frac{1}{|\Lambda_L|} \sum_{k \in \Lambda_L^*} e^{ik \cdot x - 2i\gamma(k)t} \right] \\ h_{2,t}^{(L)}(x) &= \frac{i}{2} \text{Im} \left[\frac{1}{|\Lambda_L|} \sum_{k \in \Lambda_L^*} (\gamma(k) - \gamma(k)^{-1}) e^{ik \cdot x - 2i\gamma(k)t} \right]. \end{aligned}$$

By analyzing these functions one then shows that for every $a > 0$, there exist c_a and v_a such that

$$|\sigma(T_t f, g)| \leq c_a e^{v_a |t|} \sum_{x,y} |f(x)| |g(y)| e^{-ad(x,y)}.$$

Theorem 2.1 (Ref. 10). *Let $\lambda, \omega \geq 0$. Then, for all f, g with $\text{supp } f \subset X$ and $\text{supp } g \subset Y$,*

$$\left\| [\tau_t(W(f)), W(g)] \right\| \leq C \sum_{x,y} |f(x)| |g(y)| e^{-2(d(x,y) - v|t|)},$$

with $v = 6\sqrt{\omega^2 + 4\nu\lambda}$.

3. Anharmonic lattice systems

In Ref. 10 we proved Lieb-Robinson bounds for finite systems, with estimates uniform in the volume. With the same approach one can also prove bounds for harmonic infinite systems with an anharmonic perturbation in finite volume, i.e., a model with formal Hamiltonian of the form:

$$H_\Lambda = \sum_{x \in \mathbb{Z}^\nu} p_x^2 + \omega^2 q_x^2 + \sum_{x,y \in \mathbb{Z}^\nu, |x-y|=1} \lambda(q_x - q_y)^2 + \sum_{X \subset \Lambda} \Phi(X).$$

Typical Φ are $\Phi(\{x\}) = V(q_x)$, $\Phi(\{x, y\}) = W(q_x - q_y)$, etc. but more general perturbations can be considered [11].

For the harmonic system, we can directly see how the exact expressions extend to the infinite system. For $f : \mathbb{Z}^\nu \rightarrow \mathbb{C}$ in a suitable function space ($\ell^1(\mathbb{Z}^\nu)$, or $\ell^2(\mathbb{Z}^\nu)$), it is straightforward to define

$$\tau_t(W(f)) = W(T_t f)$$

We get a Hilbert space representation by representing the Weyl operators on the Fock space generated by the $b^*(f)$ operators acting on the vacuum. On this space we have well-defined Hamiltonian such that

$$\tau_t(W(f)) = e^{itH} W(f) e^{-itH}$$

τ_t is the dynamics for the formal Hamiltonian

$$H = \sum_{x \in \mathbb{Z}^\nu} p_x^2 + \omega^2 q_x^2 + \sum_{|x-y|=1} \lambda (q_x - q_y)^2$$

and the Lieb-Robinson bounds continue to hold.

For simplicity, consider the perturbation of the form $P_\Lambda = \sum_{x \in \Lambda} V(q_x)$. Then, the perturbed dynamics, formally corresponding to $H + P_\Lambda$, and can be defined mathematically by the Dyson series:

$$\tau_t^{(\Lambda)}(W(f)) = \tau_t(W(f)) + \sum_{n=1}^{\infty} i^n \int_{0 \leq t_1 \leq t_2 \leq \dots \leq t} dt_1 \dots dt_n [\tau_{t_n}(P_\Lambda), [\dots [\tau_{t_1}(P_\Lambda), \tau_t(W(f))]]]$$

We have the following Lieb-Robinson bounds for τ_t^Λ .

Theorem 3.1 (Ref. 10). *Let $\lambda \geq 0, \omega > 0$, and V such that $\|k^2 \hat{V}(k)\|_1 < \infty$. Then, for all $f, g \in \ell^1(\mathbb{Z}^\nu)$, we have*

$$\left\| [\tau_t^\Lambda(W(f)), W(g)] \right\| \leq C \sum_{x,y} |f(x)| |g(y)| e^{-2(d(x,y)-v|t|)}$$

with

$$v = 6\sqrt{\omega^2 + 4\nu\lambda} + c\|k^2 \hat{V}(k)\|_1.$$

To show convergence, we estimate $\|\tau_t^{\Lambda_n}(W(f)) - \tau_t^{\Lambda_m}(W(f))\|$, for $\Lambda_m \subset \Lambda_n$ by considering $\tau_t^{\Lambda_n}$ as a perturbation of $\tau_t^{\Lambda_m}$. This gives

$$\tau_t^{\Lambda_n}(W(f)) = \tau_t^{\Lambda_m}(W(f)) + i \int_0^t \tau_s^{\Lambda_n} \left([P_{\Lambda_n \setminus \Lambda_m}, \tau_{t-s}^{\Lambda_m}(W(f))] \right) ds,$$

Therefore

$$\left\| \tau_t^{\Lambda_n}(W(f)) - \tau_t^{\Lambda_m}(W(f)) \right\| \leq \sum_{z \in \Lambda_n \setminus \Lambda_m} \int_0^{|t|} \left\| [V(q_z), \tau_{|t|-s}^{\Lambda_m}(W(f))] \right\| ds$$

By writing $V(q_x) = \int \hat{V}(p) W(p\delta_x) dp$, $W(p\delta_x) = e^{ipq_x}$, we can then use Theorem 3.1 for $\tau_t^{\Lambda_m}$ to obtain the convergence. The convergence is uniform on intervals $[-t_0, t_0]$. We also immediately get continuity in t of the limiting dynamics by an $\epsilon/3$ argument.

Theorem 3.2 (Ref. 11). Assume $\omega > 0$, $\|k\hat{V}(k)\|_1 < \infty$, $\|k^2\hat{V}(k)\|_1 < \infty$. For all $f \in \ell^1(\mathbb{Z}^\nu)$, and all $t \in \mathbb{R}$, the limit

$$\lim_{\Lambda \uparrow \mathbb{Z}^\nu} \tau_t^\Lambda(W(f)) = \tau_t^\infty(W(f))$$

converges in the operator norm topology and the resulting the dynamics is continuous in t in the weak operator topology.

Acknowledgments

The work reported on in this paper was supported by the National Science Foundation under Grant # DMS-0605342 and FRG grants DMS-0757581, DMS-0757424, and DMS-0757327.

References

- [1] L. Amour, P. Levy-Bruhl, and J. Nourrigat, *Dynamics and Lieb-Robinson Estimates for Lattices of Interacting Anharmonic Oscillators.*, arXiv:0904.2717.
- [2] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics. Volume 2.*, 2nd Edition (Springer-Verlag, 1997).
- [3] M. Hastings and T. Koma, Commun. Math. Phys. **265**, 781–804 (2006).
- [4] O.E. Lanford, J. Lebowitz, E. H. Lieb, J. Statist. Phys. **16**, 453–461 (1977).
- [5] E.H. Lieb and D.W. Robinson, Comm. Math. Phys. **28**, 251–257 (1972).
- [6] J. Manuceau and A. Verbeure, Commun. Math. Phys. **9**, 293–302 (1968).
- [7] B. Nachtergaele and R. Sims, Commun. Math. Phys. **265**, 119–130 (2006).
- [8] B. Nachtergaele, Y. Ogata, and R. Sims, J. Stat. Phys. **124**, 1–13 (2006).
- [9] B. Nachtergaele and R. Sims. *Locality Estimates for Quantum Spin Systems*. To appear in: Sidoravicius, Vidas (Ed.), *New Trends in Mathematical Physics. Selected contributions of the XVth International Congress on Mathematical Physics*, Springer Verlag, 2009, pp 591–614. arXiv:0712.3318.
- [10] B. Nachtergaele, H. Raz, B. Schlein, and R. Sims, Commun. Math. Phys. **286**, 1073–1098 (2009).
- [11] B. Nachtergaele, B. Schlein, R. Sims, S. Starr, and V. Zagrebnov, *On the existence of the dynamics for anharmonic quantum oscillator systems*, arXiv:0909.2249.

HEAT EQUATION FROM MICROSCOPIC DYNAMICS: A WEAK COUPLING APPROACH

C. LIVERANI

*Dipartimento di Matematica, II Università di Roma,
Roma, 00133, Italy
E-mail: liverani@mat.uniroma2.it*

S. OLLA

*CEREMADE, UMR CNRS 7534
Université Paris Dauphine
Paris Cedex, 75775 France
E-mail: olla@ceremade.dauphine.fr*

We consider the dynamics of weakly coupled anharmonic oscillators, perturbed by a energy conserving noise. The noise considered exchanges independently the components of the momentum of each oscillator. We prove that in the weak-coupling/long time limit energies of oscillators evolve following stochastic autonomous system of equations. Under a further diffusive space-time scaling of this system, the energy evolves following a non-linear heat equation.

1. The quest for heat equation, from the microscopic dynamics

The mathematical rigorous deduction, from microscopic molecular dynamics, of heat equation, or of its stationary version called Fourier's law, is one of the main open problem in non equilibrium statistical mechanics [1].

To state the problem we can consider a chain of coupled anharmonic oscillators. This is described by the Hamiltonian

$$\begin{aligned} \mathcal{H}^\Lambda &= \sum_{i \in \Lambda} \left[\frac{|p_i|^2}{2} + \frac{1}{2} \sum_{j: |i-j|=1} V(q_i - q_j) + U(q_i) \right] \\ &= \sum_{i \in \Lambda} e_i \end{aligned} \tag{1}$$

where $\Lambda \subset \mathbb{Z}^d$, q_i is the displacement of the oscillator i from its equilibrium position, and p_i its momentum (we fix masses equal to 1), U is the anharmonic potential, and V is the coupling potential. The Hamiltonian deterministic dynamics is as usual

defined by the solution of

$$\dot{q}_i = p_i, \quad \dot{p}_i = -\partial_{q_i} \mathcal{H}^\Lambda \quad (2)$$

Then the mathematical problem can be stated, in a bit imprecise way, for the macroscopic evolution of the energy in a diffusive space-time scale (for simplicity of notation we consider here the case $d = 1$):

Find a class of initial conditions such that, as $|\Lambda| = N \rightarrow \infty$, we have, for any continuous test function G with compact support,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_i G(i/N) e_i(N^2 t) = \int G(y) T(y, t) dy, \quad (3)$$

where $T(y, t)$ is the solution of the nonlinear heat equation

$$\partial_t T = \nabla \kappa(T) \nabla T \quad (4)$$

The thermal conductivity $\kappa(T)$ is defined by the Green-Kubo formula, i.e. can be obtained by the limit

$$\kappa(T) = \lim_{t \rightarrow \infty} \frac{1}{2T^2 t} \sum_i i^2 (\langle e_i(t) e_0(0) \rangle_T - \langle e_0(0) \rangle_T^2) \quad (5)$$

where $\langle \cdot \rangle_T$ stands here for the expectation of the infinite system in equilibrium at temperature T (i.e. starting with the corresponding Gibbs measure). By using space and time translation invariance of $\langle \cdot \rangle_T$, thermal conductivity can be rewritten as

$$\kappa(T) = \frac{1}{2T^2} \int_0^\infty \sum_i \langle j_{i,i+1}(t) j_{0,1}(0) \rangle_\beta dt \quad (6)$$

Here $j_{i,i+1}$ are the energy currents defined by

$$\dot{e}_i(t) = (j_{i-1,i}(t) - j_{i,i+1}(t)) \quad (7)$$

and explicitly given by

$$j_{i,i+1} = -\frac{1}{2} (p_{i+1} + p_i) \cdot \nabla V(q_{i+1} - q_i). \quad (8)$$

At the moment this problem is completely open, and it is not even clear what could be the class of initial conditions under which it holds. Even the definition of the thermal conductivity $\kappa(T)$ is a challenging problem, since it requires information on the decay of correlations for an (infinite) deterministic system.

2. Energy Conserving Noise

It is clear that adding some noise to the dynamics should help. The nature of the noise should be such that it conserves energy, is local, and should give enough mixing properties to the dynamics to guarantee the convergence of the integrals in (6). We choose the noise as random forces that conserve the kinetic energy of each atom, given by independent diffusions on the spheres $\|p_i\|^2 = \text{const}$. We consider the case

where oscillators have two degree of freedom. In order to define such diffusions, consider the vector fields

$$X_i := p_i^1 \partial_{p_i^2} - p_i^2 \partial_{p_i^1} =: J p_i \cdot \partial_{p_i},$$

and the second order operator

$$S = \sum_{i \in \Lambda} X_i^2$$

The generator of the process we are interested in is then given by

$$L_\Lambda =: A + \sigma^2 S$$

where $A = \{H^\Lambda, \cdot\}$, the usual Hamiltonian operator. Notice that the effect of this random perturbation does not change the energy currents between oscillators.

Even with such noise the above mathematical problem is open and very challenging. What is clear in this case is that initial conditions will matter little, and the theorem should be valid for a large class of initial conditions.

3. The weak coupling limit

One way to attack the problem is to perform first a weak-coupling/large times limit of the system. Multiplying the coupling interaction V by a small parameter ε , we obtain energy current of size ε , and looking at time of order $\varepsilon^{-2}t$ we have for the evolution of the energies:

$$e_i(\varepsilon^{-2}t) - e_i(0) = \sum_{|k-i|=1} \varepsilon \int_0^{\varepsilon^{-2}t} j_{i,k}(s) ds \quad (9)$$

So it is clear that it is involved a central limit theorem problem for these very degenerate Markov processes. Here is our result:

Theorem 3.1. (*C. Liverani, S. Olla, in preparation*) *In the limit $\varepsilon \rightarrow 0$, the process $\{e_i(\varepsilon^{-2}t)\}_{i \in \Lambda}$ converges, in law, to the stochastic process $\{\mathcal{E}_i\}_{i \in \Lambda}$ determined by the stochastic differential equations*

$$d\mathcal{E}_i = \sum_{|i-k|=1} \alpha(\mathcal{E}_i, \mathcal{E}_k) dt + \sum_{|i-k|=1} \gamma(\mathcal{E}_i, \mathcal{E}_k) dB_{\{i,k\}} \quad (10)$$

where $B_{\{i,k\}} = -B_{\{k,i\}}$ are independent standard Brownian motions.

The coefficients α and γ are related by

$$\alpha(a_1, a_2) = -\alpha(a_2, a_1) = \Gamma(\mathbf{a})^{-1} (\partial_{a_1} - \partial_{a_2}) [\Gamma(\mathbf{a}) \gamma(a_1, a_2)^2],$$

with

$$\Gamma(\mathbf{a}) = \prod_j Z(a_j)$$

and $Z(a)$ is a explicit function depending of U .

We have that the generator of this dynamics can be written as

$$\tilde{L} = \frac{1}{2\Gamma(\mathbf{a})} \sum_{i,j:|i-j|=1} (\partial_{a_i} - \partial_{a_j}) [\Gamma(\mathbf{a}) \gamma(a_i, a_j)^2 (\partial_{a_i} - \partial_{a_j})]$$

In order to understand the relevance of the above result, consider the harmonic case (U and V quadratic). In this case these coefficient can be explicitly computed

$$\gamma(a_i, a_j)^2 = ca_i a_j, \quad \alpha(a_i, a_j) = -3c(a_i - a_j)$$

with c a constant depending on the quadratic potentials and the corresponding stochastic differential equations became

$$\begin{aligned} d\mathcal{E}_i(t) &= 3c \sum_{|i-j|=1} -(\mathcal{E}_i(t) - \mathcal{E}_j(t)) dt + \sum_{|i-j|=1} \sqrt{c\mathcal{E}_i(t)\mathcal{E}_j(t)} dB_{\{i,j\}}(t) \\ &= 3c\Delta\mathcal{E}_i(t) dt + \sum_{|i-j|=1} \sqrt{c\mathcal{E}_i(t)\mathcal{E}_j(t)} dB_{\{i,j\}}(t) \end{aligned}$$

i.e. we have already heat equation in this time-scale, plus a diffusive term. In the non-linear case, $\alpha(a_i, a_j)$ is not a difference, and this dynamics falls in the class of conservative stochastic dynamics called *non-gradient Ginzburg-Landau models*.

4. The hydrodynamic limit

The process (10) is close the the one studied by Varadhan in [2]. In this paper Varadhan proves an *hydrodynamic limit*, i. e. that under certain condition on the initial distribution, for any test function G on $\mathbb{R}$ we have the convergence

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_i G(i/N) \mathcal{E}_i(N^2 t) = \int G(y) u(y, t) dy \quad (11)$$

where $u(y, t)$ is the solution of a nonlinear heat equation

$$\partial_t u = \nabla D(u) \nabla u \quad (12)$$

where D is given by a Green-Kubo formula similar to (6), but involving the *homogenized currents* $\alpha(a_{i+1}, a_i)$ instead of the microscopic currents $j_{i,i+1}$.

Yet our case it is not covered by such result (due to the degeneracy at zero of the diffusion coefficients and the non strict convexity of the potential of the invariant measure). In any case the extension of Varadhan's work to the present case would allow to obtain the *heat equation* in the present setting via a diffusive limit. We plan to work on such an extension in the future.

References

- [1] R.-B. F. Bonetto, J.L. Lebowitz, Fourier's law: A challenge to theorists, in *Mathematical Physics 2000*, (Imperial College Press, London,, 2000).
- [2] S. Varadhan, Nonlinear diffusion limit for a system with nearest neighbor interactions. ii, in *Asymptotic problems in probability theory: stochastic models and diffusions on fractals*, (Sanda/Kyoto, 1990), ed. H. Longman Sci. Tech., Pitman Res. Notes Math. Ser., Vol. 2831993.

KPZ UNIVERSALITY FOR KPZ

JEREMY QUASTEL

*Departments of Mathematics and Statistics, University of Toronto,
Toronto, ON M5S 1L2, Canada
E-mail: quastel@math.toronto.edu
www.math.toronto.edu/quastel*

Bertini and Giacomin [1] propose the logarithm $h(t, x) = -\log Z(t, x)$ of the solution stochastic heat equation as the solution to the Kardar-Parisi-Zhang equation. We survey recent work with Balázs and Seppäläinen showing that various quantities have the conjectured dynamic scaling exponent $z = 2/3$. We also present a slightly different version of the computation in [1] suggesting that h solves a version with Wick-ordered nonlinearity.

Keywords: Kardar-Parisi-Zhang equation; stochastic Burgers equation; scaling exponent; asymmetric simple exclusion.

1. Introduction

Despite considerable effort we are still lacking a satisfactory mathematical formulation of the Kardar-Parisi-Zhang equation,

$$\partial_t h = -\frac{1}{2}(\partial_x h)^2 + \frac{1}{2}\partial_x^2 h + \dot{W}, \quad t > 0, \quad x \in \mathbb{R} \quad (1)$$

for the height of a randomly growing interface, or the equivalent stochastic Burgers equation for $u = \partial_x h$,

$$\partial_t u = -\frac{1}{2}\partial_x u^2 + \frac{1}{2}\partial_x^2 u + \partial_x \dot{W}, \quad t > 0, \quad x \in \mathbb{R}. \quad (2)$$

Here $\dot{W}(t, x)$ is Gaussian space-time white noise.

The best methods at the present time in stochastic partial differential equations [2] can only handle smoother noises. Often (2) is studied without the derivative on the white noise, or with higher order operators than the Laplacian, in order to make it tractable [2], [3]. In [4] a version of (2) is studied where the non-linear term is replaced by an object denoted there by $u \diamond \partial_x u$. However, the resulting solutions, obtained by explicit Wiener chaos computations, do not [5] have the scaling predicted in the physical theory, which is a dynamic scaling exponent [6], [7]

$$z = 3/2. \quad (3)$$

A different approach of Bertini and Giacomin [1] is to simply define

$$h(t, x) = -\log Z(t, x) \quad (4)$$

where $Z(t, x)$ solves the stochastic heat equation

$$\partial_t Z = \partial_x^2 Z - Z \dot{W}, \quad t > 0, \quad x \in \mathbb{R}. \quad (5)$$

Unlike (1) and (2), (5) is known to be well-posed [8] with solutions continuous functions of t and x almost surely. To connect with (1) by the Hopf-Cole transformation, let $\mathcal{G}_\kappa(x) = \frac{1}{\sqrt{2\pi\kappa^2}} \exp\{-x^2/2\kappa^2\}$ and smooth out the white noise in space,

$$\dot{W}^\kappa(t, x) = \int \mathcal{G}_\kappa(x - y) \dot{W}(t, y) dy.$$

Alternatively $\dot{W}^\kappa(t, x)$ is the Gaussian field with covariance $E[\dot{W}^\kappa(t, x) \dot{W}^\kappa(s, y)] = C_\kappa(x - y) \delta(t - s)$ where $C_\kappa(x - y) = \kappa^{-1} \int \mathcal{G}_\kappa(x - u) \mathcal{G}_\kappa(y - u) du$. In particular,

$$C_\kappa(0) = \frac{1}{2} \kappa^{-1} (\pi)^{-1/2}. \quad (6)$$

Let $Z^\kappa(t, x)$ be the solution of the stochastic heat equation with the regularized noise

$$\partial_t Z^\kappa = \partial_x^2 Z^\kappa - Z^\kappa \dot{W}^\kappa, \quad t > 0, \quad x \in \mathbb{R}, \quad (7)$$

and define

$$h^\kappa(t, x) = -\log Z^\kappa(t, x). \quad (8)$$

The $h^\kappa(t, x)$ converge to $h(t, x) = -\log Z(t, x)$ and by Itô's formula,

$$\partial_t h^\kappa = -\frac{1}{2} [(\partial_x h^\kappa)^2 - C_\kappa(0)] + \frac{1}{2} \partial_x^2 h^\kappa + \dot{W}^\kappa, \quad (9)$$

the difference from (1) being the renormalization of the non-linear term.

2. Scaling exponent $z = 3/2$ for $h = -\log Z$

In joint work with M. Balázs and T. Seppäläinen [9] we consider the stochastic heat equation with $Z(0, x) = \exp\{-B(x)\}$ where $B(x)$ is a two-sided Brownian motion independent of $\dot{W}$. This corresponds to (2) being in equilibrium. (1) is almost in equilibrium except that $h(t, 0)$ is growing in time. $h(0, x) = -\log Z(0, x) = B(x)$ is statistically invariant under the transformation $h(0, x) \mapsto \epsilon^{1/2} h(0, \epsilon^{-1} x)$. Dynamical scaling exponent z means that we find a non-trivial behavior of the solution under the scaling

$$\epsilon^{1/2} h(\epsilon^{-z} t, \epsilon^{-1} x). \quad (10)$$

We are only able to study reduced quantities such the variance $\text{Var}(h(t, x))$ for $h(t, x) = \log Z(t, x)$, and the space-time correlations $S(t, x) = E[u(t, x); u(0, 0)]$ of $u(t, x) = \partial_x h(t, x)$. In fact, we only prove that for each $t \geq 0$, it is a probability measure in x , which we call $S(t, dx)$. Since $h(t, x)$ is a Brownian motion in x , $u(t, x)$ is a distribution, and it is actually surprising that the correlation function makes any sense at all. It does because in the distributional sense

$$S(t, dx) = \partial_x^2 \text{Var}(h(t, x)). \quad (11)$$

For each $t \geq 0$, $\text{Var}(h(t, x))$ is a symmetric function of x which approaches $|x|$ for large $|x|$ and never lies below $|x|$. The main result of [9] is that there exist $c_0, C_1, C_2 < \infty$, such that for $t \geq c_0$,

$$C_1 t^{2/3} \leq \text{Var}(h(t, 0)) \leq C_2 t^{2/3} \quad (12)$$

and, for $1 < m < 3$,

$$C_1 t^{2m/3} \leq \int |x|^{m-2} [\text{Var}(h(t, x)) - |x|] dx \leq C_2 t^{2m/3} \quad (13)$$

and

$$C_1 t^{2m/3} \leq \int |x|^m S(t, dx) \leq C_2 t^{2m/3}. \quad (14)$$

To a certain extent this verifies the scaling prediction $z = 2/3$ and adds strength to the assertion that $h = -\log Z$ is the relevant solution of (1).

3. Weakly asymmetric simple exclusion

The proof is via approximation by particle systems. We use the nearest neighbour exclusion process on $\mathbb{Z}$ with particles attempting jumps to the right at rate $1/2$ and to the left at rate $1/2 + \epsilon^{1/2}$. $\hat{\eta}(t, x) = 1$ or -1 depending on whether or not there is a particle at $x \in \mathbb{Z}$ at time t , and this can also be thought of as the slope of the corresponding height function

$$\zeta_\epsilon(t, x) = \begin{cases} \sum_{0 < y \leq x} \hat{\eta}(t, y) - 2N(t, 0), & x \geq 0, \\ -\sum_{x < y \leq 0} \hat{\eta}(t, y) - 2N(t, 0), & x < 0, \end{cases} \quad (15)$$

where $N(t, 0)$ is the number of particles which jump from 0 to 1 minus the number of particles which jump from 1 to 0 in the time interval $[0, t]$. As initial data we take independent Bernoulli with density $1/2$. For $x \in \mathbb{R}$ and $t \geq 0$ let $h_\epsilon(t, x)$ denote the rescaled height function:

$$h_\epsilon(t, x) = \epsilon^{1/2} (\zeta_\epsilon(\epsilon^{-2}t, \lfloor \epsilon^{-1}x \rfloor) - v_\epsilon t) \quad (16)$$

where $v_\epsilon = \frac{1}{2}\epsilon^{-3/2} - \frac{1}{4!}\epsilon^{-1/2}$. It is also shown in [1] that as $\epsilon \searrow 0$, $h_\epsilon(t, x)$ converge to $h(t, x) = -\log Z(t, x)$. To see that this is the right scaling, note that (1) is invariant under $h(t, x) \mapsto \epsilon^{1/2}h(\epsilon^{-2}t, \epsilon^{-1}x)$ together with a multiplication of the nonlinear term by $\epsilon^{1/2}$.

The correlation function $S_\epsilon(t, x) = E[\hat{\eta}(t, x); \hat{\eta}(0, 0)]$ corresponds exactly to the probabilities $P(\mathbf{x}(t) = x)$ of a *second class particle* $\mathbf{x}(t)$ starting at 0, i.e. the discrepancy between two coupled copies of the process starting in the same state except at 0, where one starts with site 0 occupied, and the other with site 0 unoccupied. When $\epsilon = 0$ the process is symmetric and the second class particle is diffusive with $|\mathbf{x}(t)| \sim t^{1/2}$. When $\epsilon > 0$ the process is superdiffusive with $|\mathbf{x}(t)| \sim t^{2/3}$. Our main result follows from the transition between the behaviors.

Theorem 3.1. For $1 \leq m < 3$, and $t \geq c_0 \epsilon^{-2}$,

$$C_1 \epsilon^{m/3} t^{2m/3} \leq E_\epsilon [|\mathbf{x}(t)|^m] \leq C_2 \epsilon^{m/3} t^{2m/3}. \quad (17)$$

The proof follows earlier bounds of [10] for general nearest neighbour exclusion, with refinements to obtain the fine behaviour as $\epsilon \searrow 0$. The idea at its most basic is as follows. We couple the exclusion η with density $1/2$ with an exclusion ζ with density $1/2 - \epsilon^{-1/6} t^{-1/3}$. The excess particles in η over ζ move as second class particles with exclusion among themselves, one of which starts at the origin and is our $\mathbf{x}(t)$. Because of the preservation of ordering, the probability that $\mathbf{x}(t) \leq -\epsilon^{1/3} t^{2/3}$ can be controlled by the probability that the current $J_{-\epsilon^{1/3} t^{2/3}}(t)$ of these second class particles crossing the space-time line between $(0, 0)$ and $(t, -\epsilon^{1/3} t^{2/3})$ is less than $\epsilon^{1/6} t^{1/3}$. There is an identity that links the variance of the current in any exclusion process with density ρ back to the first moment of the second class particle:

$$\text{Var}^\rho(J_x) = \rho(1 - \rho) E^\rho[|\mathbf{x}(t) - x|]. \quad (18)$$

Hence by Chebyshev's inequality, one can get an estimate for the first moment of the second class particle in terms of itself. This gives the upper bound.

4. Renormalized KPZ equation for $h = -\log Z$

Let $Z(t, x)$ be the solution of (5) and

$$Z_\kappa(t, x) = \langle \mathcal{G}_{x, \kappa}, Z(t) \rangle = \int \mathcal{G}_\kappa(x - y) Z(t, y) dy. \quad (19)$$

Define $h_\kappa(t, x) = -\log Z_\kappa(t, x)$. Then by Itô's formula,

$$\partial_t h_\kappa + \frac{1}{2} (\partial_x h_\kappa)^2 - \frac{1}{2} \partial_x^2 h_\kappa - \dot{W} = \left\{ Z_\kappa^{-1} \langle \mathcal{G}_{x, \kappa} Z, \dot{W} \rangle - \dot{W} \right\} + \frac{1}{2} Z_\kappa^{-2} \langle \mathcal{G}_{x, \kappa}^2, Z^2 \rangle. \quad (20)$$

For the first term, we can compute $E[l(\int \int \varphi(t, x) \{Z_\kappa^{-1} \langle \mathcal{G}_{x, \kappa} Z, \dot{W} \rangle - \dot{W}\} dx dt)^2]$ for a smooth function φ of compact support by Itô isometry to get

$$\int \int E[(\int \frac{\varphi(t, y) \mathcal{G}_\kappa(y - x) dy}{\int \mathcal{G}_\kappa(y - y') Z(t, y') dy'} Z(t, x) - \varphi(t, x))^2] dx dt, \quad (21)$$

which vanishes as $\kappa \searrow 0$ by the continuity of $Z(t, x)$.

We now compute the last term. Define $\mathcal{J}_\kappa(x) = 2\kappa \sqrt{\pi} \mathcal{G}_\kappa^2(x)$ so that $\mathcal{J}_\kappa, \kappa > 0$ is a new approximate identity. The last term is

$$\frac{1}{4} \kappa^{-1} \pi^{-1/2} \int \mathcal{J}_\kappa(x - y) e^{2(h(y) - h(x))} dy \left(\int \mathcal{G}_\kappa(x - y) e^{h(y) - h(x)} dy \right)^{-2}. \quad (22)$$

Because the $h(y) - h(x)$ are Brownian increments we can check

$$\frac{1}{2} Z_\kappa^{-2} \langle \mathcal{G}_{x, \kappa}^2, Z^2 \rangle - \frac{1}{4} \kappa^{-1} \pi^{-1/2} - c \xrightarrow{\text{weak}} 0 \quad (23)$$

where the precise value of the deterministic constant c is not relevant. It comes from essentially a quadratic variation computation. Note that one does not need the time average here, and we are only using that $h(x)$ is a two-sided Brownian motion for each $t \geq 0$. When we try to do the same thing for the quadratic term, $\frac{1}{2} (\partial_x h_\kappa)^2$ we

find that the leading order behavior comes again from a quadratic variation and matches precisely the leading order of (23). The next order term is a random field $\xi(t, x)$ which comes from the fluctuations of the quadratic variation. It contains all the interesting physics. But to understand it requires the time average as well. In other words, we have to understand the dynamics, which we do not. To summarize we have

$$\frac{1}{2}(\partial_x h_\kappa)^2 = \frac{1}{4}\kappa^{-1}\pi^{-1/2} + \xi + o(1). \quad (24)$$

If we could make sense of $:(\partial_x h_\kappa)^2: = \lim_{\kappa \rightarrow 0} \frac{1}{2}(\partial_x h_\kappa)^2 - \frac{1}{4}\kappa^{-1}\pi^{-1/2} = \xi$ then $h = -\log Z$ solves

$$\partial_t h = -\frac{1}{2}:(\partial_x h)^2: + \frac{1}{2}\partial_x^2 h + \dot{W}, \quad t > 0, \quad x \in \mathbb{R}. \quad (25)$$

The problem of proving well-posedness for (1) is now seen to be of a very different nature from a problem like well-posedness for the 3-d incompressible Navier-Stokes equations. We know that the solution is given by $h = -\log Z$; the problem is to find an appropriate formulation of (25) which fits that solution, and to prove the corresponding uniqueness.

References

- [1] L. Bertini and G. Giacomin, *Comm. Math. Phys.* **183**, 571 (1997).
- [2] G. Da Prato, *Kolmogorov equations for stochastic PDEs* Advanced Courses in Mathematics. CRM Barcelona, Advanced Courses in Mathematics. CRM Barcelona (Birkhäuser Verlag, Basel, 2004).
- [3] G. Da Prato, A. Debussche and L. Tubaro, *Electron. Comm. Probab.* **12**, 442 (2007).
- [4] H. Holden, B. Oksendal, J. Ubøe and T. Zhang, *Stochastic partial differential equations. A modeling, white noise functional approach* (Birkhauser Boston, Boston, 1996).
- [5] T. Chan, *Comm. Math. Phys.* **209**, 671 (2000).
- [6] D. Forster, D. R. Nelson and M. J. Stephen, *Phys. Rev. A (3)* **16**, 732 (1977).
- [7] K. Kardar, G. Parisi and Y. Zhang, *Phys. Rev. Lett.* **56**, 889 (1986).
- [8] J. Walsh, *An introduction to stochastic partial differential equations*, in *École d'été de probabilités de Saint-Flour, XIV-1984*, ed. P. L. Hennequin, Lecture Notes in Mathematics, Vol. 1180 (Springer-Verlag, 1986), pp. 265–439.
- [9] M. Balázs, J. Quastel and T. Seppäläinen, Scaling exponent of KPZ/Stochastic Burgers, in preparation.
- [10] M. Balázs and T. Seppäläinen, *ALEA Lat. Am. J. Probab. Math. Stat.* **6**, 1 (2009).

DERIVATION OF EFFECTIVE EVOLUTION EQUATIONS FROM MANY BODY QUANTUM DYNAMICS

BENJAMIN SCHLEIN

*DPMMS, University of Cambridge,
Cambridge, CB3 0WB, UK
E-mail: b.schlein@dpmms.cam.ac.uk*

We review some recent results concerning the derivation of effective evolution equations from first principle quantum dynamics. In particular, we discuss the derivation of the Hartree equation for mean field systems and the derivation of the Gross-Pitaevskii equation for the time evolution of Bose-Einstein condensates.

Keywords: Mean field systems, Hartree equation, Bose-Einstein condensates, Gross-Pitaevskii equation.

1. Introduction

We consider a quantum mechanical system of N spinless bosons in three dimensions. The system is described on the Hilbert space $\mathcal{H}_N = L_s^2(\mathbb{R}^{3N})$, consisting of all functions in $L^2(\mathbb{R}^{3N})$ which are symmetric with respect to arbitrary permutations of the N particles. The time evolution of a quantum mechanical system of N bosons is governed by the N -particle Schrödinger equation

$$i\partial_t\psi_{N,t} = H_N\psi_{N,t} \tag{1}$$

for the wave function $\psi_{N,t} \in \mathcal{H}_N$ of the system. Here H_N is a self-adjoint operator on $\mathcal{H}_N$, known as the Hamilton operator. We will consider Hamilton operators with two-body interactions described by a potential V , having the form

$$H_N = \sum_{j=1}^N -\Delta_{x_j} + \sum_{i<j}^N V(x_i - x_j).$$

Typically, the value of N in systems of physical interest is huge; it varies from $N \simeq 1000$ in very dilute experimental samples of Bose-Einstein condensates, up to $N \simeq 10^{23}$ in chemical experiments, or even to $N \simeq 10^{30}$ in astronomy. For such values of N , it is of course impossible to solve (1) explicitly; numerical methods are completely useless as well. Fortunately, also from the point of view of physics, it is not so important to know the precise solution to (1); it is much more important, for physicists performing experiments, to have information about the macroscopic properties of the system, which describe the typical behavior of the particles, and

result from averaging over the N particles. It turns out that the description of the macroscopic properties of the dynamics is particularly simple in so called mean field systems.

2. Mean Field Systems

A mean-field system is described by an N -body Hamilton operator of the form

$$H_N^{\text{mf}} = \sum_{j=1}^N -\Delta_j + \frac{1}{N} \sum_{i<j}^N V(x_i - x_j) \quad (2)$$

acting on $\mathcal{H}_N$. The mean-field character of the Hamiltonian is expressed by the factor $1/N$ in front of the interaction; this factor guarantees that the kinetic and potential energies are both typically of order N in the limit of large N .

We are going to study the solution $\psi_{N,t} = e^{-iH_N^{\text{mf}}t}\psi_N$ of the Schrödinger equation with factorized initial data $\psi_N = \varphi^{\otimes N}$, for some $\varphi \in L^2(\mathbb{R}^3)$.

In order to study the properties of $\psi_{N,t}$ in the limit $N \rightarrow \infty$, it is convenient to introduce marginal densities. For $k = 1, \dots, N$, the k -particle marginal density $\gamma_{N,t}^{(k)}$ associated with $\psi_{N,t}$ is defined by taking the partial trace of the density matrix $\gamma_{N,t} = |\psi_{N,t}\rangle\langle\psi_{N,t}|$ over the degrees of freedom of the last $N - k$ particles,

$$\gamma_{N,t}^{(k)} = \text{Tr}_{k+1,\dots,N} \gamma_{N,t} = \text{Tr}_{k+1,\dots,N} |\psi_{N,t}\rangle\langle\psi_{N,t}|.$$

In other words, $\gamma_{N,t}^{(k)}$ is defined as the non-negative trace class operator on $L^2(\mathbb{R}^{3k})$ with kernel given by

$$\gamma_{N,t}^{(k)}(\mathbf{x}_k; \mathbf{x}'_k) = \int d\mathbf{x}_{N-k} \bar{\psi}_{N,t}(\mathbf{x}_k, \mathbf{x}_{N-k}) \psi_{N,t}(\mathbf{x}'_k, \mathbf{x}_{N-k}). \quad (3)$$

Here we introduce the notation $\mathbf{x}_k = (x_1, \dots, x_N)$, $\mathbf{x}'_k = (x'_1, \dots, x'_N)$ and $\mathbf{x}_{N-k} = (x_{k+1}, \dots, x_N)$.

Due to the interaction, we cannot expect $\psi_{N,t}$ to be factorized. However, because of the mean-field character of the potential, it turns out that the k -particle marginal $\gamma_{N,t}^{(k)}$ associated with $\psi_{N,t}$ still factorizes in the limit $N \rightarrow \infty$, for arbitrary fixed $t \in \mathbb{R}$ and $k \in \mathbb{N}$. From the asymptotic factorization of the marginals, one can also obtain a self-consistent equation for the evolution of the condensate wave function (the one-particle orbital onto which the marginals factorize as $N \rightarrow \infty$), which turns out to be the nonlinear Hartree equation.

Theorem 2.1. *Under appropriate assumptions on the interaction potential V , let $\psi_N = \varphi^{\otimes N}$ for some $\varphi \in H^1(\mathbb{R}^3)$ with $\|\varphi\| = 1$, and $\psi_{N,t} = e^{-iH_N t}\psi_N$ (with H_N defined in (2)). Then, for every fixed $k \geq 1$ and $t \in \mathbb{R}$, we have*

$$\gamma_{N,t}^{(k)} \rightarrow |\varphi_t\rangle\langle\varphi_t|^{\otimes k} \quad \text{as } N \rightarrow \infty \quad (4)$$

in the trace-norm topology. Here φ_t is the solution to the nonlinear Hartree equation

$$i\partial_t \varphi_t = -\Delta \varphi_t + (V * |\varphi_t|^2) \varphi_t \quad (5)$$

with initial data $\varphi_{t=0} = \varphi$.

The first proof of this theorem was obtained by Spohn in [17] under the assumption of a bounded potential. The approach introduced by Spohn is based on the study of the time evolution of the reduced density matrices, which is governed by the BBGKY Hierarchy (for $k = 1, \dots, N$)

$$\begin{aligned} i\partial_t \gamma_{N,t}^{(k)} = & \sum_{j=1}^k \left[-\Delta_{x_j}, \gamma_{N,t}^{(k)} \right] + \frac{1}{N} \sum_{i < j}^k \left[V(x_i - x_j), \gamma_{N,t}^{(k)} \right] \\ & + \left(1 - \frac{k}{N} \right) \sum_{j=1}^k \text{Tr}_{k+1} \left[V(x_j - x_{k+1}), \gamma_{N,t}^{(k+1)} \right] \end{aligned} \quad (6)$$

where Tr_{k+1} denotes the partial trace over the variable x_{k+1} , and where we used the convention $\gamma_{N,t}^{(N+1)} \equiv 0$. If we fix $k \geq 1$ and we let $N \rightarrow \infty$, the BBGKY Hierarchy formally converges to the infinite hierarchy

$$i\partial_t \gamma_{\infty,t}^{(k)} = \sum_{j=1}^k \left[-\Delta_{x_j}, \gamma_{\infty,t}^{(k)} \right] + \sum_{j=1}^k \text{Tr}_{k+1} \left[V(x_j - x_{k+1}), \gamma_{\infty,t}^{(k+1)} \right]. \quad (7)$$

It is then worth noticing that this infinite hierarchy has a factorized solution. In fact, the family $\gamma_{\infty,t}^{(k)} = |\varphi_t\rangle\langle\varphi_t|^{\otimes k}$ solves (7) if and only if φ_t solves the nonlinear Hartree equation (5). This simple observation implies that in order to obtain a rigorous proof of Theorem 2.1, it is enough to complete the following three steps.

- Prove the *compactness* of the sequence of densities $\gamma_{N,t}^{(k)}$ with respect to an appropriate weak topology.
- Prove the *convergence* to the infinite hierarchy. In other words, show that every limit point of the sequence $\gamma_{N,t}^{(k)}$ solves (7).
- Prove the *uniqueness* of the solution of the infinite hierarchy.

These three steps immediately imply that $\gamma_{N,t}^{(k)} \rightarrow |\varphi_t\rangle\langle\varphi_t|^{\otimes k}$ as $N \rightarrow \infty$ (the convergence is first in a weak sense, but since the limit is a rank one projection, this automatically implies convergence in the trace norm topology).

Implementing this three step strategy becomes more difficult for potentials with singularities. In [8], Erdős and Yau extended Spohn's result to the case of a Coulomb potential $V(x) = \pm 1/|x|$. To deal with the singularity of V , they proved that arbitrary limit points $\gamma_{\infty,t}^{(k)}$ of the sequence of reduced densities $\gamma_{N,t}^{(k)}$ satisfy strong a-priori bounds of the form

$$\text{Tr} (1 - \Delta_{x_1}) \dots (1 - \Delta_{x_k}) \gamma_{\infty,t}^{(k)} \leq C^k \quad (8)$$

for some finite constant $C > 0$ and every $k \in \mathbb{N}, t \in \mathbb{R}$. Hence, it is enough to prove the uniqueness of the solution of (7) in the class of densities satisfying (8), where the Coulomb singularity can be controlled by the operator inequality $|x|^{-1} \leq C(1 - \Delta)$. A similar technique was used in [1], a joint work with A. Elgart, to handle bosons with a relativistic dispersion law interacting through a mean field Coulomb potential (the relativistic case is more complicated, because one can only prove weaker a-priori estimates).

A disadvantage of the strategy exposed above is the lack of effective bounds on the rate of convergence in (4). Following a different approach, proposed initially by Hepp in the slightly different context of the classical limit of quantum mechanics and later extended by Ginibre and Velo to a larger class of potentials (see [9, 11]), we recently managed in [16], a joint work with I. Rodnianski, to obtain effective bounds on the difference between the full Schrödinger evolution and the Hartree approximation. For potentials with at most a Coulomb singularity, we show that there exist constants $C, K > 0$ such that

$$\mathrm{Tr} \left| \gamma_{N,t}^{(1)} - |\varphi_t\rangle\langle\varphi_t| \right| \leq \frac{Ce^{Kt}}{\sqrt{N}} \quad (9)$$

for all $t \in \mathbb{R}$ (similar bounds holds for higher order marginals as well). Note that the bound on the r.h.s. of (9) is not expected to be optimal in its N dependence (in fact, for bounded potential, it was proven in [3], a joint work with L. Erdős, that the l.h.s. is at most of the order $1/N$ for every fixed t). The proof of (9) is based on a Fock space representation of the many boson system, and on the study of the time evolution of coherent states.

Recently, further progress has been achieved in the analysis of the dynamics of mean-field systems. In [12], Knowles and Pickl obtain effective estimates on the rate of convergence to the Hartree equation, for potentials with strong singularity (their proof is based on a method developed by Pickl in [15]). In [10], Grillakis, Machedon, and Margetis found second order correction to the mean-field evolution of coherent states.

3. Dynamics of Bose-Einstein condensates

Another class of systems for which an effective evolution equation can be derived consists of dilute Bose gases in the so called Gross-Pitaevskii scaling limit. The Hamiltonian of a trapped dilute Bose gas is given by

$$H_N^{\mathrm{trap}} = \sum_{j=1}^N (-\Delta_{x_j} + V_{\mathrm{ext}}(x_j)) + \sum_{i<j}^N V_N(x_i - x_j) \quad (10)$$

with $V_N(x) = N^2 V(Nx)$. Here V_{ext} is a confining potential, while $V \geq 0$ is a short range potential describing the interaction among the particles. We will denote by a_0 the scattering length of V . Recall that the scattering length is defined (for integrable potentials) as

$$8\pi a_0 = \int V(x)f(x)dx \quad (11)$$

where $f(x)$ is the solution of the zero-energy scattering equation

$$\left(-\Delta + \frac{1}{2}V \right) f = 0 \quad (12)$$

with the boundary condition $f(x) \rightarrow 1$, as $|x| \rightarrow \infty$. Note that, if a_0 is the scattering length of V , the scattering length of the rescaled potential V_N is exactly given by

$a = a_0/N$. This follows by simple scaling because, if f is the solution of (12), then $f_N(x) = f(Nx)$ solves the rescaled problem

$$\left(-\Delta + \frac{1}{2}V_N\right) f_N = 0. \quad (13)$$

In [14], Lieb, Seiringer, and Yngvason proved that, if E_N denotes the ground state energy of H_N^{trap} ,

$$\lim_{N \rightarrow \infty} \frac{E_N}{N} = \min_{\varphi \in L^2(\mathbb{R}^3): \|\varphi\|=1} \mathcal{E}_{\text{GP}}(\varphi)$$

where $\mathcal{E}_{\text{GP}}(\varphi)$ is the so called Gross-Pitaevskii energy functional given by

$$\mathcal{E}_{\text{GP}}(\varphi) = \int dx (|\nabla \varphi(x)|^2 + |V_{\text{ext}}(x)|^2 + 4\pi a_0 |\varphi(x)|^4). \quad (14)$$

In [13], Lieb and Seiringer also proved that the ground state ψ_N^{trap} of the Hamiltonian H_N^{trap} exhibits complete Bose-Einstein condensation. More precisely, they proved that, if $\gamma_N^{(1)}$ denotes the one-particle marginal associated with ψ_N^{trap} ,

$$\lim_{N \rightarrow \infty} \gamma_N^{(1)} = |\phi_{\text{GP}}\rangle\langle\phi_{\text{GP}}|$$

where ϕ_{GP} denotes the minimizer of (14) (the limit is in the trace norm topology). The interpretation of this result is straightforward; in the ground state of H_N^{trap} all particles, up to a fraction which vanishes in the limit $N \rightarrow \infty$, are in the same one-particle state with orbital ϕ_{GP} .

The question now is what happens to the condensate when the trapping potential is turned off. The next theorem, proven in a series of joint works with L. Erdős and H.-T. Yau (see [4-7]) shows that complete condensation is preserved by the Schrödinger dynamics and that the Gross-Pitaevskii theory correctly predicts the time evolution of the condensate wave function.

Theorem 3.1. *Suppose $V \geq 0$ is bounded and decays sufficiently fast and let*

$$H_N = \sum_{j=1}^N -\Delta_{x_j} + \sum_{i < j}^N V_N(x_i - x_j) \quad \text{with } V_N(x) = N^2 V(Nx). \quad (15)$$

Consider a family of initial data ψ_N with finite energy per particle

$$\langle \psi_N, H_N \psi_N \rangle \leq CN$$

and such that the one-particle marginal $\gamma_N^{(1)}$ associated with ψ_N satisfy

$$\gamma_N^{(1)} \rightarrow |\varphi\rangle\langle\varphi| \quad \text{as } N \rightarrow \infty$$

for some $\varphi \in H^1(\mathbb{R}^3)$ with $\|\varphi\| = 1$. Let $\psi_{N,t} = e^{-iH_N t} \psi_N$ and $\gamma_{N,t}^{(k)}$ be the k -particle marginal associated with $\psi_{N,t}$. Then, for every fixed $t \in \mathbb{R}$, and every $k \geq 1$, we have

$$\gamma_{N,t}^{(k)} \rightarrow |\varphi_t\rangle\langle\varphi_t|^{\otimes k} \quad \text{as } N \rightarrow \infty$$

in the trace norm topology. Here φ_t is the solution to the time dependent Gross-Pitaevskii equation

$$i\partial_t\varphi_t = -\Delta\varphi_t + 8\pi a_0|\varphi_t|^2\varphi_t \quad (16)$$

with initial data $\varphi_{t=0} = \varphi$.

Observe that the Hamiltonian (15) can be rewritten as

$$H_N = \sum_{j=1}^N -\Delta_{x_j} + \frac{1}{N} \sum_{i<j}^N v_N(x_i - x_j)$$

with $v_N(x) = N^3V(Nx)$. Formally, we have $v_N(x) \rightarrow b_0\delta(x)$ with $b_0 = \int V(x)dx$ as $N \rightarrow \infty$. It is therefore tempting to interpret (15) as a mean field Hamiltonian with a potential which depends on N and tends, as $N \rightarrow \infty$, to a delta-function. However, one should be very careful with this interpretation; physically, mean field systems are characterized by very frequent and very weak interactions. The Hamiltonian (15), on the other hand, describes a completely different regime, with very rare and very strong collisions (particles only interact when they are at distances of order N^{-1} , much smaller than the typical interparticle distance).

To understand the origin of the Gross-Pitaevskii equation (16) as the effective equation for the dynamics of the condensate wave function, consider the equation

$$i\partial_t\gamma_{N,t}^{(1)} = \left[-\Delta, \gamma_{N,t}^{(1)}\right] + (N-1)\text{Tr}_2 \left[V_N(x_1 - x_2), \gamma_{N,t}^{(2)}\right] \quad (17)$$

which describes the evolution of the one-particle marginal $\gamma_{N,t}^{(1)}$ associated with the solution $\psi_{N,t}$ of the N particle Schrödinger equation (this is the first equation of the BBGKY hierarchy for the family $\{\gamma_{N,t}^{(k)}\}_{k=1}^N$). If condensation is preserved by the time evolution, we should expect that, for large N , $\gamma_{N,t}^{(k)} \simeq |\varphi_t\rangle\langle\varphi_t|^{\otimes k}$ for both $k = 1, 2$. Inserting this Ansatz in (17), and using that $(N-1)V_N(x) \simeq N^3V(Nx) \rightarrow b_0\delta(x)$ as $N \rightarrow \infty$, we obtain the self-consistent equation

$$i\partial_t\varphi_t = -\Delta\varphi_t + b_0|\varphi_t|^2\varphi_t$$

for φ_t . This equation has the same form as (16), but a different constant in front of the nonlinearity. We get the wrong constant because, in this naive argument, we neglected the correlations characterizing the two-particle density $\gamma_{N,t}^{(2)}$. It turns out that the solution of the Schrödinger equation develops a short scale correlation structure which varies exactly on the same length scale of order N^{-1} characterizing the interaction potential. If we assume for a moment that the correlations can be described by the solution of the zero-energy scattering equation f_N (see 13), we may expect that for large but finite N the kernel of the one- and the two-particle marginals can be approximated by

$$\begin{aligned} \gamma_{N,t}^{(1)}(x_1; x'_1) &\simeq \varphi_t(x_1)\overline{\varphi}_t(x'_1) \\ \gamma_{N,t}^{(2)}(x_1, x_2; x'_1, x'_2) &\simeq f_N(x_1 - x_2)f_N(x'_1 - x'_2)\varphi_t(x_1)\varphi_t(x_2)\overline{\varphi}_t(x'_1)\overline{\varphi}_t(x'_2). \end{aligned} \quad (18)$$

Inserting this new, more precise, Ansatz in (17), we obtain another self-consistent equation for φ_t , which, by (11), is exactly the Gross-Pitaevskii equation (16).

From this heuristic discussion it is already clear that in order to obtain a rigorous proof of Theorem 3.1, one of the main steps consists in showing that the solution of the N -particle Schrödinger equation develops a short scale correlations structure, and that this structure can be described, in very good approximation, by the solution of the zero energy scattering equation f_N . In order to reach this goal, we make use of strong a-priori bounds of the form

$$\int d\mathbf{x} \left| \nabla_{x_i} \nabla_{x_j} \frac{\psi_{N,t}(\mathbf{x})}{f_N(x_i - x_j)} \right|^2 \leq C \quad (19)$$

valid for all $i \neq j$, uniformly in t and in N , and for every initial wave function ψ_N such that $\langle \psi_N, H_N^2 \psi_N \rangle \leq CN^2$ (the bounds (19) were derived in [5] under the assumption of sufficiently weak interaction potential; the case of strong potential was treated in [6], using different a-priori bounds). The estimate (19), which is proven using the conservation of the expectation of the Hamiltonian squared, identifies the short scale correlation structure of $\psi_{N,t}$. This can be used to prove the convergence of solutions of the BBGKY hierarchy towards an infinite hierarchy of equation similar to (7), but with V replaced by $8\pi a_0$.

To complete the proof of Theorem 3.1 (following the strategy outlined in Section 2), one still needs to prove the uniqueness of the solution of the infinite hierarchy; technically, this is actually the most difficult part of the proof. On the one hand, we have to prove that the limit point of the densities satisfy a-priori estimates of the form (8); the problem here is much more involved compared with the case of a mean-field Coulomb interaction because of the presence of the singular correlation structure. On the other hand, when we prove the uniqueness of the solution of the infinite hierarchy in the class of densities satisfying the a-priori estimates, we have to face the problem that, in three dimensions, the delta-interaction is not bounded by the Laplacian (in contrast to $|x|^{-1} \leq C(1 - \Delta)$, the inequality $\delta(x) \leq C(1 - \Delta)$ is not true in three dimensions). Details of the proof of the a-priori bounds can be found in [5, Section 5]. The proof of the uniqueness (in the class of densities satisfying the a-priori bounds) can be found in [4].

4. Dynamical Formation of Correlations

In the last section, we stressed the fact that the emergence of the scattering length in the Gross-Pitaevskii equation (16) is a consequence of the short scale correlation structure characterizing the solution $\psi_{N,t}$ of the N -particle Schrödinger equation. If one assumes that $\langle \psi_N, H_N^2 \psi_N \rangle \leq CN^2$, the presence of the correlation structure in the evolved wave function $\psi_{N,t}$ follows from a-priori bounds of the form (19). It turns out, however, that Theorem 3.1 can also be applied to completely factorized initial data of the form $\psi_N = \varphi^{\otimes N}$. Therefore, also for initial data with absolutely no correlations among the particles, the evolution of the condensate wave function is described by the Gross-Pitaevskii equation (16) with coupling constant proportional

to a_0 . This observation suggests that the time evolution $\psi_{N,t}$ of the completely factorized data ψ_N develops the short scale correlation structure within very small time intervals, whose length vanishes in the limit $N \rightarrow \infty$. Note that the a-priori bound (19) cannot be used to prove that $\psi_{N,t}$ contains the correct short scale structure because $\langle \varphi^{\otimes N}, H_N^2 \varphi^{\otimes N} \rangle \simeq N^3 \gg N^2$. In order to prove Theorem 3.1 for factorized initial data ψ_N we need therefore an approximation argument to replace ψ_N by the N -particle wave function $\tilde{\psi}_{N,\varepsilon} \simeq \chi(H_N \leq N\varepsilon^{-1})\psi_N$ having the correct correlation structure. We perform then the whole analysis on the evolution of $\tilde{\psi}_{N,\varepsilon}$ and only at the end, after taking $N \rightarrow \infty$, we let $\varepsilon \rightarrow 0$.

In particular, our analysis does not give any proof of the fact that the evolution of a factorized initial data develops a short scale correlation structure within short times. In this section, which is based on [2], a joint work with L. Erdős and A. Michelangeli, we provide some direct evidence supporting this claim.

It turns out that the dynamical formation of correlation can be understood as a two-body phenomenon. Consider the two-particle Hamiltonian

$$\mathfrak{h}_N^{(1,2)} = -\Delta_1 - \Delta_2 + N^2 V(N(x_1 - x_2))$$

and the factorized two-particle wave function $\psi(x_1, x_2) = \varphi(x_1)\varphi(x_2)$ for a $\varphi \in L^2(\mathbb{R}^3)$ smooth and decaying fast enough. In order to monitor the formation of correlations in a window of length $N^{-1} \leq \ell \ll 1$, we introduce the quantity

$$\mathcal{F}_N(t) = \int dx_1 dx_2 \theta(|x_1 - x_2| \leq \ell) \left| \nabla_{x_1 - x_2} \frac{e^{-it\mathfrak{h}_N^{(1,2)}} \psi(x_1, x_2)}{f_N(x_1 - x_2)} \right|^2.$$

At time $t = 0$, ψ is factorized and it is simple to show that $N\mathcal{F}_N(0) \simeq 1$. The following proposition gives an upper bound on $\mathcal{F}_N(t)$ for $t > 0$.

Proposition 4.1. *Suppose that $0 \leq t \leq N^{-1}$ and $N^{-1} \leq \ell \ll 1$. Then*

$$N\mathcal{F}_N(t) \leq \frac{(\log N^2 t)^6}{N^2 t} (N\ell)^3.$$

In particular, Proposition 4.1 implies that, if $\ell \simeq N^{-1}$, $\mathcal{F}_N(t) \ll \mathcal{F}_N(0)$ for all $N^{-2} \ll t \leq N^{-1}$; this is clear evidence for the formation of the correlation structure on length scale of order $\ell \simeq N^{-1}$ within times of order N^{-2} (note that, for technical reasons, the result is restricted to times $t \leq N^{-1}$; hence, strictly speaking, Proposition 4.1 only shows that correlations form during the time interval $N^{-2} \ll t \leq N^{-1}$).

To prove Proposition 4.1 it is useful to switch to a center of mass coordinate $\eta = (x_1 + x_2)/2$ and to a relative coordinate $x = x_2 - x_1$. In these coordinates, the Hamiltonian takes the form

$$\mathfrak{h}_N^{(1,2)} = -\frac{\Delta_\eta}{2} - 2\Delta_x + V_N(x) =: -\frac{\Delta_\eta}{2} + \mathfrak{h}_N$$

with $\mathfrak{h}_N = -2\Delta_x + V_N(x)$. If we moreover rescale the relative coordinate letting $X = Nx$, we find

$$\mathfrak{h}_N = N^2(-2\Delta_X + V(X)) =: N^2\mathfrak{h}.$$

Hence, setting $T = N^2t$, we find

$$N\mathcal{F}_N(t) = \int d\eta dX \theta(|X| \leq N\ell) \left| \nabla_X \frac{(e^{-iT\mathfrak{h}}\psi_N)(\eta, X)}{f(X)} \right|^2 \quad (20)$$

where $\psi_N(\eta, X) = \varphi(\eta + X/N)\varphi(\eta - X/N)$. In (20), the N -dependence is hidden in the initial data ψ_N which has the property of being essentially constant in X , as long as $|X| \ll N$. The formation of correlations for $T \gg 1$ is then a statement about the local relaxation of an initial data which is essentially independent of X , under the dynamics generated by $\mathfrak{h}$, to the function $f(X)$. Recall here that f is such that $\mathfrak{h}f = 0$; one can prove that, for large $|X|$, it has the form $f(X) \simeq 1 - a_0/|X|$. Defining $\omega(X) = 1 - f(X)$, we have, very formally,

$$e^{-iT\mathfrak{h}}1 = e^{-iT\mathfrak{h}}(1 - \omega) + e^{-iT\mathfrak{h}}\omega = f + e^{-iT\mathfrak{h}}\omega.$$

The second term disperses away for large T and therefore, if we only look inside a window of order one ($\ell \simeq N^{-1}$), $e^{-iT\mathfrak{h}}1 \simeq f$. This explains why the quantity under investigation becomes small for $T \gg 1$ (which is equivalent to $t \gg N^{-2}$). The main obstacle to make this formal argument rigorous is the lack of decay of ω . Since $\omega \simeq |X|^{-1}$ at infinity, the standard dispersive estimates for the evolution of ω cannot be applied. Instead, we show new dispersive estimates of the form

$$\|e^{-i\Delta t}\varphi\|_\infty \leq Ct^{-\frac{3}{2s}} \left(\|\varphi\|_s + \|\nabla\varphi\|_{\frac{3s}{s+3}} + \|\nabla^2\varphi\|_{\frac{3s}{2s+3}} \right) \quad (21)$$

for all $s \geq 3/2$ which require less decay of the initial data; the price to pay is that one needs some regularity (similar bounds can be obtained for the L^q -norms of $e^{-i\Delta t}\varphi$ for sufficiently large q). Choosing $s \geq 3$, the bound (21) can be applied to the function ω . Using Yajima's bounds on the wave operator (see [18]), estimates of the form (21) can then be proven to hold also for the dynamics generated by $\mathfrak{h}$.

Proposition 4.1 proves that the formation of correlation is a two-particle phenomenon. In order to understand the formation of correlation in the many particle setting, we would need to control the effect of the three-body interactions on the correlation structure. This is a very difficult task and, in general, not much can be rigorously established about this problem. Since, however, the many body system is very dilute, three body interactions are very rare. Therefore we can find an interval of time during which two body collisions are already effective (and lead therefore to the formation of the correlation structure) while three body collisions are still negligible.

Theorem 4.1. *Let*

$$\mathcal{G}_N(t) = \int d\mathbf{x} \theta(|x_1 - x_2| \leq \ell) \left| \frac{\psi_{N,t}(\mathbf{x})}{f_N(x_1 - x_2)} - \psi_N(\mathbf{x}) \right|^2$$

for $\psi_N = \varphi^{\otimes N}$ with $\varphi \in L^2(\mathbb{R}^3)$ smooth and decaying fast enough, and $\psi_{N,t} = e^{-iH_N t}\psi_N$, with H_N defined in (15). Then, for all $0 \leq t \leq N^{-1}$,

$$\mathcal{G}_N(t) \leq C\mathcal{G}_N(0) \left(\frac{(\log N)^{\frac{4}{5}} (N^2t)^2}{N^{\frac{1}{5}} N\ell} + \frac{(N\ell)^4}{N^2t} (\log N^2t)^6 \right).$$

In particular, the theorem implies that, when considering windows of size $\ell \simeq N^{-1}$,

$$\mathcal{G}_N(t) \ll \mathcal{G}_N(0) \quad \text{if} \quad 1 \ll N^2 t \ll N^{\frac{1}{10}}.$$

This inequality indicates that, under the many body dynamics, a completely factorized initial data develops the short scale correlation structure within times of order N^{-2} and keeps it at least up to times of order $N^{-1.9}$ (of course, we expect the correlation structure to remain intact up to times of order one, but proving this claim would require a better understanding of the effect of the many-body interactions).

References

- [1] Elgart, A.; Schlein, B.: Mean field dynamics of boson stars. *Commun. Pure Appl. Math.* **60** (2007), no. 4, 500–545.
- [2] Erdős, L.; Michelangeli, A.; Schlein, B.: Dynamical formation of correlations in a Bose-Einstein condensate. *Comm. Math. Phys.* **289** (2009), no. 3, 1171–1210.
- [3] Erdős, L.; Schlein, B.: Quantum dynamics with mean field interactions: a new approach. *J. Stat. Phys.* **134** (2009), no. 5, 859–870.
- [4] Erdős, L.; Schlein, B.; Yau, H.-T.: Derivation of the cubic nonlinear Schrödinger equation from quantum dynamics of many-body systems. *Invent. Math.* **167** (2007), 515–614.
- [5] Erdős, L.; Schlein, B.; Yau, H.-T.: Derivation of the Gross-Pitaevskii equation for the dynamics of Bose-Einstein condensate. Preprint arXiv:math-ph/0606017. To appear in *Ann. Math.*
- [6] Erdős, L.; Schlein, B.; Yau, H.-T.: Rigorous derivation of the Gross-Pitaevskii equation. *Phys. Rev. Lett.* **98** (2007), no. 4, 040404.
- [7] Erdős, L.; Schlein, B.; Yau, H.-T.: Rigorous derivation of the Gross-Pitaevskii equation with a large interaction potential. Preprint arXiv:0802.3877. To appear in *J. Amer. Math. Soc.*
- [8] Erdős, L.; Yau, H.-T.: Derivation of the nonlinear Schrödinger equation from a many body Coulomb system. *Adv. Theor. Math. Phys.* **5** (2001), no. 6, 1169–1205.
- [9] Ginibre, J.; Velo, G.: The classical field limit of scattering theory for nonrelativistic many-boson systems. I-II. *Comm. Math. Phys.* **66** (1979), no. 1, 37–76, and **68** (1979), no. 1, 45–68.
- [10] Grillakis, M.; Machedon, M.; Margetis, D.: Second-order corrections to mean field evolution for weakly interacting bosons. I. Preprint arXiv:0904.0158.
- [11] Hepp, K.: The classical limit for quantum mechanical correlation functions. *Comm. Math. Phys.* **35** (1974), 265–277.
- [12] Knowles, A.; Pickl, P.: Mean-field dynamics: singular potentials and rate of convergence. Preprint arXiv:0907.4313.
- [13] Lieb, E.H.; Seiringer, R.: Proof of Bose-Einstein condensation for dilute trapped gases. *Phys. Rev. Lett.* **88** (2002), 170409–1–4.
- [14] Lieb, E.H.; Seiringer, R.; Yngvason, J.: Bosons in a trap: a rigorous derivation of the Gross-Pitaevskii energy functional. *Phys. Rev. A* **61** (2000), 043602.
- [15] Pickl, P.: A simple derivation of mean field limits for quantum systems. Preprint arXiv:0907.4464.
- [16] Rodnianski, I.; Schlein, B.: Quantum fluctuations and rate of convergence towards mean field dynamics. *Comm. Math. Phys.* **291** (2009), no. 1, 31–61.
- [17] Spohn, H.: Kinetic equations from Hamiltonian dynamics. *Rev. Mod. Phys.* **52** (1980), no. 3, 569–615.

- [18] Yajima, K.: The $W^{k,p}$ -continuity of wave operators for Schrödinger operators. *J. Math. Soc. Japan* **47** (1995), no. 3, 551–581.

OTHER TALKS AND POSTERS

INVITED TALKS

PERSPECTIVES IN MATHEMATICAL PHYSICS

WOJCIECH DE ROECK, JÜRGE FRÖHLICH*, ALESSANDRO PIZZO, ISRAEL M. SIGAL

**ETH Zürich*

E-mail: juerg@phys.ethz.ch

I will describe some important problems in theoretical physics and illustrate in which way mathematical physics has contributed to progress in solving them. Among such problems are general aspects of quantum theory and quantum field theory, the fundamental laws of thermodynamics, the quantum theory of open systems and transport theory, etc. Special emphasis will be placed on problems related to quantum Brownian motion and quantum friction.

ASEP: INTEGRABLE STRUCTURES & LIMIT THEOREMS

CRAIG A. TRACY, HAROLD WIDOM

University of California Davis

E-mail: tracy@math.ucdavis.edu

We explain how ideas from Bethe Ansatz can be used to find exact formulas for the distribution of the current in the asymmetric simple exclusion process (ASEP). Asymptotic analysis of the resulting formulas then establishes KPZ universality for the current fluctuations. Extensions to ASEP with second class particles will also be discussed.

CONTRIBUTED TALKS

ON THE APPROACH TO EQUILIBRIUM AND MIXING PROPERTIES IN FERMI-PASTA-ULAM MODELS

ANDREA CARATI, LUIGI GALGANI, ANTONIO GIORGILLI, SIMONE PALEARI*

**Università di Milano*

E-mail: simone.paleari@unimi.it

The well-known numerical experiment of Fermi-Pasta-Ulam (FPU) was conceived in order to study the approach to equilibrium. Starting from a zero-measure class of initial data they showed a lack of equipartition of energy among linear normal modes, raising the issue of relaxation times.

At variance with the original case of FPU, we investigate the approach to equilibrium considering sets of initial data of full measure, sampled out from the microcanonical ensemble by standard Monte Carlo techniques. We give numerical evidences that the system does not have the mixing property on finite time at low energy. Mixing is tested by looking at the decay of the autocorrelations of the mode energies.

We will discuss the results obtained in [CGGP], which show a partial decay of the autocorrelation to a nontrivial profile as a function of the mode number; such profile scales with the specific energy of the system and seems to survive in the thermodynamic limit. We will also present some recent progress on the role of the special form of nonlinearity (α -model vs β -model) in these phenomena.

[CGGP] A. Carati, L. Galgani, A. Giorgilli, S. Paleari: Fermi-Pasta-Ulam phenomenon for generic initial data; Phys. Rev. E 76, 022104 (2007)

NONEQUILIBRIUM STATISTICAL MECHANICS OF (AN)HARMONIC CHAINS WITH SELF-CONSISTENT RESERVOIRS

EMMANUEL ARAÚJO PEREIRA, RICARDO FALCAO, HUMBERTO C.F. LEMOS

Universidade Federal de Minas Gerais

E-mail: emmanuel@fisica.ufmg.br

We derive an integral representation (à la Feynman-Kac) for the correlation functions of harmonic and anharmonic chains of oscillators with self-consistent reservoirs [1,2,3]. The integral formalism is inspired on a previous work for the simpler case of a nonconservative system coupled to thermal baths at the same temperature [4], and it is suitable for the study of the correlations time decay, as well as the correlations behavior in the steady state. For the harmonic case, we give a new proof for the finitude of thermal conductivity in the steady state [1,3], and for inhomogeneous systems, with graded masses or interparticle interactions, we show the absence of thermal rectification [5]. For the anharmonic models, we study (performing a non rigorous analysis) a system with soft (bounded) anharmonicity considering different constant couplings between each site and its reservoir: we make small the coupling constant for the inner sites in order to infer the behavior of the system as it is connected to two thermal baths at the boundaries only [2].

[1] Emmanuel Pereira and Ricardo Falcao, Phys. Rev. E 70, 046105 (2004)

[2] Emmanuel Pereira and Ricardo Falcao, Phys. Rev. Lett. 96, 100601 (2006)

[3] R. Falcao, A. Francisco Neto, and E. Pereira, Theor. Math. Phys. 156, 1081 (2008)

[4] P.A.F. da Veiga, M. O'Carroll, E. Pereira, and R. Schor, Commun. Math. Phys. 220, 377 (2001)

[5] Emmanuel Pereira and Humberto C. F. Lemos, Phys. Rev. E 78, 031108 (2008)

POSTERS

TATIANA V. DUDNIKOVA (Elektrostal Polytechnical Institute, Russia): **LATTICE DYNAMICS IN THE HALF-SPACE: ENERGY TRANSPORT EQUATION**

CHRISTIANE MARIA LOSERT VALIENTE KROON (Department of Physics, UCL), with IAN J. FORD: **STOCHASTIC BEHAVIOUR IN POPULATION DYNAMICS**

ALBERTO MAIOCHHI (Università degli Studi di Milano), with ANDREA CARATI: **RELAXATION TIMES FOR HAMILTONIAN ISOLATED SYSTEMS**

YOUNHYUN YEO (Konkuk University, Korea): **DENSITY NONLINEARITIES AND FIELD THEORIES IN A TOY MODEL FOR FLUCTUATING NONLINEAR HYDRODYNAMICS OF SUPERCOOLED LIQUIDS**

PDE's

(fluid dynamics, wave equation and NLSE,
Boltzmann equation, general relativity,
and elasticity)

Session organizers:

Sergiu Klainerman
Grigory Seregin

A NEW PHYSICAL-SPACE APPROACH TO DECAY FOR THE WAVE EQUATION WITH APPLICATIONS TO BLACK HOLE SPACETIMES

MIHALIS DAFERMOS

*Department of Pure Mathematics and Mathematical Statistics,
University of Cambridge, Cambridge, CB3 0WB, United Kingdom
E-mail: M.Dafermos@dpmms.cam.ac.uk*

IGOR RODNIANSKI

*Department of Mathematics, Princeton University
Fine Hall, Washington Road, Princeton, New Jersey 08544, United States*

We present a new general method for proving global decay of energy through a suitable spacetime foliation, as well as pointwise decay, starting from an integrated local energy decay estimate. The method is quite robust, requiring only physical space techniques, and circumvents use of multipliers or commutators with weights growing in t . In particular, the method applies to a wide class of perturbations of Minkowski space as well as to Schwarzschild and Kerr black hole exteriors.

Keywords: The wave equation; Black holes; General relativity

1. Introduction

The wave equation

$$\square_g \phi = 0 \tag{1}$$

on general Lorentzian background metrics g appears ubiquitously in mathematical physics, often in the context of the linearisation of a field theory governed by a system of non-linear hyperbolic p.d.e.'s. The most classical example occurs perhaps in the context of the Euler equations of fluid mechanics, in which case g represents the so-called acoustical metric. Another important source for (1) arises from problems in general relativity, where g represents the metric of spacetime, and (1) can be viewed as a poor-man's linearisation of the Einstein equations. For such metrics g , the global causal structure is often much richer than that of Minkowski space. In this context, perhaps the most interesting examples are so-called “black hole” metrics. Briefly, these are metrics which—like Minkowski space—possess a natural asymptotic structure representing future null infinity $\mathcal{I}^+$, but where the past of $\mathcal{I}^+$ has a non-trivial complement in the spacetime.

A fundamental problem for (1) on suitable metrics g is to obtain “decay estimates”, that is to say, estimates giving upper bounds for the rates of decay for ϕ toward the future, in terms of quantities computable initially on a Cauchy surface. This classical problem can be studied as an end in itself, but it is often an essential step in proving stability results for non-linear equations which give rise to (1) as described above. From the latter point of view, there is an important trade off between how much decay one proves and how robust is the argument to prove it. The history of non-linear wave equations in the past 30 years, culminating in part in the monumental stability proofs (see for instance [4]) based on robust estimates for equations of the form (1), are a testament to how important it is to get this balance right.

The now standard approach to obtaining robust decay estimates proceeds by using weighted multiplier estimates (going back to Morawetz [14]) and weighted commutators (going back to Klainerman [10]), both originating in one form or another from the symmetries of Minkowski space. In Minkowski space, full use of this method (applying both weighted multipliers and weighted commutators) obtains (a) weighted energy decay, which can be expressed for instance as

$$\begin{aligned} \int_{t=\tau} (\phi^2 + (t+r+1)^2((\partial_t + \partial_r)\phi)^2 + (|t-r|+1)^2((\partial_t - \partial_r)\phi)^2 + |\nabla\phi|^2) \\ \leq C \int_{t=0} (\phi^2 + r^2((\partial_t\phi)^2 + (\partial_r\phi)^2 + |\nabla\phi|^2)), \end{aligned} \quad (2)$$

where ∇ denotes angular derivatives, (b) pointwise decay for ϕ of the form

$$|\phi| \leq C\sqrt{E}(|t-r|+1)^{-1/2}(|t+r|+1)^{-1} \quad (3)$$

for $t \geq 0$, where E represents a higher order weighted energy at $t = 0$, and (c) pointwise estimates for derivatives of ϕ , with additional decay, for instance:

$$|\partial_t\phi| \leq C\sqrt{E}(|t-r|+1)^{-3/2}(|t+r|+1)^{-1}. \quad (4)$$

Various applications to non-linear problems and other geometric settings have required variants of the vector field method where the use of certain vector field multipliers and/or commutators is avoided and less information is obtained concerning (a), (b), or (c) above. Klainerman and Sideris [11] obtained decay results by commuting only with $t\partial_t + r\partial_r$ and angular momentum operators, and this, used together with integrated local decay estimates in the style of Morawetz [15], has proven useful in a number of problems concerning obstacles, multiple speeds, etc. [1, 9, 13]. A common feature of all these approaches, however, is the necessity of always applying at least some multipliers or commutators with weights growing in t .

In the present paper, we give a related but different approach to the problem of proving suitable decay estimates for solutions to (1) which, as we shall see, has in principle a wider range of validity and may be even more useful for non-linear problems. In particular, the method here avoids both multipliers and commutators with weights in t . Our approach does not start from scratch, but begins with certain

“more basic” estimates as ingredients, estimates which have already been proven (either classically or much more recently) in many cases of interest, by various methods. The method by which these ingredient estimates are obtained is in fact of no importance for the considerations here. As we shall see, our approach applies to the study of (1) on Minkowski space and sufficiently small perturbations thereof, but also to Schwarzschild and Kerr black hole spacetimes, and immediately yields essentially the decay consistent with non-linear stability. The common feature which these regimes share are

- (i) An integrated local energy decay estimate
- (ii) Good asymptotics towards null infinity allowing for a hierarchy of r -weighted energy identities
- (iii) Uniform boundedness of energy

Estimates of the form (i), already referred to above, were first proven by Morawetz [15] for perturbations of Minkowski space, while for Schwarzschild, estimates of the form (i) are due to several authors (see Section 4.4.1 of [8]). For slowly rotating Kerr spacetimes $(\mathcal{M}, g_{M,a})$ with parameters $|a| \ll M$, such estimates have been shown by [8, 16] (see also [2]), whereas, for the general sub-extremal Kerr case $|a| < M$, this will appear in a forthcoming paper. In the black hole case, the “integrated decay” estimate (i), to be useful here, must be understood in the sense of an estimate which does not degenerate at the horizon $\mathcal{H}^+$, an estimate which is indeed possible in view of a multiplier construction intimately connected to the celebrated red-shift effect (see [5, 8]). In the Schwarzschild and Kerr case, the integrated decay estimate necessarily degenerates, however, near the analogue of the photon sphere, and this loss is related to the trapping phenomenon; a non-degenerate estimate can nevertheless be obtained after losing differentiability. Similar phenomena occur in the simpler setting of obstacle problems in Minkowski space, to which this method also applies.

We note that the hierarchy of r -weighted energy identities (ii) applied here corresponds to the use of a multiplier of the form $r^p(\partial_t + \partial_r)$ for $0 \leq p \leq 2$.

In principle, our approach can retrieve the “full decay statements” of the traditional vector field method as would follow from the combined use of [14] and [10], except that the energy decay statements will in general lose derivatives if the integrated local energy decay statement does so. In this short note, we shall here only give the first step of the approach. This begins by obtaining τ^{-2} decay of the energy flux through a suitable foliation Σ_τ (whose leaves approach null infinity), where τ is the natural parameter of the foliation. This is precisely the analogue of (2). (See the discussion of Section 7.) This then leads immediately to pointwise results (in Section 5) which are the analogue of

$$|r^{1/2}\phi| \leq Ct^{-1}, \quad |r\phi| \leq Ct^{-1/2}.$$

Again, no use is made of commutation with vector fields with weights in t . We outline in Section 8 how to extend the method to achieve the analogue of (3).

We consider the case of Minkowski space first in Section 3, followed by Schwarzschild in Section 4, to illustrate the structure of the argument in each of these settings. As shall be apparent, the nature of the argument is such that elements (i)–(iii), suitably interpreted, are sufficient and a very general theorem can in fact be formulated. This general theorem includes in particular the Kerr case. See Section 6. We leave the precise formulation of the most general assumptions for a follow-up paper.

2. Notation

Let $(\mathcal{M}, g)$ be a Lorentzian manifold and ϕ a solution to (1) on $\mathcal{M}$. We recall the energy-momentum tensor $\mathbb{T}_{\mu\nu}[\phi] = \phi_{,\mu}\phi_{,\nu} - \frac{1}{2}g_{\mu\nu}\phi^{,\alpha}\phi_{,\alpha}$. Given a vector field X , we define the currents $J_\mu^X[\phi] = \mathbb{T}_{\mu\nu}[\phi]X^\nu$, $K^X[\phi] = \mathbb{T}^{\mu\nu}[\phi]\pi_{\mu\nu}^X$, where $\pi_{\mu\nu}^X = \frac{1}{2}\mathcal{L}_X g_{\mu\nu}$. Recall that $\nabla^\mu J_\mu^X[\phi] = K^X[\phi]$. If X is Killing, then $K^X[\phi] = 0$, in which case $J_\mu^X[\phi]$ is a conserved current.

Integrands without an explicit measure of integration are to be understood with respect to the induced volume form. For the case of null hypersurfaces Σ , integrands of the form $\int_\Sigma J_\mu^X n_\Sigma^\mu$ are to be understood with respect to a choice of volume form on Σ and corresponding normalisation for n^μ such that the integral represents the correct term in the energy identity.

3. The Minkowski case

Let us start by considering the wave equation (1) in Minkowski space. For convenience, we may assume for this discussion that ϕ is smooth and compactly supported on a Cauchy hypersurface. This will ensure that all integrals are a priori defined. By a density argument, our results will only require the finiteness of the quantity on the right hand side of the final estimate (7) and the relation $\phi \rightarrow 0$ at spacelike infinity.

Let (t, r, θ, φ) denote a choice of standard spherical polar coordinates on Minkowski space. We shall let ∇ denote the induced covariant derivative on the spheres of constant r , and Δ the induced Laplacian. We will also define the null coordinates $u = t - r$, $v = t + r$. Finally, let us introduce the notation T to denote the coordinate vector field ∂_t with respect to (t, r) coordinates.

For our argument it will be useful to consider the following foliation: Fix $R > 0$ and consider the hypersurfaces Σ_τ defined as the union of the spacelike $\{t = \tau\} \cap \{r \leq R\}$ and the null $\{u = (\tau - R)\} \cap \{v \geq (\tau + R)\}$.

The following estimate can be shown by the original method of Morawetz [15]:

$$\int_\tau^\infty \int_{r \leq R} (J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha + \phi^2) \leq C_R \int_{\Sigma_\tau} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha. \quad (\text{ILED-Mink})$$

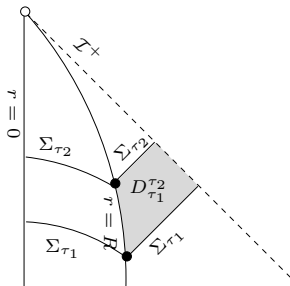
The initials of the label stand for “integrated local energy decay”. Since $n^\alpha = T^\alpha$ in $r \leq R$, we could replace the first term on the right hand side more explicitly with the quantity $(\partial_t \phi)^2 + (\partial_r \phi)^2 + |\nabla \phi|^2$, but we prefer this more geometric formulation

to compare with other situations. Our goal is to use (ILED-Mink) to derive decay for the energy flux through Σ_τ , and finally, pointwise decay estimates.

We rewrite the wave equation in the form

$$-\partial_u \partial_v \psi + \Delta \psi = 0, \quad \psi := r\phi,$$

multiply the equation by $r^p \partial_v \psi$ and integrate by parts in the region $D_{\tau_1}^{\tau_2}$ bounded by the two null hypersurfaces $u_1 = (\tau_1 - R)$, $u_2 = (\tau_2 - R)$ and the timelike hypersurface $r = R$:



We obtain (suppressing an additional spherical integration $\int_{\theta, \varphi} \sin \theta d\theta d\varphi$ from the notation)

$$\begin{aligned} & \int_{\Sigma_{\tau_2} \cap \{r \geq R\}} r^p (\partial_v \psi)^2 dv + \int_{D_{\tau_1}^{\tau_2}} r^{p-1} (p(\partial_v \psi)^2 + (2-p)|\nabla \psi|^2) du dv + \int_{\mathcal{I}_{\tau_1-R}^{\tau_2-R}} r^p |\nabla \psi|^2 du \\ = & \int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^p (\partial_v \psi)^2 dv + \int_{\tau_1}^{\tau_2} r^p (|\nabla \psi|^2 - (\partial_v \psi)^2) d\tau|_{r=R}. \end{aligned} \quad (\text{p-WE-Mink})$$

The initials “WE” denote “weighted energy”. Remember that the r^2 factor of the volume form is included in the quadratic terms in ψ as $\psi = r\phi$. Thus, the r^p and r^{p-1} factors should be viewed as weights. We see that the left hand side of the above identity is positive definite for $p \leq 2$.

Let us first apply (p-WE-Mink) with $p = 2$. Then, observing that the last term on the right hand side of (p-WE-Mink) can be controlled (after a bit of averaging in R) by the left hand side of (ILED-Mink), we obtain

$$\int_{\Sigma_{\tau_2} \cap \{r \geq R\}} r^2 (\partial_v \psi)^2 dv + \int_{D_{\tau_1}^{\tau_2}} r (\partial_v \psi)^2 du dv \leq \int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2 (\partial_v \psi)^2 dv + C \int_{\Sigma_{\tau_1}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha.$$

This implies that we can find a dyadic sequence of $\tau_n \rightarrow \infty$ with the property that

$$\int_{\Sigma_{\tau_n} \cap \{r \geq R\}} r (\partial_v \psi)^2 dv \leq C \tau_n^{-1} \left[\int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2 (\partial_v \psi)^2 dv + \int_{\Sigma_{\tau_1}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \right].$$

We now apply (p-WE-Mink) with $p = 1$ to the region $D_{\tau_{n-1}}^{\tau_n}$ to obtain

$$\begin{aligned}
& \int_{\Sigma_{\tau_n} \cap \{r \geq R\}} r(\partial_v \psi)^2 dv + \int_{D_{\tau_{n-1}}^{\tau_n}} ((\partial_v \psi)^2 + |\nabla \psi|^2) du dv \\
& \leq \int_{\Sigma_{\tau_{n-1}} \cap \{r \geq R\}} r(\partial_v \psi)^2 dv + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \\
& \leq C \tau_n^{-1} \left[\int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2(\partial_v \psi)^2 dv + \int_{\Sigma_{\tau_1}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \right] + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha.
\end{aligned}$$

Adding a multiple of the estimate (ILED-Mink), we obtain

$$\begin{aligned}
& \int_{\tau_{n-1}}^{\tau_n} \int_{r \leq R} (J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha + \phi^2) + \int_{\Sigma_{\tau_n} \cap \{r \geq R\}} r(\partial_v \psi)^2 dv + \int_{D_{\tau_{n-1}}^{\tau_n}} ((\partial_v \psi)^2 + |\nabla \psi|^2) du dv \\
& \leq C \tau_n^{-1} \left[\int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2(\partial_v \psi)^2 dv + \int_{\Sigma_{\tau_1}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \right] + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha. \quad (5)
\end{aligned}$$

We now observe that

$$\int_{v_n}^\infty (\partial_v(r\phi))^2 dv = \int_{v_n}^\infty [r^2(\partial_v \phi)^2 + 2r\partial_v \phi \phi + \phi^2] dv = \int_{v_n}^\infty r^2(\partial_v \phi)^2 dv - r\phi^2|_{v=v_n}.$$

Substituting this into (5) we obtain

$$\begin{aligned}
& \int_{\tau_{n-1}}^{\tau_n} \int_{\Sigma_\tau} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \leq C \tau_n^{-1} \left[\int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2(\partial_v \psi)^2 dv + \int_{\Sigma_{\tau_1}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \right] \\
& + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha. \quad (6)
\end{aligned}$$

Finally, in view also of the energy bound

$$\int_{\Sigma_\tau} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \leq \int_{\Sigma_{\tau'}} J_\alpha^T[\phi] n_{\Sigma_{\tau'}}^\alpha, \quad (\text{EB-Mink})$$

which holds for all $\tau \geq \tau'$, the estimate (6) easily implies that

$$\int_{\Sigma_\tau} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \leq C \tau^{-2} \left[\int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2(\partial_v \psi)^2 \sin \theta d\theta d\phi dv + \int_{\Sigma_{\tau_1}} J_\alpha^T[\phi] n_{\Sigma_\tau}^\alpha \right] \quad (7)$$

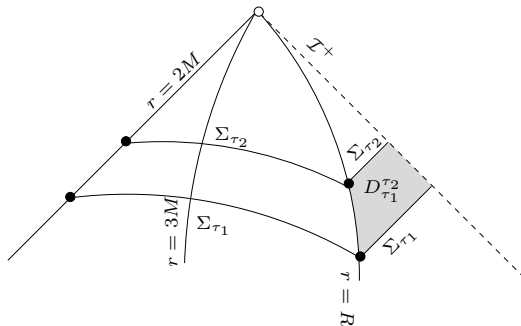
for any $\tau \geq \tau_1$.

Remark 3.1. Note that (7) essentially corresponds to (2) of the discussion of the vector field method. In view of the absence of weights in t in the multiplier used for

the proof, it is already immediate from the above that (7) follows similarly on sufficiently small non-stationary perturbations of Minkowski space. See the discussion in Section 7.

4. The Schwarzschild case

We now consider the wave equation (1) on a Schwarzschild background $(\mathcal{M}, g)$ with parameter M . See Section 2 of [8]. Let $\mathcal{D}$ denote the closure of (a connected component of) the domain of outer communications, let Σ_τ be a translation-invariant family of hypersurfaces connecting the event horizon $\mathcal{H}^+$ and null infinity $\mathcal{I}^+$ as below:



For definiteness, one can choose the family Σ_τ explicitly as follows: Given a constant $R > 3M$, first define a Regge-Wheeler coordinate $r^* = r + 2M \log(r - 2M) - R$, the null coordinates $u = t - r^*$, $v = t + r^*$ and the coordinate $t^* = t + 2M \log(r - 2M)$. Then Σ_τ can be defined to coincide with $t^* = \tau$ for $r \leq R$, and to coincide with $u = \tau$ for $r \geq R$. Let us define $D_{\tau_1}^{\tau_2}$ analogously as before, i.e., with the present normalisation of the null coordinates: $D_{\tau_1}^{\tau_2} = \{r \geq R\} \cap \{\tau_1 \leq u \leq \tau_2\}$.

It will be useful to fix a translation-invariant timelike vector field N on $J^+(\Sigma_{\tau_1})$ such that $N = \partial_t$ for $r \geq R$, say. (The coordinate vector field ∂_{t^*} in (r, t^*) coordinates will do for instance.) Let us note that (EB-Mink) is replaced by

$$\int_{\Sigma_\tau} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha \leq C \int_{\Sigma_{\tau'}} J_\alpha^N[\phi] n_{\Sigma_{\tau'}}^\alpha. \quad (\text{EB-Schw})$$

This non-degenerate energy boundedness statement was originally proven in [5] but can in fact be shown independently of decay results, as in [7] and Section 3 of [8].

The analogue of (ILED-Mink) is

$$\int_\tau^\infty \int_{r \leq R} (\chi J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha + (\partial_r \phi)^2 + \phi^2) \leq C \int_{\Sigma_\tau} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha, \quad (\text{ILED1-Schw})$$

where χ is a weight which vanishes quadratically at $r = 3M$. This type of estimate, which does not degenerate on the horizon, was originally proven in [5]. For background, see Section 4.4.1 of [8]. Commuting with T , we may obtain for instance

$$\int_\tau^\infty \int_{r \leq R} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha \leq C \int_{\Sigma_\tau} J_\alpha^N[T\phi] n_{\Sigma_\tau}^\alpha + C \int_{\Sigma_\tau} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha. \quad (\text{ILED2-Schw})$$

It is essential here that one has the non-degenerate quantity $J^N[\phi]$ on the horizon.

Before adapting the argument of Section 3 to Schwarzschild, it remains to derive the analogue of (p-WE-Mink).

As before, we may rewrite the wave equation (1) as

$$-\frac{1}{1-\frac{2M}{r}}\partial_u\partial_v\psi + \Delta\psi - \frac{2M}{r^3}\psi = 0, \quad \psi := r\phi.$$

Proceeding as before, we generate the following additional terms in the analogue of (p-WE-Mink): a boundary term at null infinity of the right sign, proportional to $2M \int_{\mathcal{I}_{\tau_1-R}^{\tau_2-R}} r^{p-3}|\psi|^2 du$ and the bulk terms $(3-p)M \int_{D_{\tau_1}^{\tau_2-R}} r^{p-4}\psi^2 du dv$,

$\int_{D_{\tau_1}^{\tau_2}} r^{p-1} \left[p - \frac{2M}{r}\right] |\partial_v\psi|^2 du dv$. Both expressions have the right sign as long as

$0 < p \leq 3$ and r is sufficiently large. Note also the new $(1 - 2M/r)$ weights in the analogue of (p-WE-Mink), which of course will play no role for large r .

Suppressing again the spherical integration from the notation, we thus obtain, for R sufficiently large, the estimate

$$\begin{aligned} & \int_{\Sigma_{\tau_2} \cap \{r \geq R\}} r^p (\partial_v \psi)^2 dv + \int_{D_{\tau_1}^{\tau_2}} r^{p-1} (p(\partial_v \psi)^2 + (2-p)|\nabla \psi|^2) dudv + \int_{\mathcal{I}_{\tau_1-R}^{\tau_2-R}} r^p |\nabla \psi|^2 \\ & \leq C \left(\int_{\Sigma_{\tau_2} \cap \{r \geq R\}} r^p (\partial_v \psi)^2 dv + \int_{\tau_1}^{\tau_2} r^p (|\nabla \psi|^2 + (\partial_v \psi)^2) d\tau|_{r=R} \right). \end{aligned} \quad (\text{p-WE-Schw})$$

Given (ILED1-Schw), (ILED2-Schw), (p-WE-Schw), (EB-Schw), one repeats the Minkowski argument until before (5).

At this point, one adds the estimate (ILED2-Schw) to obtain

$$\begin{aligned} & \int_{\tau_{n-1}}^{\tau_n} \int_{r \leq R} (J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha + \phi^2) + \int_{\Sigma_{\tau_n} \cap \{r \geq R\}} r (\partial_v \psi)^2 dv + \int_{D_{\tau_{n-1}}^{\tau_n}} ((\partial_v \psi)^2 + |\nabla \psi|^2) dudv \\ & \leq C \tau_n^{-1} \left[\int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2 (\partial_v \psi)^2 dv + \int_{\Sigma_{\tau_1}} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha \right] \\ & + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^N[T\phi] n_{\Sigma_\tau}^\alpha. \end{aligned} \quad (8)$$

One obtains as before

$$\begin{aligned} \int_{\tau_{n-1}}^{\tau_n} \int_{\Sigma_\tau} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha &\leq C\tau_n^{-1} \left[\int_{\Sigma_{\tau_1} \cap \{r \geq R\}} r^2 (\partial_v \psi)^2 dv + \int_{\Sigma_{\tau_1}} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha \right] \\ &\quad + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha + C \int_{\Sigma_{\tau_{n-1}}} J_\alpha^N[T\phi] n_{\Sigma_\tau}^\alpha \\ &\leq C\tau_n^{-1} \int_{\Sigma_{\tau_1}} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha + C \int_{\Sigma_{\tau_1}} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha + C \int_{\Sigma_{\tau_1}} J_\alpha^N[T\phi] n_{\Sigma_\tau}^\alpha. \end{aligned}$$

Identifying a good dyadic sequence, applying the boundedness statement (EB-Schw), and using the above inequality again with ϕ replaced by $T\phi$, we obtain finally

$$\int_{\Sigma_\tau} J_\alpha^N[\phi] n_{\Sigma_\tau}^\alpha \leq C\tau^{-2} \sum_{i=0}^2 \int_{\Sigma_{\tau_1}} r^2 J_\alpha^N[T^i \phi] n_{\Sigma_\tau}^\alpha. \quad (9)$$

5. Pointwise estimates

Starting from the energy decay bounds (7), (9) proven above, one can obtain pointwise decay estimates after additional commutations and applications of weighted Sobolev inequalities. To obtain decay near null infinity, we shall require commutator vector fields with weights in r , but, again in accordance with the philosophy of our method, no weights in t will be used.

Specifically, we apply commutations with cut-off versions of Ω_i , supported only near infinity, together with additional commutations with T and the red-shift vector field N (or equivalently Y) and an appeal to elliptic estimates (see Section 5.3.8 of [8]). From this approach, we obtain finally:

$$|r^{1/2}\phi| \leq C\sqrt{E}\tau^{-1}, \quad |r\phi| \leq C\sqrt{E}\tau^{-1/2}, \quad (10)$$

where

$$E = \sum_{|\alpha| \leq 2} \sum_{\Gamma=T, N, \Omega_i} \int_{\Sigma_0} r^2 (J_\mu^N(\Gamma^\alpha \phi) + J_\mu^N(\Gamma^\alpha T\phi) + J_\mu^N(\Gamma^\alpha TT\phi)) n_{\Sigma_0}^\mu.$$

6. A general formulation and the Kerr case

It should be clear from the structure of the above argument that the properties (i), (ii), (iii), as captured by suitable estimates in the form of (ILED1-Schw)–(ILED2-Schw), (p-WE-Schw), (EB-Schw), are in fact sufficient for obtaining pointwise-in-time energy decay of the form (9). In particular, we note that the stationarity of the metric is not used directly, as long as one has the above estimates (i), (ii), (iii), and (in the case where there is trapping), as long as one can absorb the errors arising from commuting with T . We shall give a general formulation in a follow-up paper. The results apply to a wide class of background metrics and can include obstacles with trapped rays, black holes, etc.

Of particular importance are the Kerr spacetimes $(\mathcal{M}, g_{a,M})$. See Section 5.1 of [8] for a discussion. In the case $|a| \ll M$, the estimate associated to (ii) follows by the results of [8, 16], whereas the estimate associated to (iii) follows from [7]. For the general case $|a| < M$, the estimates associated to (i) and (ii) will follow by forthcoming results. On the other hand, a computation reveals that the analogue of (i) follows as in the Schwarzschild case. Thus, one obtains precisely (9) for Kerr, where T is the stationary Killing field, N is a globally timelike translation-invariant vector field, and Ω_i are angular momentum operators on an ambient background Schwarzschild metric.

In the general formulation, to obtain pointwise estimates as in (10), the assumptions necessary for (9) have to be supplemented with assumptions regarding the existence of sufficiently many good commutators. Again, we leave the most general formulation to a follow-up paper. We remark here, however, that the necessary properties can be immediately explicitly verified for Kerr, for the entire range $|a| < M$, in view of the timelike nature of the span of ∂_t and ∂_ϕ outside the horizon, and the positivity of the surface gravity ensuring the validity of the commutation Theorem 7.2 of [8]. Thus, as in Schwarzschild, one obtains the pointwise (10) for Kerr, where the vector fields are interpreted as in the previous paragraph.

7. Discussion

As mentioned before, the energy decay estimate (7) corresponds exactly to what one can obtain in Minkowski space by applying the multiplier J_μ^Z associated to the generator of inverted time translations $Z = u^2\partial_u + v^2\partial_v = (t^2 + r^2)\partial_t + 2tr\partial_r$. This argument for obtaining decay estimates in Minkowski space goes back to Morawetz [14]. Analogues of the current J_μ^Z can be constructed in the Schwarzschild case [3, 5] and slowly rotating Kerr case $|a| \ll M$ [2, 8] and were used, in conjunction with an integrated decay type inequality, to obtain decay of energy in Σ_τ . Already, however, in the Kerr case, there is a loss in the power of decay which one can obtain by this method, which is in principle proportional to the parameter a . See Section 5.3.6 of [8]. The difficulty with these constructions arises from error terms in the region $r \leq R$ which inherit large weights in t . These problems could be even more profound in the setting of non-linear applications. This issue was in fact our motivation for the present work.

It is amusing to compare the role of null infinity $\mathcal{I}^+$ in the present argument with the role of the cosmological horizon in the Schwarzschild-de Sitter case studied in [6]. There, the analogue of the iteration process applied here could be continued indefinitely to yield decay bounds faster than any polynomial rate. This is precisely related to the red-shift. Indeed, this is reflected already in the Schwarzschild case at the event horizon by the fact that $J^N[\phi]$ appears without degeneration at $r = 2M$ on both sides of inequality (ILED1-Schw).

We mention finally a forthcoming Fourier-based method of Tataru and collaborators for obtaining pointwise estimates for stationary metrics starting again from

(i) and good asymptotic behaviour of the metric^a.

8. Refinements: the full decay of the vector field method

As explained in the introduction, the “full vector field method” yields faster point-wise decay of the form (3). One way to prove such a result starting from our (9) would be to apply one commutation with the scaling vector field $t\partial_t + r\partial_r$. (See work of Luk [12] for the Schwarzschild case.) This would be contrary, however, to the philosophy of our current approach, which is to avoid any multipliers and commutators with weights in t . It turns out that our approach can in fact be extended to yield more decay without such commutations, by instead commuting with ∂_v (again understood in a null (u, v) coordinate system), and Ω_i which allows applying the hierarchy of estimates (p-WE-Mink) or (p-WE-Schw) for larger weights in r , i.e. for larger values of the parameter p . This leads to additional decay for a higher-order energy flux through Σ_τ , which can then be transformed into the usual decay for ϕ using also ideas of [12]. This will be discussed in a forthcoming follow-up paper.

Acknowledgments

This work was conducted when I.R. was visiting Cambridge in February 2009. I.R. thanks the University of Cambridge for hospitality. M.D. is supported in part by a grant from the ERC. I.R. is supported in part by NSF grant DMS-0702270.

References

- [1] S. Alinhac *On the Morawetz–Keel–Smith–Sogge Inequality for the Wave Equation on a Curved Background*, Publ. RIMS, Kyoto Univ. **42** (2006), 705–720
- [2] L. Andersson and P. Blue *Hidden symmetries and decay for the wave equation on the Kerr spacetime*, arXiv:0908.2265
- [3] P. Blue and J. Sterbenz *Uniform decay of local energy and the semi-linear wave equation on Schwarzschild space* Comm. Math. Phys. **268** (2006), no. 2, 481–504
- [4] D. Christodoulou and S. Klainerman *The global nonlinear stability of the Minkowski space* Princeton University Press, 1993
- [5] M. Dafermos and I. Rodnianski *The redshift effect and radiation decay on black hole spacetimes* Comm. Pure Appl. Math. **52** (2009), 859–919.
- [6] M. Dafermos and I. Rodnianski *The wave equation on Schwarzschild-de Sitter spacetimes*, arXiv:0709.2766v1 [gr-qc]
- [7] M. Dafermos and I. Rodnianski *A proof of the uniform boundedness of solutions to the wave equation on slowly rotating Kerr backgrounds*, arXiv:0850.4309v1 [gr-qc]
- [8] M. Dafermos and I. Rodnianski *Lectures on black holes and linear waves*, arXiv:0811.0354 [gr-qc]
- [9] M. Keel, H. Smith, and C. Sogge *Almost global existence for quasilinear wave equations in three space dimensions*, J. Amer. Math. Soc. **17** (2004), 109–153
- [10] S. Klainerman *Uniform decay estimates and the Lorentz invariance of the classical wave equation* Comm. Pure Appl. Math. **38** (1985), 321–332

^aTalk of D. Tataru, Oberwolfach, October 2009

- [11] S. Klainerman and T. Sideris *On almost global existence for nonrelativistic wave equations in 3D* Comm. Pure Appl. Math. **49** (1996), 307–321
- [12] J. Luk *Improved decay for solutions to the linear wave equation on a Schwarzschild black hole*, arXiv:0906.5588
- [13] J. Metcalfe and C. D. Sogge *Global existence of null-form wave equations in exterior domains* Math. Z. (2007), 521–549
- [14] C. S. Morawetz *The limiting amplitude principle* Comm. Pure Appl. Math. **15** (1962) 349–361
- [15] C. S. Morawetz *Time decay for the nonlinear Klein-Gordon equations* Proc. Roy. Soc. Ser. A **206** (1968) 291–296
- [16] D. Tataru and M. Tohaneanu *Local energy estimate on Kerr black hole backgrounds*, arXiv:0810.5766

NONLINEAR REGULARIZING EFFECT FOR HYPERBOLIC PARTIAL DIFFERENTIAL EQUATIONS

FRANÇOIS GOLSE

*Centre de Mathématiques Laurent Schwartz, Ecole polytechnique
91128 Palaiseau cedex France
E-mail: golse@math.polytechnique.fr*

The Tartar-DiPerna compensated compactness method, used initially to construct global weak solutions of hyperbolic systems of conservation laws for large data, can be adapted in order to provide some regularity estimates on these solutions. This note treats two examples: (a) the case of scalar conservation laws with convex flux, and (b) the Euler system for a polytropic, compressible fluid, in space dimension one.

Keywords: Hyperbolic systems, Compensated compactness, Regularizing effect, Scalar conservation law, Isentropic Euler system

1. Motivation

Hyperbolic PDEs such as the wave equation are known to propagate singularities, unlike parabolic (or elliptic) PDEs, whose solutions are more regular than the corresponding data. Besides, in the context of hyperbolic PDEs, nonlinearities are responsible for the build-up of finite time singularities in the form of shock waves. Therefore, the notion of a “nonlinear regularizing effect” for hyperbolic PDEs may seem somewhat of a paradox.

Yet it has been known since the work of P. Lax [8, 9] that the evolution semigroup defined by the entropy solution $u \equiv u(t, x) \in \mathbf{R}$ of a scalar conservation law

$$\partial_t u + \partial_x f(u) = 0, \quad (t, x) \in \mathbf{R}_+^* \times \mathbf{R}$$

with strictly convex flux f is compact in $L^1(\mathbf{R})$ for each $t > 0$. On the other hand, if the flux f is linear, solving the equation above explicitly by the method of characteristics shows that whichever singularities exist in the initial data $u|_{t=0}$ are propagated and persist in the solution $u(t, x)$ for all $t > 0$. This simple example suggests that some type of nonlinearities may indeed induce limited regularization effects in hyperbolic PDEs. The purpose of the present note is to investigate that question on two examples: (a) the case of a scalar conservation law with convex flux as above, and (b) the case of the Euler system for the dynamics of a polytropic, compressible fluid.

2. Regularizing effect for scalar conservation laws

Consider the Cauchy problem

$$\begin{cases} \partial_t u + \partial_x f(u) = 0, & x \in \mathbf{R}, t > 0, \\ u|_{t=0} = u^{in}, \end{cases} \quad (1)$$

with unknown $u \equiv u(t, x) \in \mathbf{R}$ and flux $f \in C^2(\mathbf{R}; \mathbf{R})$, and assume without loss of generality that $f(0) = f'(0) = 0$.

One of the methods for constructing entropy solutions of Eq. (1) is based on the compensated compactness method proposed by Tartar [13]. A striking feature in Tartar's argument is that he obtains the compactness of some approximating sequence converging to the entropy solution of (1) *without* using any variant of the Ascoli-Arzelà theorem^a based on Sobolev (or Besov) regularity estimates.

Our main purpose in the present note is to present a method for obtaining nonlinear regularization effects in the context of hyperbolic PDEs that is inspired from Tartar's compensated compactness argument, and follows it very closely.

Theorem 2.1. *Let $a, R > 0$ and assume that $f''(v) \geq a$ for all $v \in \mathbf{R}$, while the initial data $u^{in} \in L^\infty(\mathbf{R})$ satisfies $u^{in}(x) = 0$ for a.e. $|x| \geq R$. Then, the entropy solution u belongs to the Besov space $B_{\infty,loc}^{1/4,4}(\mathbf{R}_+^* \times \mathbf{R})$; in other words*

$$\int_0^\infty \int_{\mathbf{R}} \chi(t, x)^2 |u(t, x) - u(t + s, x + y)|^4 dx dt = O(|s| + |y|)$$

as $|s| + |y| \rightarrow 0$, for each compactly supported $\chi \in C^1(\mathbf{R}_+^* \times \mathbf{R})$.

Before giving the proof of this estimate, let us compare it with earlier results in the literature.

As is well known, the optimal regularity result for Eq. (1) was obtained by Lax [9], who proved that the entropy solution $u \in BV_{loc}(\mathbf{R}_+^* \times \mathbf{R})$, as a consequence of the Lax-Oleinik one-sided inequality

$$\partial_x u(t, x) \leq \frac{1}{at}, \quad (t, x) \in \mathbf{R}_+^* \times \mathbf{R}.$$

Unfortunately, this inequality is specific to the case of scalar conservation laws in space dimension 1 with nondegenerate convex flux.

More recently, Lions-Perthame-Tadmor [10] and Jabin-Perthame [7] obtained a Sobolev regularity estimate, by using a “kinetic formulation” of the scalar conservation law (1), together with some appropriate “velocity averaging” result. Specifically, they proved that $u \in W_{loc}^{s,p}(\mathbf{R}_+^* \times \mathbf{R})$ for all $s < \frac{1}{3}$ and $1 \leq p < \frac{3}{2}$.

On the other hand, a very interesting contribution of DeLellis and Westdickenberg [2] shows that one cannot obtain better regularity in the scale of Besov spaces than $B_\infty^{1/r,r}$ for $r \geq 3$ or $B_r^{1/3,r}$ for $1 \leq r < 3$, by using *only* that the entropy production is a bounded Radon measure, without using that it is a *positive* Radon measure.

^aThe same is true of the argument used by Lax in Ref. [8].

Our result in Theorem 2.1, like the one of Lions-Perthame-Tadmor or of Jabin-Perthame, does not use the positivity of the entropy production, and therefore belongs to the DeLellis-Westdickenberg optimality class.

Sketch of the proof. The proof is split in two steps. We henceforth denote $\mathbf{D}_{s,y}\phi(t, x) := \phi(t + s, x + y) - \phi(t, x)$.

Step 1: Let u be the entropy solution of Eq. (1), and consider the two vector fields $B := (u, f(u))$ and $E := (f(u), g(u))$, where $g'(v) = f'(v)^2$ for each $v \in \mathbf{R}$. That u is the entropy solution of Eq. (1) entails the two following equalities:

$$\operatorname{div}_{t,x} B = 0, \quad \text{and} \quad \operatorname{div}_{t,x} E = -\mu,$$

where μ is a bounded Radon measure on $\mathbf{R}_+^* \times \mathbf{R}$. A variant of the Murat-Tartar div-curl lemma [12] leads to the inequality

$$\int_0^\infty \int_{\mathbf{R}} \chi^2 \mathbf{D}_{s,y} E \cdot J \mathbf{D}_{s,y} B dt dx \leq C(|s| + |y|),$$

where J denotes the rotation of an angle $\pi/2$, the function χ is C^1 with compact support in $\mathbf{R}_+^* \times \mathbf{R}$ and $C = C(\|u\|_\infty, \|\mu\|_1) > 0$. (The notations $\|u\|_{L^\infty}$ and $\|\mu\|_1$ designate respectively the sup norm of u and the total mass of μ .)

Step 2: The integrand in the l.h.s. of the inequality above is of the form

$$\begin{aligned} (w - v)(g(w) - g(v)) - (f(w) - f(v))^2 \\ = \int_v^w d\lambda \int_v^w f'(\lambda)^2 d\lambda - \left(\int_v^w f'(\lambda) d\lambda \right)^2 \geq 0 \end{aligned}$$

by the Cauchy-Schwarz inequality, as observed by Tartar [13]. In fact, the r.h.s. of the identity above can be written as the double integral:

$$\begin{aligned} (w - v)(g(w) - g(v)) - (f(w) - f(v))^2 &= \int_v^w \int_v^w (f'(\zeta) - f'(\xi)) f'(\zeta) d\xi d\zeta \\ &= \frac{1}{2} \int_v^w \int_v^w (f'(\zeta) - f'(\xi))^2 d\xi d\zeta \geq \frac{1}{2} \int_v^w \int_v^w a^2(\zeta - \xi)^2 d\xi d\zeta = \frac{a^2}{12} |w - v|^4, \end{aligned}$$

and substituting this lower bound in the inequality obtained at the end of Step 1 above entails the claimed $B_{\infty,loc}^{1/4,4}$ estimate. $\square$

Remark 2.1. The same method also works for degenerate convex fluxes, for which $f''(v) \geq 0$ for all $v \in \mathbf{R}$, but may have finitely many zeros $v_1, \dots, v_n$ of finite order — meaning that $f''(v) = O((v - v_k)^{2\beta_k})$ as $v \rightarrow v_k$ for some positive integer β_k . See Ref. [5].

Remark 2.2. The proof sketched above uses only the entropy condition with f as the entropy density. By using a family of entropies (e.g. all Kruzhkov's entropies) one can improve the argument above and obtain a Besov regularity estimate in $B_\infty^{1/3,3}$, known to be optimal according to DeLellis-Westdickenberg [2]. See Ref. [6].

3. The Euler system for polytropic compressible fluids

The Euler system governs the evolution of the density $\rho \equiv \rho(t, x) \geq 0$ and velocity field $u \equiv u(t, x) \in \mathbf{R}$ of a polytropic compressible fluid:

$$\begin{cases} \partial_t \rho + \partial_x(\rho u) = 0, & \rho|_{t=0} = \rho^{in}, \\ \partial_t(\rho u) + \partial_x(\rho u^2 + \kappa \rho^\gamma) = 0, & u|_{t=0} = u^{in}. \end{cases} \quad (2)$$

We assume that this Cauchy problem is posed for all $x \in \mathbf{R}$ and $t > 0$. The pressure is $p(\rho) = \kappa \rho^\gamma$, and, by a convenient choice of units, one can assume that $\kappa = (\gamma - 1)^2/4\gamma$. This system is known to be hyperbolic with characteristic speeds

$$\lambda_\pm = u \pm \theta \rho^\theta, \quad \text{where } \theta = \frac{\gamma - 1}{2}.$$

Besides, along C^1 solutions (ρ, u) , Euler's system assumes the diagonal form

$$\begin{cases} \partial_t w_+ + \lambda_+ \partial_x w_+ = 0, \\ \partial_t w_- + \lambda_- \partial_x w_- = 0, \end{cases}$$

where $w_\pm \equiv w_\pm(\rho, u)$ are the Riemann invariants

$$w_+ := u + \rho^\theta > u - \rho^\theta =: w_-.$$

In 1983, DiPerna [3, 4] managed to extend Tartar's method to a certain class of nonlinear hyperbolic systems with two equations in space dimension one including Euler's system (2). He proved that, given $\bar{\rho} > 0$ and assuming that $\rho^{in} - \bar{\rho}$ and u are of class C^2 and compactly supported on $\mathbf{R}$, and that $\rho^{in} > 0$ on $\mathbf{R}$, there exists a least one entropy solution of (2) defined for all $t \geq 0$ and $x \in \mathbf{R}$. DiPerna's original proof could handle only exponents of the form $\gamma = 1 + 1/(2n + 1)$ for each $n \in \mathbf{N}$. The case of an arbitrary $\gamma \in (1, 3]$ was subsequently settled by Chen [1] and Lions-Perthame-Souganidis [11].

Definition 3.1. Given an open set $\mathcal{O} \subset \mathbf{R}_+^* \times \mathbf{R}$, an entropy solution $(\rho, \rho u)$ of Eq. (2) is called admissible on $\mathcal{O}$ if and only if there exist constants $C > 0$, $u^* > 0$ and $0 < \rho_* < \rho^*$ such that $\rho_* \leq \rho \leq \rho^*$ and $|u| \leq u^*$ on $\mathcal{O}$, and, for each smooth entropy-entropy flux pair (ϕ, ψ) for the system Eq. (2),

$$\iint_{\mathcal{O}} |\partial_t \phi(\rho, \rho u) + \partial_x \psi(\rho, \rho u)| \leq C \|D^2 \phi\|_{L^\infty([\rho_*, \rho^*] \times [-\rho^* u^*, \rho^* u^*])}.$$

Any DiPerna solution whose artificial viscous approximation with viscosity $\epsilon > 0$ satisfies $\rho_\epsilon \geq \rho_*$ uniformly on $\mathcal{O}$ as $\epsilon \rightarrow 0$ is admissible on $\mathcal{O}$. Yet, the global existence of admissible solutions for initial data of arbitrary size remains an open problem at the time of this writing.

Theorem 3.1. Assume that $\gamma \in (1, 3)$ and let $\mathcal{O} \subset \mathbf{R}_+^* \times \mathbf{R}$ be open. Any admissible solution of Euler's system (2) on $\mathcal{O}$ satisfies

$$\iint_{\mathcal{O}} |(\rho, u)(t + s, x + y) - (\rho, u)(t, x)|^2 dx dt = O(\ln(|s| + |y|)^{-2})$$

as $|s| + |y| \rightarrow 0$.

The only regularity result for large data known prior to this one is due to Lions-Perthame-Tadmor [10] and Jabin-Perthame [7], for the only case $\gamma = 3$. Using a kinetic formulation of Eqs. (2) and some appropriate velocity averaging argument, they proved that ρ and $\rho u \in W_{loc}^{s,p}(\mathbf{R}_+ \times \mathbf{R})$ for all $s < \frac{1}{4}$ and $1 \leq p \leq \frac{8}{5}$. Unfortunately, the structure of the compressible Euler system (2) prevents any obvious extension of their method to the case $\gamma \in (1, 3)$. While we doubt that the regularity obtained in Theorem 3.1 is optimal, some depletion of nonlinear interactions may occur when $\gamma = 3$, since the Euler system in Riemann invariants coordinates is then decoupled into two independent Hopf (i.e. inviscid Burgers) equations. This could account for the better regularity obtained when $\gamma = 3$.

The proof of Theorem 3.1 (see Ref. [5]) is again inspired from the compensated compactness method in Ref. [3] for hyperbolic systems. It uses two special features of Eq. (2). First, the characteristic speeds are linear in terms of the Riemann invariants: $(\lambda_+, \lambda_-) = (w_+, w_-)\mathcal{A}$, where the matrix $\mathcal{A}$ is symmetric. Moreover $\mathcal{A}$ is definite positive for $\gamma > 1$, and, whenever $\gamma \in (1, 3)$, satisfies the stronger coercivity property

$$(\sinh X, \sinh Y)\mathcal{A}(X, Y)^T \geq \frac{\gamma-1}{2}(X \sinh X + Y \sinh Y), \quad X, Y \in \mathbf{R}.$$

The second property of the Euler system (2) used in the proof is that the vector field $(w_+, w_-) \mapsto (\partial_{w_-} \lambda_+ / (\lambda_+ - \lambda_-), \partial_{w_+} \lambda_- / (\lambda_- - \lambda_+))$ is a gradient.

4. Final remarks

Thus the Tartar-DiPerna compensated compactness method can be used to establish new regularizing effects in the context of hyperbolic systems of conservation laws. Open questions include (a) the case of scalar conservation laws in space dimension larger than one, (b) the case of more general pressure laws in the Euler system, and (c) the case of solutions of the Euler system with vanishing density.

References

- [1] G.Q. Chen, Proc. Amer. Math. Soc. **125** (1997), 2981–2986.
- [2] C. DeLellis, M. Westdickenberg, Ann. Inst. H. Poincaré Anal. Non Lin. **20** (2003), 1075–1085.
- [3] R. DiPerna, Arch. Rational Mech. Anal. **82** (1983), 27–70.
- [4] R. DiPerna, Comm. Math. Phys. **91** (1983), 1–30.
- [5] F. Golse, in preparation.
- [6] F. Golse, B. Perthame, in preparation.
- [7] P.-E. Jabin, B. Perthame, ESAIM Control Optim. Calc. Var. **8** (2002), 761–774.
- [8] P. Lax, Comm. Pure and Appl. Math. **7** (1954), 159–194.
- [9] P. Lax, Comm. Pure and Appl. Math. **10** (1957), 537–566.
- [10] P.-L. Lions, B. Perthame, E. Tadmor, J. Amer. Math. Soc. **7** (1994), 169–191.
- [11] P.-L. Lions, B. Perthame, P. Souganidis, Comm. Pure Appl. Math. **49** (1996), 599–638.
- [12] F. Murat, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **5** (1978), 489–507.
- [13] L. Tartar, in *Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. IV*, Res. Notes in Math., 39, Pitman, Boston, Mass.-London, 1979.

ON THE MAXWELL-KLEIN-GORDON SYSTEM IN LORENZ GAUGE

SIGMUND SELBERG* and ACHENEF TESFAHUN

*Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim, Norway*

** E-mail: sigmund.selberg@math.ntnu.no*

A well-known result due to Klainerman and Machedon says that the Maxwell-Klein-Gordon system (MKG), when written relative to the Coulomb gauge, is globally well-posed for finite-energy initial data. The proof of that result relies crucially on the null structure of the main bilinear terms of MKG in Coulomb gauge. It appears to have been widely believed that such a structure is not present in Lorenz gauge, but recent work by the authors shows that it is, and this enables us to prove finite-energy global well-posedness directly in Lorenz gauge.

Keywords: Maxwell-Klein-Gordon equations, Lorenz gauge, well-posedness, dispersive estimates, null forms

1. Introduction

The Maxwell-Klein-Gordon system (MKG) is a classical nonlinear field theory describing a spin-0 particle interacting with the electromagnetic field. It is in fact a special case of the Yang-Mills-Higgs nonlinear gauge field theory with (abelian) gauge group $SO(2) \simeq \mathbb{S}^1 \simeq U(1)$. The gauge freedom allows one to impose an additional gauge condition on the electromagnetic potential. Here we shall compare the Lorenz and Coulomb gauge conditions, each of which makes MKG a nonlinear system of PDEs, but with rather different properties. As we shall see, MKG in Lorenz gauge is a system of nonlinear hyperbolic equations, whereas in Coulomb gauge it is a combination of nonlinear hyperbolic and elliptic equations.

We are interested in the Cauchy problem for MKG starting from given initial data at time $t = 0$. MKG has a conserved total energy, which plays a decisive role in the question of global existence and regularity of solutions. The first breakthrough in this respect was achieved by Eardley and Moncrief [1], who proved global well-posedness (GWP) of MKG (in temporal gauge) for an initial data space corresponding to one degree of L^2 -differentiability more than the energy space (in fact, they proved this for the more general Yang-Mills-Higgs system). From a physical point of view it is of course of interest to know whether the same result holds for the energy space itself, a question posed by Strauss [2] and answered affirmatively

by Klainerman and Machedon [3] (who later also did the Yang-Mills case [4]).

A key point about the result of Klainerman and Machedon [3] is that they worked in Coulomb gauge, the reason being that in that gauge they could demonstrate a special structure of the bilinear terms in MKG, a so-called *null structure*, without which well-posedness in the energy space would be just out of reach. Since that seminal work, it appears to have been widely believed that such a structure is not present in Lorenz gauge, but recent work by the authors [5] shows that it is, and this enables us to prove finite-energy GWP of MKG in that gauge.

Our search for a Lorenz gauge null structure in MKG was inspired by the recent discovery, by D’Ancona, Foschi and Selberg [6], that the important Maxwell-Dirac system in Lorenz gauge has such a structure.

MKG (in three space dimensions, as considered here) is energy subcritical, meaning that the critical (scale invariant) data regularity is below the energy regularity; in fact, it is half a derivative below energy. Local well-posedness (LWP) almost down to the scale invariant regularity has been proved by Sterbenz and Machedon [7], using Coulomb gauge (see also the work of Cuccagna [8]). Whether the analogous result holds in Lorenz gauge is an open question, but one can certainly prove LWP some way below energy in Lorenz gauge (this is clear from the analysis in Ref. 5).

We will now discuss some key features of the proof of finite-energy GWP in Lorenz gauge, comparing along the way with the Coulomb gauge argument of Klainerman and Machedon. We start by presenting the MKG system.

2. The Maxwell-Klein-Gordon system

Let us set the stage: We are on the Minkowski space-time $\mathbb{R}^{1+3} = \mathbb{R}_t \times \mathbb{R}_x^3$, whose points we denote $(x^\mu)_{\mu=0,1,2,3}$, where $t = x^0$ is time and $x = (x^1, x^2, x^3)$ is the spatial position. We write $\partial_\mu = \frac{\partial}{\partial x^\mu}$, and $\partial_t = \partial_0$. The spatial gradient is denoted ∇ , and ∂ stands for the space-time gradient. Indices are raised and lowered using the metric $m = (m^{\mu\nu}) = \text{diag}(-1, 1, 1, 1)$. Repeated upper/lower Greek (resp. Roman) indices are implicitly summed over 0, 1, 2, 3 (resp. 1, 2, 3). For example, the spatial Laplacian can be written as $\Delta = \partial^j \partial_j$ and the wave operator as $\square = \partial^\mu \partial_\mu = -\partial_t^2 + \Delta$.

MKG is a nonlinear coupling of the Maxwell and Klein-Gordon equations. First recall Maxwell’s equations for the electric and magnetic fields $\mathbf{E}, \mathbf{B}: \mathbb{R}^{1+3} \rightarrow \mathbb{R}^3$:

$$\nabla \cdot \mathbf{E} = \rho, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0, \quad \nabla \times \mathbf{B} - \partial_t \mathbf{E} = \mathbf{J}. \quad (\text{M})$$

The charge density ρ and current density $\mathbf{J}$ will be specified shortly, but for the moment we consider them to be given. Formally, the middle two equations in (M) are equivalent to the existence of a 4-potential $A = (A_0, A_1, A_2, A_3) = (A_0, \mathbf{A})$ such that $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = \nabla A_0 - \partial_t \mathbf{A}$, and then (M) becomes, with $(J_\mu) = (-\rho, \mathbf{J})$,

$$\square A_\mu - \partial_\mu (\partial^\nu A_\nu) = -J_\mu. \quad (\text{M}')$$

The potential is only determined up to a *gauge transformation* $A \rightarrow A' = A + \partial\chi$ for any (sufficiently smooth) $\chi: \mathbb{R}^{1+3} \rightarrow \mathbb{R}$, and two potentials related in this way

must be considered equivalent. A solution of (M') (for given J) must therefore be understood as an equivalence class, and this implies that we have *gauge freedom*: We can choose a representative A with desirable properties, by adding a *gauge condition*. Popular gauge conditions include (i) temporal: $A_0 = 0$, (ii) Coulomb: $\nabla \cdot \mathbf{A} = 0$ and (iii) Lorenz: $\partial^\nu A_\nu = 0$. Obvious advantages of Lorenz gauge are its Lorentz [sic] invariance and the fact that (M') simplifies to $\square A_\mu = -J_\mu$.

Next, we recall the Klein-Gordon equation. In the absence of an electromagnetic field, it reads $\square\phi = m^2\phi$, where $\phi : \mathbb{R}^{1+3} \rightarrow \mathbb{C}$ is the wave function and $m > 0$ is the rest mass. The associated conservation law is $\partial_t \rho + \nabla \cdot \mathbf{J} = 0$, where $\rho = -\text{Im}(\phi \overline{\partial_t \phi})$ and $\mathbf{J} = \text{Im}(\phi \overline{\nabla \phi})$. Coupling to an external field represented by a potential A satisfying Maxwell's equations is achieved by the minimal substitution $\partial_\mu \rightarrow D_\mu^{(A)} \equiv \partial_\mu - iA_\mu$, which gives us the MKG system:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = \rho, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0, \quad \nabla \times \mathbf{B} - \partial_t \mathbf{E} = \mathbf{J} \\ D_\mu^{(A)} D^{(A)\mu} \phi = m^2 \phi, \end{array} \right. \quad (\text{MKG})$$

where $\rho = -\text{Im}(\phi \overline{D_0^{(A)} \phi})$, $\mathbf{J} = \text{Im}(\phi \overline{\nabla^{(A)} \phi})$, $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = \nabla A_0 - \partial_t \mathbf{A}$.

There is a conserved energy $\mathcal{E}(t) = \frac{1}{2} \int_{\mathbb{R}^3} (|D^{(A)}|^2 + m^2 |\phi|^2 + |\mathbf{E}|^2 + |\mathbf{B}|^2) dx$, and as discussed in the introduction, we are interested in the well-posedness of MKG starting from initial data whose energy is finite.

3. MKG in Lorenz and Coulomb gauges — a comparison

MKG is invariant under the gauge transformation $\phi \rightarrow e^{i\chi} \phi$ and $A_\mu \rightarrow A_\mu + \partial_\mu \chi$. The observables $\mathbf{E}$, $\mathbf{B}$, ρ and $\mathbf{J}$ are unaffected by this transformation. Thus, we have gauge freedom, allowing us to add a gauge condition of our choice.

MKG in Coulomb gauge reads (here $\mathbf{P}$ is the projection onto divergence-free vector fields on $\mathbb{R}^3$):

$$\left\{ \begin{array}{l} \square\phi = m^2\phi + 2i\mathbf{A} \cdot \nabla\phi - 2iA_0\partial_t\phi - i(\partial_t A_0)\phi + A_\mu A^\mu \phi \\ \Delta A_0 = |\phi|^2 A_0 - \text{Im}(\phi \overline{\partial_t \phi}) \\ \square\mathbf{A} = -\mathbf{P}(\text{Im}(\phi \overline{\nabla \phi}) + |\phi|^2 \mathbf{A}) \\ \nabla \cdot \mathbf{A} = 0, \end{array} \right. \quad (\text{MKGC})$$

whereas in Lorenz gauge MKG takes the form

$$\left\{ \begin{array}{l} \square\phi = m^2\phi + 2iA_\mu \partial^\mu \phi + A_\mu A^\mu \phi \\ \square A = -\text{Im}(\phi \overline{\partial \phi}) - A|\phi|^2 \\ \partial^\mu A_\mu = 0. \end{array} \right. \quad (\text{MKGL})$$

In both gauges, finite energy data corresponds roughly to H^1 data for (A, ϕ) , so we need to investigate whether local well-posedness (LWP) holds for such data.

A simple toy model for the above systems is

$$\square u = uDu + u^3, \quad (u, \partial_t u)|_{t=0} = (u_0, u_1) \in H^s \times H^{s-1},$$

where D stands for any directional derivative. But the work of Lindblad [9] shows that generic equations of this form are ill-posed for our target regularity $s = 1$. On the other hand, LWP holds for all $s > 1$ by dispersive estimates for the wave equation. The moral is that one needs special structure in the bilinear term, uDu , to reach $s = 1$. (The cubic term turns out to be easy to treat, since it contains no derivatives.)

4. Energy and dispersive estimates

As a warm-up, let us consider first the cubic equation $\square u = u^3$. Defining the norm $X_T(u) = \sup_{0 \leq t \leq T} (\|u(t)\|_{H^1} + \|\partial_t u(t)\|_L^2)$, we apply the energy inequality for the wave equation to obtain (here $X_0 = \|u_0\|_{H^1} + \|u_1\|_{L^2}$)

$$X_T(u) \leq C \left(X_0 + \int_0^T \| [u(t)]^3 \|_{L^2} dt \right) \leq CX_0 + C'T [X_T(u)]^3$$

where we used the Sobolev embedding $H^1 \hookrightarrow L^6$ in the last step. LWP for H^1 data then follows by a standard argument, so here the target $s = 1$ was easy to reach.

Now let us try the same for $\square u = uDu$. The energy inequality gives

$$X_T(u) \leq C \left(X_0 + \int_0^T \|u(t)Du(t)\|_{L^2} dt \right) \leq C \left(X_0 + X_T(u) \int_0^T \|u(t)\|_{L^\infty} dt \right),$$

but Sobolev embedding only allows $\|u(t)\|_{L^\infty} \leq C_s \|u(t)\|_{H^s}$ for $s > 3/2$, so we get LWP for data in H^s with $s > 3/2$ (whereas our target is $s = 1$). This classical result can however be improved (see Ref. 10) to $s > 1$ by using dispersive estimates.

The sticking point is $\int_0^T \|u(t)\|_{L^\infty} dt$. Estimating this by $T \sup_{0 \leq t \leq T} \|u(t)\|_{L^\infty}$ followed by spatial Sobolev is not a good idea, as we saw above. Instead, one must exploit integrability in time due to dispersive effects. Consider first the 0-th iterate $u^{(0)}$, satisfying $\square u^{(0)} = 0$ with data (u_0, u_1) . If the data are C_c^∞ , the wave decays due to dispersion (the wave front of a light signal is a sphere expanding with unit speed, so the finite amount of energy in the signal is dispersed over an area increasing like t^2 , hence the amplitude of the signal must attenuate over time, and in fact the rate of decay is $O(1/t)$). Essentially due to this effect, the following (nonsharp Strichartz type estimate) holds:

$$\int_0^T \|u^{(0)}(t)\|_{L^\infty} dt \leq C_{T,s} (\|u_0\|_{H^s} + \|u_1\|_{H^{s-1}}) \quad \text{iff } s > 1, \quad (1)$$

allowing one to bound the next iterate $u^{(1)}$ (satisfying $\square u^{(1)} = u^{(0)}Du^{(0)}$) in $C_t H^s$ (we omit the details).

But what about the subsequent iterates? We used a dispersive estimate for the *free* wave equation $\square u = 0$, but it is only the 0-th iterate that satisfies this equation. The remarkable fact, however, is that by foliating Fourier space by null cones (the null cone is the characteristic variety of the wave operator $\square$) any Strichartz type

estimate for the free wave equation implies a corresponding estimate for all Schwartz functions $v(t, x)$. By using this well-known fact it can be seen that (1) implies

$$\int_0^T \|v(t)\|_{L^\infty} dt \leq C_{T,s} \|v\|_{s,b} \equiv \|\langle |\tau| - |\xi| \rangle^{2b} \langle \xi \rangle^{2s} \tilde{v}(\tau, \xi)\|_{L^2_{\tau,\xi}} \quad (2)$$

for all $s > 1$, $b > 1/2$. Here $\tilde{v}(\tau, \xi)$ is the space-time Fourier transform of $v(t, x)$ and $\langle \xi \rangle = (1 + |\xi|^2)^{1/2}$. The completion of $\mathcal{S}(\mathbb{R}^{1+3})$ with respect to the norm $\|v\|_{s,b}$ will be denoted $H^{s,b}$.

The estimate (2) allows one to prove LWP of $\square u = uDu$ for $s > 1$ by iterating in the norm $\|u\|_{H_T^{s,b}}$, where $b = 1/2 + \varepsilon$ for $\varepsilon > 0$ sufficiently small, and $H_T^{s,b}$ denotes the restriction of $H^{s,b}$ to $0 \leq t \leq T$. As remarked, $s > 1$ is optimal for the toy model. For the actual MKG system, on the other hand, the energy regularity (corresponding roughly to $s = 1$) can be reached due to special structure (so-called *null structure*) of the bilinear terms, as we now discuss.

5. Null structure in bilinear terms of MKG

Consider first the bilinear term $\mathbf{A} \cdot \nabla \phi$. Recall the splitting

$$\mathbf{A} = \underbrace{-\Delta^{-1} \nabla \times \nabla \times \mathbf{A}}_{\text{divergence-free}} + \underbrace{\Delta^{-1} \nabla (\nabla \cdot \mathbf{A})}_{\text{curl-free}} \equiv \mathbf{A}^{\text{df}} + \mathbf{A}^{\text{cf}}.$$

Klainerman and Machedon observed that (here $\mathbf{w} = \Delta^{-1} \nabla \times \mathbf{A}$)

$$\mathbf{A}^{\text{df}} \cdot \nabla \phi = -(\nabla \times \mathbf{w}) \cdot \nabla \phi = -\epsilon^{ijk} \partial_j w_k \partial_i \phi = \epsilon^{kji} \partial_j w_k \partial_i \phi = (\nabla w_k \times \nabla \phi)^k.$$

This suggests that instead of $\square u = uDu$ it is more realistic to consider the model

$$\square u = \left(\frac{\nabla}{|\nabla|} u \times \nabla u \right)_i. \quad (3)$$

This turns out to be much better behaved than $\square u = uDu$. To see why, we first recall the well-known fact that the worst product interaction of waves is when both waves concentrate along a null ray in Fourier space. That is, if we denote the space-time frequencies of the two waves by $X = (\tau, \xi)$ and $Y = (\lambda, \eta)$, the worst case is when X and Y are null and collinear. But in that situation the symbol of the right hand side of (3) vanishes! (This type of cancellation property is known in general as *null structure*, since it has to do with the geometry of the null cone.) Indeed, the symbol is (here θ is the angle between the spatial frequencies ξ and η) $(\xi/|\xi|) \times \eta = O(|\eta| \sin \theta)$, and this vanishes because collinearity implies $\theta = 0$ or π .

Now, in Coulomb gauge, $\mathbf{A} = \mathbf{A}^{\text{df}}$, hence $\mathbf{A} \cdot \nabla \phi$ has a good structure. (In MKGL there is also a term $A_0 \partial_t \phi$ of course, but this is easier to treat, since A_0 satisfies an elliptic equation with quite good regularity properties.)

What then about Lorenz gauge? The term $A_\mu \partial^\mu \phi$ can easily be seen to have a null structure. Indeed, on the Fourier transform side this term becomes a convolution integral of

$$\widetilde{A}_\mu(X) \widetilde{\partial^\mu \phi}(Y) = iY^\mu \widetilde{A}_\mu(X) \widetilde{\phi}(Y). \quad (4)$$

But the Lorenz gauge condition, $\partial^\mu A_\mu = 0$, says that $X^\mu \widetilde{A}_\mu(X) = 0$, so if the space-time frequencies X and Y become parallel, then also $Y^\mu \widetilde{A}_\mu(X) = 0$, hence (4) vanishes. It is possible to quantify this null cancellation in terms of the angle between the space-time frequencies, but it turns out to be more convenient to get rid of the temporal frequencies in the null form symbol, by splitting A and ϕ using standard spectral decompositions for the wave and Klein-Gordon equations; see Ref. 5 for the details.

In addition to the null structure in the term $A_\mu \partial^\mu \phi$ in Lorenz gauge, there is also a null structure in the wave equations for $\mathbf{E}$ and $\mathbf{B}$, independently of the choice of gauge. Using these null structures, and product laws for the $H^{s,b}$ spaces (see Ref. 11 for a discussion of such estimates), we are then able to prove the desired finite-energy result for MKG in Lorenz gauge; see Ref. 5.

References

- [1] D. M. Eardley and V. Moncrief, *Comm. Math. Phys.* **83**, 171 (1982).
- [2] W. A. Strauss, *Nonlinear wave equations*, CBMS Regional Conference Series in Mathematics, Vol. 73 (Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1989).
- [3] S. Klainerman and M. Machedon, *Duke Math. J.* **74**, 19 (1994).
- [4] S. Klainerman and M. Machedon, *Ann. of Math.* **142**, 39 (1995).
- [5] S. Selberg and A. Tesfahun, Finite-energy global well-posedness of the Maxwell-Klein-Gordon system in Lorenz gauge, preprint 2009 (submitted).
- [6] P. D’Ancona, D. Foschi and S. Selberg, Null structure and almost optimal local well-posedness of the Maxwell-Dirac system, to appear in American Journal of Math.
- [7] M. Machedon and J. Sterbenz, *J. Amer. Math. Soc.* **17**, 297 (2004).
- [8] S. Cuccagna, *Comm. Partial Differential Equations* **24**, 851 (1999).
- [9] H. Lindblad, *Amer. J. Math.* **118**, 1 (1996).
- [10] G. Ponce and T. C. Sideris, *Comm. Partial Differential Equations* **18**, 169 (1993).
- [11] P. D’Ancona, D. Foschi and S. Selberg, Atlas of products for wave-Sobolev spaces, preprint (2009).

A REGULARITY THEOREM FOR ENERGY CRITICAL WAVE-MAPS

JACOB STERBENZ

*Department of Mathematics,
University of California, San Diego (UCSD)
La Jolla, CA 92093-0112 USA
E-mail: jsterben@math.ucsd.edu*

DANIEL TATARU

*Department of Mathematics,
University of California, Berkeley (UCB)
Berkeley, CA 94720-3840 USA
E-mail: tataru@math.berkeley.edu*

In this article we review our recent proof of global regularity for the Wave-Map equations from $(2 + 1)$ dimensional Minkowski space into arbitrary compact manifolds $\mathcal{M}$.

Keywords: Wave-Map; Critical Regularity; Dispersion/Collapse Dichotomy

1. Introduction

Wave-Maps (WM) may be thought of as an analog of the scalar wave equation with values in a manifold Riemannian $\mathcal{M}$. If $\mathcal{M}$ is compact, then by Nash's theorem we may isometrically embed it in some Euclidean space $\mathbb{R}^N$, in which case the WM equations are formally critical points of the constrained variational problem:

$$D\mathcal{L}[\Phi] = 0, \quad \Phi \in \mathcal{M} \subseteq \mathbb{R}^N,$$

were (again formally):

$$\mathcal{L}[\Phi] := \int_{\mathbb{R}^{n+1}} \partial^\alpha \Phi \cdot \partial_\alpha \Phi \, dx dt.$$

Here $\alpha = 0, \dots, n$ are Lorentzian indices. In terms of a fixed embedding for $\mathcal{M} \subseteq \mathbb{R}^N$ the Euler-Lagrange equations are written as:

$$\partial^\alpha \partial_\alpha \Phi = -\mathcal{S}(\Phi)(\partial^\alpha \Phi, \partial_\alpha \Phi), \tag{1}$$

where $\mathcal{S} : T\mathcal{M} \times T\mathcal{M} \rightarrow N\mathcal{M}$ is the second fundamental form on $\mathcal{M}$.

Wave-Maps into $\mathcal{M} = \mathbb{S}^2$, the so called “sigma-models”, arise in simplified quantum-field theories (see [1]). They are also related to long time evolution problems for anti-ferromagnetic chains (see [2]). On the other hand, WM into $\mathcal{M} = \mathbb{H}^2$

arise as simplified models for Einstein's equations with a $U(1)$ symmetry imposed (see [3]).

A fundamental property of the WM evolution is that it preserves the usual Dirichlet energy:

$$E[\phi](t) = \int_{\mathbb{R}^n} (|\partial_t \Phi(t)|^2 + |\nabla_x \Phi(t)|^2) dx. \quad (2)$$

Equally important is that solutions to WM are invariant with respect to the change of scale:

$$\Phi(t, x) \rightsquigarrow \Phi(\lambda t, \lambda x), \quad \lambda > 0.$$

From these two properties, one sees that the spatial dimension $n = 2$ plays the distinguished role of being *energy critical*. That is, the energy is dimensionless with respect to rescalings of the above form for $n = 2$. One interesting consequence of this is that energy conservation does not necessarily prevent blowup solutions (compare to [4]), although energy size is an overall indicator of the amount of interaction needed to form a blowup. Our results quantify this relation precisely. The remainder of this article refers to the $n = 2$ case of (1).

1.1. *Focusing vs. Defocusing Classes of Manifolds and Solitons*

For constant curvature targets $\mathcal{M}$, one may split the WM equations into two families: “focusing” and “defocusing”. These correspond to positive and negative (resp.) target curvature. In the focusing case there may be non-trivial “soliton” type solutions. For example, if $\mathcal{M} = \mathbb{S}^2$ then any rational function $\Phi_0(z) : \mathbb{R}^2 \approx \mathbb{C} \rightarrow \mathbb{S}^2$ is an example of such an object. Such a map solves the time independent analog of (1), the so called Harmonic-Map equation. Moreover, WM are Lorentz invariant so one may form traveling wave solutions from Φ_0 , moving at any velocity less than light and in any direction, simply by boosting. For example, boosting along the x^1 axis we obtain the explicit time dependant solutions:

$$\Phi_0(x^1, x^2) \rightsquigarrow \Phi_0(\sqrt{\lambda^2 - 1}t + \lambda x^1, x^2), \quad \lambda > 1.$$

This justifies the terminology “soliton”.

In the case of negative curvature, for example $\mathcal{M} = \mathbb{H}^2$ or some quotient^a thereof, it is well known that there are no non-trivial harmonic maps, and hence no solitons in our terminology.

We pause to remark here that it has been known for some time that positive (spherical) or negative target curvature controls the critical existence/regularity theory of the corresponding elliptic (see [5]) and parabolic problems (see [6] and [7]). Our results are a direct analog of this well established elliptic/parabolic theory.

^aOf course one cannot isometrically embed a non-compact hyperbolic target smoothly into $\mathbb{R}^N$, but often the study of a compact quotient suffices for basic regularity questions.

Another fruitful analogy is to compare WM with the family of mass (resp.) energy critical non-linear Schrödinger (resp.) wave equations in 2D (resp.) 3D:

$$\frac{1}{i}\partial_t u - \Delta_x u = \lambda |u|^2 u, \quad \partial^\alpha \partial_\alpha u = -\lambda |u|^4 u. \quad (3)$$

In the focusing case, $\lambda > 0$, one has the existence of non-linear bound states u_0 which can also be made into traveling waves via the symmetries of the equations. For both equations, these solitons are unstable and lead to mass (resp.) energy concentration type blowup solutions (see [8] and [9]). However, if the size of the initial data is less than that of the groundstate soliton u_0^* in a critical conserved norm, then one has global well-posedness (see [10] and [11]). On the other hand, in the defocusing case, $\lambda < 0$, there are no solitons and one has global regularity for a wide class of large initial data (see [12]^b and [13]).

Finally, returning to the case of energy critical WM, it has recently been shown that in the focusing case soliton instability leads to blowup in finite time (see [14] and [15]). Taken together with the results reviewed here, this more or less completes the regularity picture for critical WM.

2. Main Results

Our main regularity results are the following:

Theorem 2.1 (A General Regularity Theory). *Let $\Phi[0] = (\Phi_0, \Phi_1) \in \mathcal{M} \times T\mathcal{M}$ be a (possibly large) C^∞ set of initial data, with $\mathcal{M}$ compact and supporting no non-trivial harmonic Maps $\Psi : \mathbb{S}^2 \rightarrow \mathcal{M}$. Then there is a unique global C^∞ solution $\Phi(t)$ with this initial data. Furthermore, if $\widetilde{\mathcal{M}}$ covers such a compact $\mathcal{M}$ then any C^∞ data $\Phi[0] = (\Phi_0, \Phi_1) \in \widetilde{\mathcal{M}} \times T\widetilde{\mathcal{M}}$ also yields a unique smooth global solution into $\widetilde{\mathcal{M}}$.*

This in turn follows from a more precise result for compact targets, as well as a simple covering argument and the finite speed of propagation. Specifically, we have shown that:

Theorem 2.2 (The Bubbling Conjecture). *If a wave-map Φ to a compact target $\mathcal{M}$ blows up in finite time, there must be a subsequence of boosted rescalings $\Phi^{(n)}$ which converge strongly to a non-trivial Harmonic-Map $\Psi : \mathbb{S}^2 \rightarrow \mathcal{M}$. Furthermore, the energy radius of the blowup must shrink at a faster than linear (i.e. self-similar) rate.*

Finally, it is to be expected that in the absence of soliton type behavior one should have asymptotic wave profiles that are in some sense linear. In particular there should be eventual dispersion of the entire solution. A quantitative, although somewhat weak, form of this statement is contained in the following:

^bFor the NLS, a defocusing non-linearity makes the problem subcritical, so the analogy is somewhat weak.

Theorem 2.3 (Large Data Scattering). *If there are no non-trivial Harmonic-Maps $\Psi : \mathbb{S}^2 \rightarrow \mathcal{M}$ with $\mathcal{M}$ compact, then the global solution $\Phi(t)$ belongs to $L_t^6(\dot{W}_x^{6, \frac{1}{2}})$.*

We remark that all of the above results can be made conditional on the size of the energy. That is, in the cases where $\mathcal{M}$ does support non-trivial harmonic maps, one has global regularity and dispersion below the ground-state energy.

2.1. Concurrent Results

Along with the above mentioned results, there have been several recent papers specializing to the case of constant curvature hyperbolic targets. Following a long series of works (see [16], [17], [18], [19], [20], [21]), T. Tao provides global regularity and scattering for critical WM into all hyperbolic spaces $\mathbb{H}^n$. A similar result was also recently obtained for the case of the hyperbolic plane $\mathbb{H}^2$ by J. Krieger and W. Schlag (see [22]).

2.2. Brief Overview of the Proof

Our proof is broken down into two modular components, appearing in two separate papers:

- (1) We prove an ϵ -regularity theorem which says that WM remain regular with the expected dispersive behavior as long as a weak scale invariant norm is sufficiently small. This norm is much coarser than the energy norm (2), and therefore allows for large energy solutions that are dispersed in phase space (i.e. either physical or frequency space). This result appears in [23].
- (2) In our second paper [24], we establish a series of results that provide a dichotomy between energy concentration and energy dispersion. Roughly speaking, these results say that in the absence of the previous “energy dispersed” scenario, one must have concentration of energy into a non-trivial entire harmonic map $\Phi_0 : \mathbb{R}^2 \rightarrow \mathcal{M}$. This uncovers a regularity picture consistent with what is known for the Harmonic-Map heat flow [6].

3. The ϵ -Regularity Theorem

The precise statement is expressed using a dispersive norm S introduced in [25], [26], and [27]:

Theorem 3.1 (The Energy Dispersed Regularity Theorem). *There exists two functions of the energy, $0 < \epsilon(E) \ll 1$ and $F(E) \gg 1$, such that if Φ is a Wave-Map defined on an interval I with $\sup_k \|P_k \Phi\|_{L_{t,x}^\infty[I]} \leq \epsilon$, then one also has $\|\Phi\|_{S[I]} \leq F$. Furthermore, if $\|\Phi\|_{S[I]} < \infty$ then Φ extends in a regular way to a larger interval $J \supset I$.*

An detailed account of the proof of this result is well beyond the scope of the present article, and we shall give only the briefest outline for experts having some familiarity with the subject:

3.1. Renormalization

At the critical regularity level, the equations (1) are no longer perturbative in the sense that the right hand side of the equation can be directly separated. Instead, one needs to exploit the geometric nature of the equation to rewrite it in the frequency localized form:

$$\square_{A_{<k}} P_k \Phi = \text{Error} , \quad (4)$$

where $\square_{A_{<k}}$ is a magnetic wave equation with $\mathfrak{so}(N)$ connection $A_{<k}$ made up from the lower frequency components of Φ , and Error is a term which may be treated via perturbation theory, at least when the energy is small. This idea was originally introduced by T. Tao (see [26]), and later refined and generalized in the work of the second author (see [27]).

It turns out that the curvature of $A_{<k}$ is sufficiently regular under the assumptions of Theorem 3.1, but *it is not necessarily small*. However, we are able to provide a large data gauge transformation that flattens this connection enough so that the resulting equation can be controlled, although this procedure does not completely reduce things to a perturbative problem as in the small energy case.

3.2. Null-Form Estimates

The Error term from line (4) contains many trilinear expressions of the form:

$$\mathcal{N}(\Phi^{(1)}, \Phi^{(2)}, \Phi^{(3)}) = \Phi^{(1)} \partial^\alpha \Phi^{(2)} \partial_\alpha \Phi^{(3)} ,$$

which are called *null-forms*. Such expressions were originally studied in the work of Klainerman/Machedon (see [28] and [29]) to provide a subcritical regularity theory for equations of the form (1). This leads to a sophisticated family of function spaces and estimates, finalized in their most delicate and precise form in the work of the second author (see [25]) and the subsequent work of T. Tao (see [26]).

However, these works are not enough to prove satisfactory bounds under the weak assumptions of Theorem 3.1. Instead, it is necessary to also use the refined null-form estimates for transverse waves of T. Wolff and T. Tao (see [30] and [31]). Finally, to transfer these estimates to the WM setting we also need some further function space technology, the so called (U, V) -spaces, developed by the second coauthor and collaborators in a series of works (see [32] and [33]).

3.3. Induction and Linearization

In addition to renormalization and multilinear estimates, the proof of Theorem 3.1 requires an additional linearization and induction argument. The inductive part is

similar in spirit of Bourgain’s “induction on energy” strategy (see [34]), although its purpose in our work takes on a somewhat different role.

Roughly speaking, the idea is to first work with equation (4) for very low frequencies $k \ll 0$, and then to inductively pass the estimates thus gained to higher and higher wave numbers k . Unfortunately a direct application of this strategy would lead to uncontrolled divergences unless one can first show that the magnetic wave equation $\square_{A_{<k}}$ is essentially perturbative on sufficiently small time intervals I , even for large curvature connections $A_{<k}$. This is not entirely possible, but satisfactory bounds are available via a delicate argument that requires some additional modifications of the standard trilinear null-form estimates of [26], used in conjunction with the conservative nature of the magnetic wave equation evolution.

4. Concentration vs. Dispersion Dichotomies

The main results here are all of the following form:

Theorem 4.1 (Concentration-Dispersion Dichotomy). *Let Φ be a wave-map defined in the interior of a backward cone $r < -t$. Then exactly one of the following two possibilities holds:*

- (1) *There exists a sufficiently small time $t_0 < 0$ and an extension of Φ to the slab $[t_0, 0] \times \mathbb{R}^2$ such that:*

$$\sup_k \sup_{t \in [t_0, 0]} \|P_k \Phi(t)\|_{L_x^\infty} \leq \epsilon(E) .$$

- (2) *There exists a sequence of points $(t_n, x_n) \rightarrow 0$ and a sequence of scales $t_n^{-1} \lambda_n \rightarrow 0$ such that the sequence of functions:*

$$\Phi^{(n)}(x) = \Phi(t_n, \lambda_n x + x_n) ,$$

converges strongly on every compact ball $\{|x| \leq R\}$ to a harmonic map $\Psi : \mathbb{R}^2 \rightarrow \mathcal{M}$ of non-trivial energy.

This Theorem is proved in a sequence of steps. The first is to use the regularity Theorem 3.1 to detect a non-trivial pocket of coherent energy in the case where option (1) fails. Second, we use several *non-perturbative* energy estimates^c to show that this pocket of energy must propagate in such a way that it remains in a strictly time-like cone $r < (1 - \gamma)|t|$, where $0 < \gamma \ll 1$. At the same time, we use an additional weighted energy estimate to show that a time-like component of the energy density of Φ is converging to zero. Taken together these imply that in the absence of case (1) in the above Theorem, we must have that:

- (i) The Wave-Map Φ contains a pocket of energy that is asymptotically uniformly time-like.

^cThese estimates are all provided via a multiplier approach, and are similar to estimates obtained by M. Grillakis, T. Tao, and Kenig/Merle (see [16], [35], and [11]).

- (ii) After an additional sequence of Lorentz boosts, this pocket of time-like energy is becoming stationary in the sense that $\partial_t \Phi \rightarrow 0$.

From these two facts taken together, and with the use of an additional rescaling and compactness argument, we are able to produce a non-trivial entire harmonic map $\Phi_0 : \mathbb{R}^2 \rightarrow \mathcal{M}$.

5. Open Problems

The following is a brief list of open problems that are now more tractable given the current state of our understanding. However, the last two problems are still very much out of reach as of this writing:

- Provide a detailed description of the dynamics in non-equivariant blowups into $\mathbb{S}^2$ close to ground-state energy.
- Prove the existence and stability of “bubble-trees”, i.e. multisoliton blowup solutions.
- Establish a partial regularity theory for supercritical WM; e.g. prove finiteness of $\mathcal{H}^{n-2}(\text{sing})$ (Hausdorff dimension) where *sing* is the singular set at the first blowup time.
- Establish the regularity picture for critical 4D Yang-Mills (instanton threshold conjecture). Here the linearized gauge field interaction is *much* less trivial than for WM.
- Establish a regularity theory for the $U(1)$ symmetric Einstein equations of an expanding cosmology. There are perhaps some interesting new quasilinear effects here, e.g. (unstable) naked singularities. This is probably a good candidate for the detailed study of low regularity quasilinear wave equations.
- Is there a partial regularity result for supercritical WM into negative targets that is *better* than what one would expect from scaling; i.e. finite $\mathcal{H}^{n-2}(\text{sing})$? In particular, is there some notion of Federer’s dimension reduction argument in this case?

References

- [1] R. Rajaraman, *Solitons and instantons* (North-Holland Publishing Co., Amsterdam, 1982). An introduction to solitons and instantons in quantum field theory.
- [2] J. Shatah and C. Zeng, *Comm. Math. Phys.* **262**, 299 (2006).
- [3] Y. Choquet-Bruhat and V. Moncrief, *Ann. Henri Poincaré* **2**, 1007 (2001).
- [4] S. Klainerman and M. Machedon, *Ann. of Math. (2)* **142**, 39 (1995).
- [5] J. Sacks and K. Uhlenbeck, *Ann. of Math. (2)* **113**, 1 (1981).
- [6] M. Struwe, *Comment. Math. Helv.* **60**, 558 (1985).
- [7] K.-C. Chang, W. Y. Ding and R. Ye, *J. Differential Geom.* **36**, 507 (1992).
- [8] F. Merle and P. Raphael, *Ann. of Math. (2)* **161**, 157 (2005).
- [9] J. Krieger, W. Schlag and D. Tataru, *Duke Math. J.* **147**, 1 (2009).
- [10] M. I. Weinstein, *Comm. Math. Phys.* **87**, 567 (1982/83).
- [11] C. E. Kenig and F. Merle, *Acta Math.* **201**, 147 (2008).

- [12] T. Cazenave, *Semilinear Schrödinger equations*, Courant Lecture Notes in Mathematics, Vol. 10 (New York University Courant Institute of Mathematical Sciences, New York, 2003).
- [13] M. G. Grillakis, *Ann. of Math. (2)* **132**, 485 (1990).
- [14] J. Krieger, W. Schlag and D. Tataru, *Invent. Math.* **171**, 543 (2008).
- [15] I. Rodnianski and J. Sterbenz, On the formation of singularities in the critical $O(3)$ sigma-model, to appear in *Annals of Math.*
- [16] T. Tao, Geometric renormalization of large energy wave maps, in *Journées “Équations aux Dérivées Partielles”*, (École Polytech., Palaiseau, 2004) pp. Exp. No. XI, 32.
- [17] T. Tao, Global regularity of wave maps III. Large energy from $\mathbb{R}^{2+1}$ to hyperbolic spaces., arXiv:0805.4666.
- [18] T. Tao, Global regularity of wave maps IV. Absence of stationary or self-similar solutions in the energy class., arXiv:0806.3592.
- [19] T. Tao, Global regularity of wave maps V. Large data local well-posedness in the energy class., arXiv:0808.0368.
- [20] T. Tao, An inverse theorem for the bilinear L^2 Strichartz estimate for the wave equation., arXiv:0904.2880.
- [21] T. Tao, Global regularity of wave maps VI. Minimal energy blowup solutions, in preparation.
- [22] J. Krieger and W. Schlag, Concentration compactness for critical wave-maps, arXiv:0908.2474.
- [23] J. Sterbenz and D. Tataru, Energy dispersed large data wave maps in 2+1 dimensions, arXiv:0906.3384.
- [24] J. Sterbenz and D. Tataru, Regularity of wave-maps in dimension $2 + 1$, preprint.
- [25] D. Tataru, *Amer. J. Math.* **123**, 37 (2001).
- [26] T. Tao, *Comm. Math. Phys.* **224**, 443 (2001).
- [27] D. Tataru, *Amer. J. Math.* **127**, 293 (2005).
- [28] S. Klainerman and M. Machedon, *Comm. Pure Appl. Math.* **46**, 1221 (1993).
- [29] S. Klainerman and M. Machedon, *Duke Math. J.* **81**, 99 (1995), A celebration of John F. Nash, Jr.
- [30] T. Wolff, *Ann. of Math. (2)* **153**, 661 (2001).
- [31] T. Tao, *Math. Z.* **238**, 215 (2001).
- [32] H. Koch and D. Tataru, *Comm. Pure Appl. Math.* **58**, 217 (2005).
- [33] H. Koch and D. Tataru, *Int. Math. Res. Not. IMRN*, Art. ID rnm053, 36 (2007).
- [34] J. Bourgain, *Global solutions of nonlinear Schrödinger equations*, American Mathematical Society Colloquium Publications, Vol. 46 (American Mathematical Society, Providence, RI, 1999).
- [35] M. G. Grillakis, On the wave map problem, in *Nonlinear wave equations (Providence, RI, 1998)*, , Contemp. Math. Vol. 263 (Amer. Math. Soc., Providence, RI, 2000) pp. 71–84.

OTHER TALKS

INVITED TALKS

ON THE UNIQUENESS OF STATIONARY BLACK HOLES IN VACUUM

ALEXANDRU IONESCU

University of Wisconsin, Madison

E-mail: ionescu@math.wisc.edu

It is widely expected that the domain of outer communication of a four-dimensional regular, stationary, black hole solution of the Einstein vacuum equations has to be isometric to the domain of outer communication of a Kerr black hole. So far the conjecture is known to be true in the case of real-analytic spacetimes, with suitable regularity assumptions, through the work of Hawking, Carter, and Robinson.

I will discuss some recent work aimed at removing the real-analyticity assumption: a uniqueness theorem for regular spacetimes that satisfy a suitable identity on the two-dimensional bifurcation sphere, and a uniqueness theorem for regular spacetimes that are assumed to be "close" to some Kerr spacetime.

H-PRINCIPLE AND FLUID DYNAMICS

CAMILLO DE LELLIS*, L. SZÉKÉLYHIDI

**Institut für Reine Mathematik, Universität Zürich*

E-mail: delellis@math.uzh.ch

The Euler equations are perhaps the oldest system of partial differential equations derived in fluid dynamics. The h principle is a concept introduced in the seventies by Gromov to unify several counterintuitive phenomena in differential geometry. Two famous instances of the h -principle are the Nash-Kuiper C^1 isometric embeddings and the Smale's Eversion Theorem.

LANDAU DAMPING

CÉDRIC VILLANI

Institut Henri Poincaré, Paris

E-mail: cvillani@umpa.ens-lyon.fr

The celebrated and mysterious Landau damping predicts relaxation of the kinetic distribution of a classical collisionless plasma, without any diffusion or irreversibility. For the past sixty years, Landau damping has been treated exclusively for the linearized model, with the exception of very partial results. In a joint work with Clement Mouhot, we close this gap by presenting a solution to the nonlinear damping problem. Unexpected mathematical and physical developments will be discussed.

WELLPOSEDNESS OF THE TWO AND THREE DIMENSIONAL FULL WATER WAVE PROBLEM

SIJUE WU

University of Michigan, Ann Arbor

E-mail: sijue@umich.edu

We consider the problem of global in time existence and uniqueness of solutions of the 2-D infinite depth full water wave equation. It is known that this equation has a solution for a time period $[0, T/\epsilon]$ for initial data of type $\epsilon\Phi$, where T depends only on Φ . We show that for such data there exists a unique solution for a time period $[0, e^{T/\epsilon}]$. This is achieved by better understandings of the nature of the nonlinearity of the water wave equation. The paper is now accepted by Invent. Math. and is available at <http://www.springerlink.com/content/c3v28534883m7131/fulltext.pdf>

CONTRIBUTED TALKS

CRITICAL POINTS AND LEVEL SETS OF SOLUTIONS TO ELLIPTIC PDE'S

ALBERTO ENCISO

ETH Zürich

E-mail: alberto.enciso@math.ethz.ch

The study of the qualitative properties of the solutions to the equation

$$\Delta u = 0 \quad \text{in } \mathbf{R}^n \setminus \overline{\Omega}, \quad u|_{\partial\Omega} = 1,$$

which vanish at infinity was pioneered by Faraday and Maxwell in the 19th century. In physics, the level sets of the function u are equipotential surfaces, whereas its critical points correspond to the equilibrium positions for the motion of a charged particle in presence of the conducting surface $\partial\Omega$. The relevance of Faraday's lines of force, which are simply the integral curves of the gradient vector field $-\nabla u$, was unveiled in Maxwell's celebrated treatise, where they led to the first hints of what is nowadays known as Morse theory.

In this talk we will use related ideas to analyze some geometrical properties of the solutions to the above exterior boundary problem, Ω being a bounded domain with C^2 connected boundary. We prove that the critical set of u can be nonempty (in fact, of codimension 3) even when Ω is contractible, thereby settling a question posed by Kawohl, discuss sufficient geometric criteria for the absence of critical points in this problem and analyze the properties of the critical set for generic domains. Time permitting, related

problems on Riemannian manifolds will be discussed as well. Our results hinge on a combination of classical potential theory, transversality techniques and the qualitative theory of dynamical systems, and are mainly extracted from the references

- [1] A. Enciso, D. Peralta-Salas, Critical points and level sets in exterior boundary problems, *Indiana Univ. Math. J.*, in press.
- [2] A. Enciso, D. Peralta-Salas, Critical points and generic properties of Green functions on complete manifolds, submitted for publication.

ON SOME CLASSES OF THE PARTIAL DIFFERENTIAL EQUATIONS WITH NON-TRIVIAL SYMMETRY GROUPS

VASYL M. FEDORCHUK*, VOLODYMYR I. FEDORCHUK

**Pedagogical University, Institute of Mathematics, Krakow*
E-mail: vasfed@gmail.com

In many cases the mathematical models of various processes can be described by means of partial differential equations (PDEs) in the spaces of different dimensions and different types. It is well known that many of the PDEs, which are useful in theoretical and mathematical physics, have non-trivial symmetry groups (see, for example, [1,2,3]).

The knowledge of the subgroup structure of the local Lie groups of the point transformations allow us to construct and investigate PDEs with non-trivial symmetry groups (see, for example, [1,2,3,4,5]).

Some new results obtained while applying of the subgroup structure of the local Lie groups of the point transformations to construct and investigate the PDEs with non-trivial symmetry groups will be presented.

- [1] W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., 1977).
- [2] P.J. Olver, *Applications of Lie Groups to Differential Equations* (Springer-Verlag, New York, 1986).
- [3] W.I. Fushchych and A. G. Nikitin, *Symmetries of Equations of Quantum Mechanics* (Allerton Press Inc., New York, 1994).
- [4] S. Lie, G. Scheffers, *Vorlesungen über Differentialgleichungen mit bekannten infinitesimalen Transformationen* (Leipzig, 1891).
- [5] L. V. Ovsyannikov, *Group Analysis of Differential Equations* (Academic Press, New York, 1982).

THE HARDY INEQUALITY AND THE ASYMPTOTIC BEHAVIOUR OF THE HEAT EQUATION IN TWISTED DOMAINS

DAVID KREJČÍŘÍK

Nuclear Physics Institute ASCR
E-mail: david@ujf.cas.cz

In this talk we revise a recently established Hardy inequality in twisted quantum waveguides on the background of transience of the Brownian motion. We begin by recalling the classical Hardy inequality and its relation to geometric, spectral, stochastic and other properties of the underlying Euclidean space. After discussing the complexity of the problem when reformulated for quasicylindrical subdomains, we focus on the prominent class of tubes. As the main result, we show that the geometric deformation of twisting

yields an improved decay rate for solutions of the heat equation in three-dimensional tubes of uniform cross-section. This is a joint work with Enrique Zuazua.

A RIGIDITY PROPERTY OF THE DEVELOPMENT OF CONFORMALLY FLAT DATA IN GENERAL RELATIVITY

JUAN ANTONIO VALIENTE KROON

School of Mathematical Sciences Queen Mary, University of London

E-mail: j.a.valiente-kroon@qmul.ac.uk

There is available a regular finite initial value problem for the conformal Einstein field equations which is obtained by using the propagation equations implied by the conformal Einstein field equation in a particular conformal gauge. This conformal gauge is based on certain conformal invariants (conformal geodesics) and renders a finite representation of the conformal boundary of the spacetime, where its location is known a priori. Given a time symmetric initial data set for the vacuum Einstein field equations which is conformally flat in a neighbourhood of null infinity, I use this regular initial value problem to show that its solutions extend smoothly through the sets where null infinity "touches" spatial infinity if and only if the initial data coincides with data which is exactly Schwarzschild in a neighbourhood of infinity. The relevance of this result is put in context by noticing that the Schwarzschild spacetime is the only vacuum spacetime with conformally flat hypersurfaces. The main result is obtained from a careful analysis of the algebraic structure of the solutions of certain transport equations implied by the restriction of the conformal Einstein field equations to the "cylinder at spatial infinity" – a total characteristic hypersurface which arises in this conformal representation of the spacetime. Possible generalisations of this result to initial data sets which are neither conformally flat nor time symmetric are discussed.

[1] H. Friedrich. Gravitational fields near space-like and null infinity. *J. Geom. Phys.* 24, 83 (1998)

[2] J.A. Valiente Kroon. A new class of obstructions to the smoothness of null infinity. *Comm. Math. Phys.* 244, 133 (2004).

[3] J.A. Valiente Kroon. A rigidity property of the development of conformally flat data in General Relativity. In preparation.

BOUNDARY DIFFERENTIABILITY OF SOLUTIONS OF ELASTIC PLASTIC PROBLEMS WITH LINEAR HARDENING

JENS FREHSE, DOMINIQUE LÖBACH

Institut für Angewandte Mathematik, Universität Bonn

E-mail: loebach@iam.uni-bonn.de

We consider the quasistatic variational inequality of elastic plastic deformation with linear kinematic and isotropic hardening. The basic domain $\Omega \subset \mathbb{R}^n$ has $C^{1,1}$ -boundary. Near the Dirichlet boundary we obtain the full tangential derivatives of the stresses and hardening parameters σ, ξ in L^2 . For the normal derivatives we obtain Nishitskii Space differentiability of order $\frac{1}{2}$. Related results for the displacements u are presented.

MATHEMATICAL ANALYSIS OF UNSTEADY FLOWS OF FLUIDS WITH PRESSURE, SHEAR-RATE AND TEMPERATURE DEPENDENT MATERIAL MODULI

MIROSLAV BULÍČEK, JOSEF MÁLEK*, KUMBAKONAM R. RAJAGOPAL

**Faculty of Mathematics and Physics, Charles University Prague*

E-mail: malek@karlin.mff.cuni.cz

In his treatise titled “The physics of high pressures” (1931), Bridgman carefully documented that the viscosity and the thermal conductivity of most liquids depend on the pressure and the temperature. The relevant experimental studies show that even at high pressures the variations of the values in the density are insignificant in comparison to that of the viscosity, and it is thus reasonable to assume that the liquids in question are incompressible fluids with pressure dependent viscosities.

We rigorously investigate the mathematical properties (such as the existence and regularity) of unsteady three-dimensional internal flows of such incompressible fluids. The model is expressed through a system of partial differential equations representing the balance of mass, the balance of linear momentum, the balance of energy and the equation for the entropy production. Assuming that we have Navier’s slip at the impermeable boundary we establish the long-time existence of a (suitable) weak solution when the data are large.

GLOBAL PROPERTIES OF EXPANDING SPACETIMES WITH CYLINDRICAL GRAVITATIONAL WAVES

MAKOTO NARITA

Okinawa National College of Technology

E-mail: narita@okinawa-ct.ac.jp

The strong cosmic censorship conjecture (SCCC) is one of the most important and unsolved problems in classical mathematical relativity. Recently, the validity of SCCC has been shown under certain assumptions, e.g. spherical, homogeneous, T^3 -Gowdy symmetries. To extend these results, we will consider expanding spacetimes with unpolarized cylindrically symmetric gravitational waves, which are proposed by Gowdy and Edmonds. The metric is $g = e^{2(\eta-U)}(-dt^2 + dr^2) + e^{2U}(dx + A dy)^2 + e^{-2U}R^2 dy^2$, where $\eta = \eta(t, r)$, $U = U(t, r)$, $A = A(t, r)$ and $R = tr$. We prove a global in time existence theorem for the corresponding Einstein equations. Into the past direction, it is shown that the solutions to the Einstein equations become asymptotically velocity-dominated near singularities. The Fuchsian algorithm, which is developed by Kichenassamy and Rendall, is used. The square of Riemann curvature blows up in the case. Into the future direction, it is proved that the spacetimes are future geodesically complete. Thus, the validity of SCCC is shown.

ON POSITIVE SOLUTIONS AND LIOUVILLE THEOREMS FOR P-LAPLACIAN-TYPE EQUATIONS

YEHUDA PINCHOVER

Department of Mathematics, Technion – Israel Institute of Technology

E-mail: pincho@tx.technion.ac.il

Let Ω be a domain in $\mathbb{R}^d$, $d \geq 2$, $1 < p < \infty$, and $V \in L_{\text{loc}}^\infty(\Omega)$. In this talk we shall discuss a few aspects of relations between functional-analytic properties of the functional

$$Q(u) := \frac{1}{p} \int_{\Omega} (|\nabla u|^p + V|u|^p) \, dx \quad u \in C_0^\infty(\Omega),$$

and properties of positive solutions of the equation

$$-\Delta_p(u) + V|u|^{p-2}u = 0 \quad \text{in } \Omega,$$

where $\Delta_p(u) := \nabla \cdot (|\nabla u|^{p-2} \nabla u)$ is the celebrated p -Laplacian. As an application we shall present Liouville theorems for such equations.

STEADY COMPRESSIBLE NAVIER–STOKES–FOURIER SYSTEM

PIOTR B. MUCHA, MILAN POKORNÝ*

** Faculty of Mathematics and Physics, Charles University Prague*

E-mail: pokorny@karlin.mff.cuni.cz

We study steady flow of a compressible heat conducting fluid in a bounded domain $\Omega \subset \mathbb{R}^3$. We consider either the slip boundary condition or the homogeneous Dirichlet boundary condition for the velocity and so-called Newton's boundary condition for the temperature. For the pressure law $p(\rho, \theta) \sim \rho^\gamma + \theta\rho$ with $\gamma > 7/3$ we show that under reasonable technical assumptions on the data of the problem, there is a weak solution to the above mentioned system. Moreover, for the slip boundary condition and $\gamma > 3$ the solution is such that the density $\rho \in L^\infty(\Omega)$, the velocity $\mathbf{v} \in W^{1,q}(\Omega)$ and the temperature $\theta \in W^{1,q}(\Omega)$ for any $1 \leq q < \infty$.

ABSTRACT POLYMER GAS WITH ATTRACTIVE PAIR INTERACTIONS

ALDO PROCACCI

Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

E-mail: aldo@mat.ufmg.br

Two convergence criteria of cluster expansion for abstract polymer gases with general pair interaction (i.e. not necessarily repulsive and not necessarily hard-core) are presented. The first one it is a generalization of the Kotecký-Preiss criterion for stable pair interactions. The second one is a generalization of the new criterion given in [Fernandez and Procacci, Comm. Math. Phys. 274 (2007)].

DELTA-SHOCK WAVES AND MEDIUMS WHICH CAN BE TREATED AS PRESSURELESS CONTINUUMS

VLADIMIR M. SHELKOVICH

*Department of Mathematics, St. Petersburg State University
of Architecture and Civil Engineering*

E-mail: shelkv@yahoo.com

There are “nonclassical” situations where, in contrast to Lax’s and Glimm’s classical results, the Cauchy problem for a quasilinear system of conservation laws may admit delta-shock wave type solutions [1]. They are solutions such that their components may contain Dirac functions.

We study delta-shocks in the multidimensional system of conservation laws

$$\rho_t + \nabla \cdot (\rho F(U)) = 0, \quad (\rho U)_t + \nabla \cdot (\rho N(U)) = 0, \quad x \in R^n, \quad t \geq 0, \quad (1)$$

where F and N are given vector and tensor fields, respectively, $\rho = \rho(x, t) \in R$, $U(x, t) \in R^n$. The zero-pressure gas dynamics is a well-known particular case of (1).

We derive the balance laws describing mass, momentum, and energy transport between the volume outside of the delta-shock wave front and the delta-shock moving wave front. We prove that these processes are going on in such a way that the mass of the delta-shock wave front is an increasing quantity, while the energy of the volume (outside of the delta-shock wave front) and the total energy are nonincreasing quantities. The Cauchy problem related to propagation of delta-shock waves is solved.

The systems of the type (1) are appropriate for modeling and studying singular problems in mediums which can be treated as pressureless continuums (dusty gases, two-phase flows with solid particles or droplets, granular gases).

[1] V.M. Shelkovich, δ - and δ' -shock types of singular solutions to systems of conservation laws and the transport and concentration processes, *Uspekhi Mat. Nauk*, 63:3, 2008, 73-146. English transl. in *Russian Math. Surveys*, 63:3, 2008, 473-546.

P-ADIC PSEUDO-DIFFERENTIAL EQUATIONS AND WAVELETS

MARIA SKOPINA

Department of Applied Mathematics and Control Processes, St. Petersburg State University

E-mail: skopina@MS1167.spb.edu

p-Adic pseudo-differential equations are used in p-adic models instead of differential equations. Such equations arise in some applications, for example, in modeling interbasin kinetics of macromolecules and for the description of the genetic code.

A p-adic analog of the Haar basis and some other p-adic wavelet systems were recently found. It turned out that these wavelets have the following wonderful property: they are eigenfunctions of some p-adic pseudo-differential operators. Due to this property, a method for finding the exact solutions of p-adic pseudo-differential equations was developed in [1]. In particular, they solved the Cauchy problem for the Schrödinger-like semi-linear equations $\frac{\partial u(x, t)}{\partial t} + Au(x, t) + u(x, t)|u(x, t)|^{2m} = 0$, $u(x, 0) = u^0(x)$, where time t is a real, $x \in Q_p^n$, A is a pseudo-differential operator. Solutions of the above Cauchy problem is soliton-like and describes diffusion processes in Q_p^n . So, wavelets are useful for the problems of p-adic mathematical physics.

To develop a general scheme for the construction of p-adic wavelets we study p-adic multiresolution analyses (MRAs). It turned out that the MRA theory in p-adics very differs from real setting. Only 1-periodic test functions may be taken as orthogonal scaling functions. Moreover, all these functions generate the same Haar MRA. On the other hand, there are infinitely many different orthogonal bases generated by the p-adic Haar MRA. All these bases were described [2]. We also suggested a method for the construction of wavelet frames and Riesz bases generated by non-orthogonal scaling functions.

- [1] S. Albeverio, A.Yu. Khrennikov, V.M. Shelkovich, p -Adic semi-linear evolutionary pseudo-differential equations in the Lizorkin space, Dokl. RAN, 415, no. 3, 2007, 295-299.
- [2] V.M. Shelkovich, M. Skopina, p -Adic Haar multiresolution analysis and pseudo-differential operators, to appear in J. of Fourier Analysis and Applications.

POSTERS

BINDU ANUBHA BAMBAH (School of Physics, University of Hyderabad), with CHANDRA-SEKHAR MUKKU: **KNOTTED CONFIGURATIONS IN NON-ABELIAN MAGNETO-FLUIDS**

NIKOLAY BELYAKOV (Belfort, France): **SINGULAR INTEGRAL TRANSFORM FOR SEMI-SPACE WITH MULTI-LAYER COATING**

DENIS BORISOV (Bashkir State Pedagogical University, Ufa): **ON THE SPECTRUM OF TWO-DIMENSIONAL PERIODIC OPERATOR WITH A LOCALIZED PERTURBATION**

MIKHAIL BORSUK (University of Warmia and Mazury, Olsztyn): **THE ELLIPTIC TRANSMISSION PROBLEM IN NONSMOOTH DOMAINS**

MARIA SANTOS BRUZON (Department of Mathematics, University of Cadiz), with MARIA LUZ GANDARIAS: **TRAVELLING WAVE SOLUTIONS OF REACTION-DIFFUSION-CONVECTION EQUATIONS**

MIROSLAV BULÍČEK (Charles University, Prague), with PIOTR GWIAZDA, JOSEF MÁLEK, and AGNIESZKA SWIERCZEWSKA-GWIAZDA: **ON SCALAR HYPERBOLIC LAWS WITH DISCONTINUOUS FLUX**

FLORICA IOANNA DRAGOMIRESCU (Politehnica din Timisoara), with D.A. BISTRAM, S. MUNTEAN, and R.F. SUSAN-RESIGA: **ON A POLYNOMIALS BASED TAU METHOD IN A SWIRLING FLOW DOWNSTREAM A FRANCIS TURBINE RUNNER**

ALEXANDRU DUMITRACHE (IMSAM, Romanian Academy), with FLORIN FRUNZULICA and HORIA DUMITRESCU: **A NUMERICAL STUDY OF COMBUSTION PHENOMENON IN A CYLINDRICAL COMBUSTION CHAMBER**

FLORIN FRUNZULICA (Politehnica Buchares), with ALEXANDRU DUMITRACHE: **ADAPTIVE METHOD USING CONTROLLED GRID DEFORMATION**

MARIA LUZ GANDARIAS (Department of Mathematics, University of Cadiz), with MARIA SANTOS BRUZON: **NONCLASSICAL POTENTIAL SYMMETRIES FOR SOME HEAT EQUATIONS**

ÓSCAR J. GARAY (University of the Basque Country), with J. ARROYO and J.J. MENCIA: **CLOSED EXTREMA OF CURVATURE ENERGIES IN THE 2-SPHERE**

Q-HEUNG CHOI (Inha University, Korea), with TACKSUN JUNG: **PERIODIC SOLUTIONS FOR NONLINEAR PARABOLIC SYSTEMS**

VALERY IMAYKIN (Scientific Research Institute of Innovative Strategies in General Education Development), with ALEXANDER OMECH, HERBERT SPOHN, and BORIS VAINBERG: **SOLITON-TYPE ASYMPTOTICS FOR WAVE-PARTICLE SYSTEMS**

TACKSUN JUNG (Kunsan National University), with Q-HEUNG CHOI: **MULTIPLE SOLUTIONS FOR A CLASS OF THE SYSTEMS OF THE CRITICAL GROWTH SUSPENSION BRIDGE EQUATIONS**

- YILDIRAY KESKIN (Department of Mathematics, Selçuk University), with GALIP OTURANC, ONUR KARAOGLU, and SEMA SERVI: **REDUCED DIFFERENTIAL TRANSFORM METHOD FOR EXACT SOLUTIONS OF LAPLACE EQUATION**
- AYSE BETUL KOC (Department of Mathematics, Selçuk University), with A. KURNAZ: **SOLITON SOLUTIONS OF THE GENERALIZED HIROTO-SATSUMA COUPLED KDV SYSTEM BY THE REDUCED DIFFERENTIAL TRANSFORMATION**
- ANATOLY N. KOCHUBEI (Institute of Mathematics, Kyiv): **DISTRIBUTED ORDER CALCULUS AND EQUATIONS OF ULTRASLOW DIFFUSION**
- YACHENG LIU (Harbin Engineering University), with RUNZHANG XU: **FAMILY OF POTENTIAL WELLS AND ITS APPLICATIONS TO BOUSSINESQ EQUATIONS**
- MANUEL LUNA-LAYNEZ (University of Sevilla), with J. CASADO-DIAZ and F.J. SUÁREZ-GRAU: **ASYMPTOTIC BEHAVIOR OF A VISCOUS FLUID NEAR A ROUGH WALL WITH VERY SMALL ASPERITIES**
- FAUSTINO MAESTRE-CABALLERO (University of Sevilla): **DINAMIC OPTIMAL DESIGN PROBLEM FOR A GENERAL COST**
- NIKOLAI A. MAGNITSKIY (Institute for Systems Analysis, RAS), with N.M. EVSTINGEEV and S.V. SIDOROV: **ON NATURE OF TURBULENCE IN THREE DIMENSIONAL RAYLEIGH-BENARD CONVECTION**
- RAYISA MOISEYENKO (Donetsk Nat. University), with V.I. STOROZHEV: **ANTIPLANE WAVE SCATTERING ON TUNNEL CYLINDRICAL VOIDS WITH ELASTIC INCLUSIONS OF SECTOR CROSS-SECTION IN ISOTROPIC BODY**
- RAJFA MUSEMIĆ (University of Sarajevo, Faculty of Mechanical Engineering), with MIRJANA VUKOVIĆ: **A NUMERICAL TREATMENT OF AXISYMMETRIC DIFFUSION PROBLEM – DEVELOPMENT AND ADAPTATION**
- DIRK PAULY (University of Duisburg-Essen, University of Jyväskylä): **LOW FREQUENCY ASYMPTOTICS FOR MAXWELL EQUATIONS**
- OLEKSII PLIUKHIN (Institute of Mathematics, Kyiv): **NEW CONDITIONAL SYMMETRIES AND EXACT SOLUTIONS OF REACTION-DIFFUSION SYSTEMS WITH POWER DIFFUSIVITIES**
- ALEXANDER REZOUNENKO (Kharkov University, Dept of Mechanics and Mathematics): **INVESTIGATIONS OF PARTIAL DIFFERENTIAL EQUATIONS WITH DISCRETE AND DISTRIBUTED STATE-DEPENDENT DELAYS**
- M. MANUELA RODRIGUES (University of Aveiro, Dept of Mathematics), with E.M. ROCHA: **THE CONVERGENCE ANALYSIS OF THE DECOMPOSITION METHOD FOR THE (1+1)-PARABOLIC PROBLEM IN NONUNIFORM MEDIA**
- SEBASTIAN SZYBKA (Jagellonian University Krakow), with TADEUSZ CHMAJ: **FRACTAL THRESHOLD BEHAVIOR IN VACUUM GRAVITATIONAL COLLAPSE**
- MARCO ANTONIO TANECO-HERNÁNDEZ (Instituto de Física y Matemáticas, Universidad Michoacana), with ANATOLI E. MERZON: **NONLINEAR SCATTERING IN THE LAMB SYSTEM WITH ZERO-MASS**
- SVETLANA TOKAREVA (Belfort, France): **NUMERICAL SIMULATION OF FLUID FLOW WITH DISCONTINUOUS GALERKIN METHODS**
- RENATA TROYAN (Donetsk National University): **CHANGING PROPERTIES OF NORMAL WAVES DISPERSION SPECTRUMS BY USING WAVEGUIDE SECTION GEOMETRY**
- VLADIMIR V. VASILYEV (Bryansk State University): **ON DIFFERENT BOUNDARY VALUE PROBLEMS FOR PSEUDODIFFERENTIAL EQUATIONS IN NON-SMOOTH DOMAINS**

DMITRY VOROTNIKOV (Voronezh State University): **MATHEMATICAL STUDY OF EQUATIONS OF VISCOELASTIC DIFFUSION IN POLYMERS**

RUNZHANG XU (Harbin Engineering University), with WEINING WU: **VARIATIONAL METHOD TO EVOLUTION EQUATION WITH COMBINED POWER-TYPE NONLINEARITIES**

Probability Theory

(including random matrices and SLE)

Session organizers:

Yuval Peres
Wendelin Werner

GLAUBER DYNAMICS FOR SPIN SYSTEMS AT HIGH AND CRITICAL TEMPERATURES

EYAL LUBETZKY

*Theory Group of Microsoft Research,
Redmond, WA 98052, USA
E-mail: eyal@microsoft.com*

The spectral gap of the continuous-time heat-bath Glauber dynamics for the Ising model on the lattice is believed to exhibit the following behavior. For some critical-inverse temperature β_c , the mixing-time of the dynamics is logarithmic in the surface-area for $\beta < \beta_c$, polynomial for $\beta = \beta_c$ and exponential for $\beta > \beta_c$. Furthermore, for $\beta < \beta_c$ the mixing time is sharply concentrated and the dynamics exhibits the *cutoff phenomenon*, an abrupt convergence to equilibrium. We survey the recent progress in confirming this picture for various underlying geometries and spin system models.

1. Introduction

The *Ising Model* on a finite graph $G = (V, E)$ with inverse-temperature $\beta \geq 0$ and no external magnetic field is defined as follows. Its set of possible *configurations* is $\Omega = \{\pm 1\}^V$, where each $\sigma \in \Omega$ assigns positive or negative *spins* to the vertices of G . In the free boundary case, the probability that the system is at a given configuration σ is given by the *Gibbs distribution* $\mu_G(\sigma) = Z^{-1}(\beta) \exp\left(\beta \sum_{xy \in E} \sigma(x)\sigma(y)\right)$, where $Z(\beta)$ is the partition function. In the presence of a boundary condition $\tau \in \{\pm 1\}^{\partial V}$ (that fixes the spins of some subset $\partial V \subset V$ of the sites), we let $\mu_G^\tau(\sigma)$ denote the Gibbs measure conditioned on σ agreeing with τ on ∂V .

Heat-bath Glauber dynamics is one of the most practiced methods for sampling the Ising model: It is the Markov chain whose discrete-time version is given as follows. At each step a vertex is chosen uniformly at random and its spin is updated according to μ_G^τ conditioned on the spins of all the other vertices (this chain is reversible with respect to the Gibbs distribution μ_G^τ). The continuous-time version of the dynamics associates each site with an independent rate-1 Poisson clock, determining the update times of this site as above (by this definition the continuous dynamics is $|V|$ times faster than the discrete dynamics).

In the classical Ising model, the underlying geometry is $(\mathbb{Z}/n\mathbb{Z})^2$, the square lattice of side-length n , and there exists a critical inverse-temperature $\beta_c > 0$ at which the static Gibbs distribution exhibits a phase-transition with respect to long-range correlations between spins.

Supported by many experiments and studies in the theory of dynamical critical

phenomena, physicists believe that the spectral-gap of the continuous-time dynamics on lattices has the following *critical slowing down* behavior (e.g., [1–4]): At high temperatures ($\beta < \beta_c$) the inverse-gap is $O(1)$, at the critical β_c it is polynomial in the surface-area and at low temperatures it is exponential in it. This is known for $\mathbb{Z}^2$ except at the critical β_c , and establishing the order of the gap at criticality seems extremely challenging.

Furthermore, not only is $\text{gap}^{-1} = O(1)$ at high temperatures, but it is conjectured (cf. [5]) that the dynamics mixes abruptly in this regime and exhibits *cutoff*, an intriguing phenomenon discovered by Diaconis and by Aldous (see, e.g., [6]). Consider an infinite family of ergodic finite Markov chains (X_t) with an implicit parameter n , and let $t_{\text{MIX}}(\epsilon)$ be the minimum $t \geq 0$ where the heat-kernel H_t associated with (X_t) is within a total-variation distance of ϵ from stationarity. We say that (X_t) has cutoff iff for some sequence $w_n = o(t_{\text{MIX}}(1/e))$ we have the following: For any $0 < \epsilon < 1$ there exists some $c_\epsilon > 0$ such that $t_{\text{MIX}}(\epsilon) - t_{\text{MIX}}(1 - \epsilon) \leq c_\epsilon w_n$ for all n . That is, there is a sharp transition in the convergence of the given chains to equilibrium at time $(1 + o(1))t_{\text{MIX}}(1/e)$. In this case, the sequence w_n is called a *cutoff window* and the sequence $t_{\text{MIX}}(1/e)$ is called the *cutoff point*.

Though many families of chains are believed to exhibit cutoff, proving the occurrence of this phenomenon is often an extremely challenging task and there are relatively few examples for which cutoff has been rigorously shown. One of the first such examples was the Ising model at infinite temperature ($\beta = 0$) on a system of size n , where it was shown by Aldous [7] that cutoff occurs at $\frac{1}{2} \log n$ with a window of $O(1)$. On the square lattice $(\mathbb{Z}/n\mathbb{Z})^2$ it was unknown whether cutoff occurs at high temperatures. In fact, it was even unknown whether there is cutoff for the 1-dimensional case (see e.g. [8] for this longstanding open problem of Peres).

In what follows we describe some of the recent developments in establishing the above behavior of the Glauber dynamics in various underlying geometries. While the above picture is also conjectured to hold (and partly verified in various geometries) for other spin system models such as the Potts model, gas hard-core model etc., here we will mainly focus on the Ising model.

2. Glauber dynamics for the mean-field Ising model

The *Curie-Weiss* model corresponds to the case where the underlying geometry is the complete graph on n vertices. The study of this model (see, e.g., [5, 9–11]) is motivated by the fact that its behavior approximates that of the Ising model on high-dimensional tori. It is convenient in this case to re-scale the parameter β , so that the stationary measure μ_n satisfies $\mu_n(\sigma) \propto \exp\left(\frac{\beta}{n} \sum_{x < y} \sigma(x)\sigma(y)\right)$. It is well known that for any fixed $\beta > 1$, the Glauber dynamics (X_t) mixes in exponential time (cf., e.g., [12]), whereas for any fixed $\beta < 1$ the inverse-gap is $O(1)$ (see [13] and also [14]). Levin, Luczak and Peres [5] established that the mixing-time at the critical point $\beta = 1$ has order $\sqrt{n}$ and that for fixed $0 < \beta < 1$ there is cutoff at time $[2(1 - \beta)]^{-1} \log n$ with window $O(1)$.

The critical window in which the dynamics transitions from abrupt mixing with cutoff to a power-law dependence on n and finally to exponentially slow mixing was studied in [15]. It was shown there that the critical window has order $1/\sqrt{n}$ about which t_{MIX} interpolates smoothly between the subcritical and supercritical regimes:

Theorem 2.1 (cf. [15]). *Let $\delta = \delta(n) > 0$ be such that $\delta^2 n \rightarrow \infty$ with n . The Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 - \delta$ exhibits cutoff at time $(1/2\delta) \log(\delta^2 n)$ with window size $1/\delta$. In addition, the spectral gap of the dynamics in this regime is $(1+o(1))\delta$, where the $o(1)$ -term tends to 0 as $n \rightarrow \infty$.*

Theorem 2.2 (cf. [15]). *Let $\delta = \delta(n)$ satisfy $\delta = O(1/\sqrt{n})$. The mixing time of the Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 \pm \delta$ has order $\sqrt{n}$, and does not exhibit cutoff. In addition, the spectral gap of the dynamics in this regime has order $1/\sqrt{n}$.*

Theorem 2.3 (cf. [15]). *Let $\delta = \delta(n) > 0$ be such that $\delta^2 n \rightarrow \infty$ with n . The mixing-time of Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 + \delta$ does not exhibit cutoff, and has order $t_{\text{MIX}} = \frac{1}{\delta} \exp(\frac{n}{2} [\beta \zeta^2 - \zeta \log(\frac{1+\zeta}{1-\zeta}) - \log(1-\zeta^2)])$, where ζ is the unique positive solution of $\tanh(\beta x) = x$. In particular, in the special case $\delta \rightarrow 0$, the order of the mixing time is $(1/\delta) \exp((\frac{3}{4} + o(1))\delta^2 n)$, where the $o(1)$ -term tends to 0 as $n \rightarrow \infty$. In addition, the spectral gap of the dynamics in this regime has order $1/t_{\text{MIX}}$.*

The aforementioned picture for the behavior of the Glauber dynamics is therefore completely verified for this model.

3. Ising model on the Bethe Lattice

The important case of the Ising model on a regular tree, known as the Bethe lattice, has been intensively studied (e.g., [16–26]). On this canonical example of a non-amenable graph (one whose boundary is proportional to its volume) the model exhibits a rich behavior. For example, it has two distinct critical inverse-temperatures: one for uniqueness of the Gibbs state and another for the purity of the free-boundary state. The latter, β_c , coincides with the critical spin-glass parameter.

Previous results on the Ising model on a regular tree imply that the correct parameter to play the role of the surface-area is the tree-height h : It was shown in [16] and in [25] that the inverse-gap is $O(1)$ for $\beta < \beta_c$ and exponential in h for $\beta > \beta_c$, yet its critical behavior remained unknown until very recently: It was shown in [27] that the inverse-gap at criticality is indeed polynomial in h . Furthermore, this holds under any boundary condition, and an analogous result holds for t_{MIX} .

Theorem 3.1 (cf. [27]). *Fix $b \geq 2$ and let $\beta_c = \text{arctanh}(1/\sqrt{b})$ denote the critical inverse-temperature for the Ising model on the b -ary tree of height h . Then there exists some constant $c > 0$ independent of b , so that the following holds: For any boundary condition τ , the continuous-time Glauber dynamics for the above critical Ising model satisfies $\text{gap}^{-1} \leq t_{\text{MIX}} = O(h^c)$.*

The phase-transition of the inverse-gap of the dynamics on the Bethe lattice is thus completely verified (other than the complete graph, the regular tree is the only geometry where power-law mixing for the critical Ising model has been verified). It remains open to determine whether or not there is cutoff at high temperatures.

4. Recent progress for Ising model on lattices at high temperature

A series of breakthrough papers by Aizenman, Dobrushin, Holley, Shlosman, Stroock et al. (cf., e.g., [13, 28–41]) starting from the late 1970's has developed the theory of the convergence rate to stationarity of the Glauber dynamics for the Ising models on lattices. While many of these results involved the spectral-gap of the dynamics in the high temperature regime, the asymptotics of the L^1 mixing time in this regime remained unknown. Peres (cf. [5, 8]) conjectured that for any d the Glauber dynamics for the Ising model on $(\mathbb{Z}/n\mathbb{Z})^d$ should exhibit cutoff.

This was very recently confirmed in [42] for any dimension and any temperature such that there is strong spatial mixing: For $\mathbb{Z}^2$ this carries all the way to the critical temperature.

Theorem 4.1 (cf. [42]). *Let $\beta_c = \frac{1}{2} \log(1+\sqrt{2})$ be the critical inverse-temperature for the Ising model on $\mathbb{Z}^2$. Then the continuous-time Glauber dynamics for the Ising model on $(\mathbb{Z}/n\mathbb{Z})^2$ at inverse-temperature $0 \leq \beta < \beta_c$ with periodic boundary conditions has cutoff at $\lambda_\infty^{-1} \log n$ with a window of $O(\log \log n)$, where λ_∞ is the spectral gap of the dynamics on the infinite-volume lattice.*

Theorem 4.2 (cf. [42]). *Let $d \geq 1$ and consider the continuous-time Glauber dynamics for the ferromagnetic Ising model on $(\mathbb{Z}/n\mathbb{Z})^d$ with periodic boundary conditions, inverse-temperature β and external field h . Suppose that β, h are such that there is strong spatial mixing. Then the dynamics exhibits cutoff at $(d/2\lambda_\infty) \log n$ with a window of $O(\log \log n)$, where λ_∞ is the spectral gap of the dynamics on the infinite-volume lattice.*

The proofs of these results hinge on a new technique for translating L^1 -mixing to L^2 -mixing of projections of the chain, which enables the application of logarithmic-Sobolev inequalities. The technique is general and its applications for proving cutoff carry to other monotone and anti-monotone spin-systems, e.g. gas hard-core, Potts, anti-ferromagnetic Ising, arbitrary boundary conditions, etc.

Theorem 4.3 (cf. [42]). *The cutoff result given in Theorem 4.2 for the ferromagnetic Ising model also holds for the anti-ferromagnetic Ising model.*

Theorem 4.4 (cf. [42]). *Let $d \geq 1$. The following holds for the Glauber dynamics for the gas hard-core model on $(\mathbb{Z}/n\mathbb{Z})^d$ with fugacity β and periodic boundary conditions. If β is such that there is strong spatial mixing, then the dynamics exhibits cutoff at $(d/2\lambda_\infty) \log n$ with window of $O(\log \log n)$, where λ_∞ is the spectral gap of the dynamics on the infinite volume lattice.*

References

- [1] P. Hohenberg and B. Halperin, *Rev. Mod. Phys.* **49**, 435 (1977).
- [2] K. B. Lauritsen and H. C. Fogedby, *J. Statist. Phys.* **72**, 189 (1993).
- [3] F. Martinelli, Lectures on Glauber dynamics for discrete spin models, in *Lectures on probability theory and statistics*, (Springer, Berlin, 1999) pp. 93–191.
- [4] F.-G. Wang and C.-K. Hu, *Phys. Rev. E* **56**, 2310 (1997).
- [5] D. A. Levin, M. Luczak and Y. Peres, *Probab. Theory Related Fields* **146**, 223 (2008).
- [6] P. Diaconis, *Proc. Nat. Acad. Sci. U.S.A.* **93**, 1659 (1996).
- [7] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, in *Seminar on probability, XVII*, (Springer, Berlin, 1983) pp. 243–297.
- [8] D. Levin, Y. Peres and E. Wilmer, *Markov chains and mixing times* (American Mathematical Society, 2008).
- [9] R. S. Ellis, *Entropy, large deviations, and statistical mechanics* (Springer-Verlag, Berlin, 2006). Reprint of the 1985 original.
- [10] R. S. Ellis and C. M. Newman, *Z. Wahrsch. Verw. Gebiete* **44**, 117 (1978).
- [11] R. S. Ellis, C. M. Newman and J. S. Rosen, *Z. Wahrsch. Verw. Gebiete* **51**, 153 (1980).
- [12] R. B. Griffiths, C.-Y. Weng and J. S. Langer, *Phys. Rev.* **149**, 301 (1966).
- [13] M. Aizenman and R. Holley, Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime, in *Percolation theory and ergodic theory of infinite particle systems*, (Springer, New York, 1987) pp. 1–11.
- [14] R. Bubley and M. Dyer, *Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS 1997)*, 223.
- [15] J. Ding, E. Lubetzky and Y. Peres, *Comm. Math. Phys.* **289**, 725 (2009).
- [16] N. Berger, C. Kenyon, E. Mossel and Y. Peres, *Probab. Theory Related Fields* **131**, 311 (2005).
- [17] P. M. Bleher, J. Ruiz and V. A. Zagrebnov, *J. Statist. Phys.* **79**, 473 (1995).
- [18] J. M. Carlson, J. T. Chayes, L. Chayes, J. P. Sethna and D. J. Thouless, *J. Statist. Phys.* **61**, 987 (1990).
- [19] J. M. Carlson, J. T. Chayes, L. Chayes, J. P. Sethna and D. J. Thouless, *Europhys. Lett.* **106**, 355 (1988).
- [20] J. M. Carlson, J. T. Chayes, J. P. Sethna and D. J. Thouless, *J. Statist. Phys.* **61**, 1069 (1990).
- [21] W. Evans, C. Kenyon, Y. Peres and L. J. Schulman, *Ann. Appl. Probab.* **10**, 410 (2000).
- [22] D. Ioffe, Extremality of the disordered state for the Ising model on general trees, in *Trees (Versailles, 1995)*, (Birkhäuser, Basel, 1996) pp. 3–14.
- [23] D. Ioffe, *Lett. Math. Phys.* **37**, 137 (1996).
- [24] R. Lyons, *Comm. Math. Phys.* **125**, 337 (1989).
- [25] F. Martinelli, A. Sinclair and D. Weitz, *Comm. Math. Phys.* **250**, 301 (2004).
- [26] R. Pemantle and Y. Peres, *Ann. Probab.* (to appear).
- [27] J. Ding, E. Lubetzky and Y. Peres, *Comm. Math. Phys.* (to appear).
- [28] R. L. Dobrushin and S. B. Shlosman, *J. Statist. Phys.* **46**, 983 (1987).
- [29] R. Holley, On the asymptotics of the spin-spin autocorrelation function in stochastic Ising models near the critical temperature, in *Spatial stochastic processes*, (Birkhäuser Boston, Boston, MA, 1991) pp. 89–104.
- [30] R. A. Holley and D. W. Stroock, *J. Statist. Phys.* **46**, 1159 (1987).
- [31] R. A. Holley and D. W. Stroock, *Comm. Math. Phys.* **123**, 85 (1989).
- [32] T. M. Liggett, *Interacting particle systems* (Springer-Verlag, Berlin, 2005). Reprint of the 1985 original.

- [33] S. L. Lu and H.-T. Yau, *Comm. Math. Phys.* **156**, 399 (1993).
- [34] F. Martinelli and E. Olivieri, *Comm. Math. Phys.* **161**, 447 (1994).
- [35] F. Martinelli and E. Olivieri, *Comm. Math. Phys.* **161**, 487 (1994).
- [36] F. Martinelli, E. Olivieri and R. H. Schonmann, *Comm. Math. Phys.* **165**, 33 (1994).
- [37] D. W. Stroock and B. Zegarliński, *Comm. Math. Phys.* **144**, 303 (1992).
- [38] D. W. Stroock and B. Zegarliński, *J. Funct. Anal.* **104**, 299 (1992).
- [39] D. W. Stroock and B. Zegarliński, *Comm. Math. Phys.* **149**, 175 (1992).
- [40] B. Zegarliński, *J. Funct. Anal.* **105**, 77 (1992).
- [41] B. Zegarliński, *Lett. Math. Phys.* **20**, 173 (1990).
- [42] E. Lubetzky and A. Sly (preprint).

ON THE SCALING LIMIT OF RANDOM PLANAR MAPS WITH LARGE FACES

JEAN-FRANCOIS LE GALL and GRÉGORIE MIERMONT

*Laboratoire de Mathématiques, Université de Paris-Sud,
Bât. 425, 91405 Orsay, France*

*E-mail: {Jean-Francois.Legall, Gregory.Miermont}@math.u-psud.fr
www.math.u-psud.fr*

We discuss asymptotics for large random planar maps under the assumption that the distribution of the degree of a typical face is in the domain of attraction of a stable distribution with index $\alpha \in (1, 2)$. We view the vertex-set of a map as a metric space by endowing it with the usual graph distance. When the number n of vertices of the map tends to infinity, this metric space, rescaled by the factor $n^{-1/2\alpha}$, converges in distribution as $n \rightarrow \infty$, at least along suitable subsequences, towards a limiting random compact metric space whose Hausdorff dimension is equal to 2α .

This is a short presentation of the article [1], to which the interested reader is referred for more details.

Keywords: Random planar maps, random trees, scaling limits, stable tree

1. Introduction

A planar map is a proper embedding of a finite connected graph in the two-dimensional sphere. For technical reasons, it is convenient to deal with rooted planar maps, meaning that there is a distinguished oriented edge called the root edge. One is interested in the “shape” of the graph and not in the particular embedding that is considered: More rigorously, two rooted planar maps are identified if they correspond via an orientation-preserving homeomorphism of the sphere. The faces of the map are the connected components of the complement of edges, and the degree of a face counts the number of edges that are incident to it. Large random planar graphs are of interest in particular in theoretical physics, where they serve as models of random geometry [2].

A simple way to generate a large random planar map is to choose it uniformly at random from the set of all rooted p -angulations with n faces (a planar map is a p -angulation if all faces have degree p). It is conjectured that the scaling limit of uniformly distributed p -angulations with n faces, when n tends to infinity (or equivalently when the number of vertices tends to infinity), does not depend on the choice of p and is given by the so-called Brownian map. Since the pioneering work of Chassaing and Schaeffer [3], there has been several results supporting this conjecture. Marckert and Mokkadem [4] introduced the Brownian map and proved

a weak form of the convergence of rescaled uniform quadrangulations towards the Brownian map. A stronger version, involving convergence of the associated metric spaces in the sense of the Gromov-Hausdorff distance, was derived in Le Gall [5] in the case of uniformly distributed $2p$ -angulations. Because the distribution of the Brownian map has not been fully characterized, the convergence results of [5] require extracting suitable subsequences. Proving the uniqueness of the distribution of the Brownian map is one of the key open problems in the area.

2. Boltzmann distributions and maps with large faces

A more general way of choosing a large planar map at random is to use Boltzmann distributions. Here, we restrict our attention to bipartite maps, where all face degrees are even. Given a sequence $q = (q_1, q_2, q_3, \dots)$ of nonnegative real numbers and a bipartite planar map $\mathbf{m}$, the associated Boltzmann weight is

$$W_q(\mathbf{m}) = \prod_{f \in F(\mathbf{m})} q_{\deg(f)/2} \quad (1)$$

where $F(\mathbf{m})$ denotes the set of all faces of $\mathbf{m}$, and $\deg(f)$ is the degree of the face f . One can then generate a large planar map by choosing it at random in the set of all planar maps with n vertices (or with n faces) with probability weights that are proportional to $W_q(\mathbf{m})$.

Such distributions arise naturally (possibly in slightly different forms) in problems involving statistical physics models on random maps, like the $O(N)$ loop model, percolation, Ising or Potts models. This gives an important motivation for the present work. It is explained in [1, Section 8] how one obtains random maps with Boltzmann distributions, similar to the ones considered here, by considering the so-called *external gaskets* associated with each element of the configuration space. It is to be expected that the scaling limits discussed in Section 3 below appear as limits of such external gasket configuration in critical models.

Assuming certain integrability conditions on the sequence of weights, Marckert and Miermont [6] obtain a variety of limit theorems for large random bipartite planar maps chosen according to these Boltzmann distributions. These results are extended in Miermont [7] and Miermont and Weill [8] to the non-bipartite case, including large triangulations. In all these papers, limiting distributions are described in terms of the Brownian map. Therefore these results strongly suggest that the Brownian map should be the universal limit of large random planar maps, under the condition that the distribution of the degrees of faces satisfies some integrability property.

Here, we consider Boltzmann distributions such that, even for large n , a random planar map with n vertices will have “macroscopic” faces, which in some sense will remain present in the scaling limit. This leads to a (conjectured) scaling limit which is different from the Brownian map. In fact our limit theorems involve new random processes that are closely related to the stable trees of [9], in contrast to the construction of the Brownian map [4, 5], which is based on Aldous’ CRT.

For technical reasons, we consider planar maps that are both rooted and pointed: In addition to the root edge, there is a distinguished vertex denoted by v_* . Under appropriate conditions on the sequence of weights q , formula (1) defines a finite measure W_q on the set of all rooted and pointed planar maps. We let $\mathbf{P}_q$ be the probability measure obtained by normalizing W_q . When the law of the number $\#V(\mathbf{m})$ of vertices under $\mathbf{P}_q$ has a heavy tail (as opposed to an exponential tail), we say that q is critical (a precise definition can be found in [1, Section 2]). Our main assumption is as follows.

Assumption 2.1. *The critical weights q in (1) are chosen in such a way that under $\mathbf{P}_q$, the distribution μ of (half) the degree of a typical face (namely, the face located to the left of the root edge) is in the domain of attraction of a stable distribution with index α , for some $\alpha \in (1, 2)$, i.e. $\mu(k)$ behaves like $k^{-\alpha-1}$ when $k \rightarrow \infty$.*

3. Scaling limits

We equip the vertex set $V(\mathbf{m})$ of a planar map $\mathbf{m}$ with the graph distance d_{gr} , and we would like to investigate the properties of this metric space when $\mathbf{m}$ is distributed according to $\mathbf{P}_q$ and conditioned to be large. For every integer $n \geq 1$, denote by M_n a random planar map distributed according to $\mathbf{P}_q(\cdot \mid \#V(\mathbf{m}) = n)$. Our goal is to get information about typical distances in the metric space $(V(M_n), d_{\text{gr}})$ when n is large, and if at all possible to prove that these (suitably rescaled) metric spaces converge in distribution as $n \rightarrow \infty$ in the sense of the Gromov-Hausdorff distance. Our main result in this vein is the following.

Theorem 3.1. *Under Assumption 2.1, for every $n \geq 1$, let M_n be distributed according to $\mathbf{P}_q(\cdot \mid \#V(\mathbf{m}) = n)$, respectively according to $\mathbf{P}_q(\cdot \mid \#V(\mathbf{m}) \geq n)$. From every strictly increasing sequence of integers, one can extract a subsequence along which*

$$(V(M_n), n^{-1/2\alpha} d_{\text{gr}}) \xrightarrow[n \rightarrow \infty]{(d)} (\mathbf{M}_\infty, \delta_\infty)$$

where $(\mathbf{M}_\infty, \delta_\infty)$ is a random compact metric space and the convergence holds in distribution in the Gromov-Hausdorff sense. Furthermore, the Hausdorff dimension of $(\mathbf{M}_\infty, \delta_\infty)$ is a.s. equal to 2α .

The Hausdorff dimension 2α should be compared with the value 4 for the dimension of the Brownian map [5]. The fact that the Hausdorff dimension is bounded above by 2α follows from Hölder continuity properties of the distance process, which is briefly discussed at the end of Section 4 below. The proof of the corresponding lower bound is more involved and depends on some properties of the stable tree and its coding by Lévy processes, which are investigated in [9]. Similarly as in the case of the convergence to the Brownian map, the extraction of a subsequence in Theorem 3.1 is needed because the limiting distribution is not characterized. Other results on the limiting space are obtained in [1], like the convergence in distribution of the so-called profile of distances seen from v_* .

4. Methods

Our key tool is the Bouttier-Di Francesco-Guitter bijection [10] between bipartite planar maps and certain labeled trees called mobiles. A mobile is a (rooted) plane tree, where vertices at even distance, respectively at odd distance, from the root are called white, resp. black, and white vertices are assigned integer labels that satisfy certain simple rules. See Figure 1 for an example and basic properties.

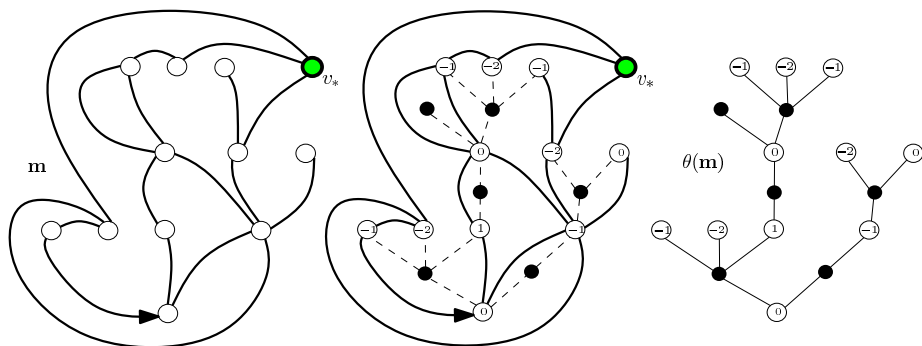


Fig. 1. In the Bouttier-Di Francesco-Guitter bijection, a (rooted and pointed) planar map $\mathbf{m}$ corresponds to a mobile $\theta(\mathbf{m})$ in such a way that each face of $\mathbf{m}$ is associated with a black vertex of $\theta(\mathbf{m})$ and each vertex of $\mathbf{m}$ (with the exception of the distinguished vertex v_*) is associated with a white vertex of $\theta(\mathbf{m})$. Moreover, the degree of a face of $\mathbf{m}$ is exactly twice the degree of the associated black vertex in the mobile $\theta(\mathbf{m})$.

We use a nice feature of the Bouttier-Di Francesco-Guitter bijection: Up to an additive constant depending on $\mathbf{m}$, the distance between v_* and an arbitrary vertex $v \in V(\mathbf{m}) \setminus \{v_*\}$ coincides with the label of the white vertex of $\theta(\mathbf{m})$ associated with v . Thus, in order to understand the asymptotic behavior of distances from v_* in the map M_n , it suffices to get information about labels in the mobile $\theta(M_n)$ when n is large. To this end, we first consider the tree $\mathcal{T}(M_n)$ obtained by ignoring the labels in $\theta(M_n)$. Under our basic assumption, the results of [9] can be applied to prove that the tree $\mathcal{T}(M_n)$ converges in distribution, modulo a rescaling of distances by the factor $n^{-(1-1/\alpha)}$, towards the so-called stable tree with index α . The stable tree can be defined by a suitable coding from the sample path of a centered stable Lévy process with no negative jumps and index α , under an appropriate excursion measure. The preceding convergence to the stable tree is however not sufficient for our purposes, since we are primarily interested in labels. Note that, under the assumptions made in [6] on the weight sequence q (and in particular in the case of uniformly distributed $2p$ -angulations), the rescaled trees $\mathcal{T}(M_n)$ converge towards the CRT, and the scaling limit of labels is described in [6] as Brownian motion indexed by the CRT, or equivalently as the Brownian snake driven by a normalized Brownian excursion. In our “heavy tail” setting however, the scaling limit of the labels is *not* Brownian motion indexed by the stable tree, but is given by a new random process

of independent interest, which we call the continuous distance process.

Let us give an informal presentation of the distance process. We view the stable tree as the genealogical tree for a continuous population, and the distance of a vertex from the root is interpreted as its generation in the tree. Fix a vertex a in the stable tree. Among the ancestors of a , countably many of them, denoted by $b_1, b_2, \dots$ correspond to a sudden creation of mass in the population: Each b_k has a macroscopic number $\delta_k > 0$ of “children”, and one can also consider the quantity $r_k \in [0, \delta_k]$, which is the rank among these children of the one that is an ancestor of a . The preceding description is informal in our continuous setting (there are no children) but can be made rigorous thanks to the ideas developed in [9] and in particular to the coding of the stable tree by a Lévy process. We then associate with each vertex b_k a Brownian bridge $(B_k(t))_{t \in [0, \delta_k]}$ (starting and ending at 0) with duration δ_k , independently when k varies, and we set

$$D(a) = \sum_{k=1}^{\infty} B_k(r_k). \quad (2)$$

The resulting process $D(a)$ when a varies in the stable tree is the continuous distance process. As a matter of fact, since vertices of the stable trees are parametrized by the interval $[0, 1]$ (using the coding by a Lévy process), it is more convenient to define the continuous distance process as a process $(D_t)_{t \in [0, 1]}$ indexed by the interval $[0, 1]$.

Much of the technical work contained in [1] is devoted to proving that the rescaled labels in the mobile $\theta(M_n)$ converge in distribution to the continuous distance process. The proper rescaling of labels involves the multiplicative factor $n^{-1/2\alpha}$ instead of $n^{-1/4}$ in earlier work. This indicates that the typical diameter of our random planar maps M_n is of order $n^{1/2\alpha}$ rather than $n^{1/4}$ in the case of maps with faces of bounded degree. This also gives a good intuition on the fact that the scaling limit arising in Theorem 3.1 has Hausdorff dimension 2α .

References

- [1] J.-F. Le Gall and G. Miermont, arXiv:0907.3262 (2009).
- [2] J. Ambjørn, B. Durhuus and T. Jonsson, *Quantum geometry. A statistical field theory approach*. Cambridge Monographs on Mathematical Physics, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 1997).
- [3] P. Chassaing and G. Schaeffer, *Probab. Theory Related Fields* **128**, 161 (2004).
- [4] J.-F. Marckert and A. Mokkadem, *Ann. Probab.* **34**, 2144 (2006).
- [5] J.-F. Le Gall, *Invent. Math.* **169**, 621 (2007).
- [6] J.-F. Marckert and G. Miermont, *Ann. Probab.* **35**, 1642 (2007).
- [7] G. Miermont, An invariance principle for random planar maps, in *Fourth Colloquium on Mathematics and Computer Sciences CMCS'06*, Discrete Math. Theor. Comput. Sci. Proc., AG (Nancy, 2006) pp. 39–58 (electronic).
- [8] G. Miermont and M. Weill, *Electron. J. Probab.* **13**, no. 4, 79 (2008).
- [9] T. Duquesne and J.-F. Le Gall, *Astérisque* **281**, vi+147 (2002).
- [10] J. Bouttier, P. Di Francesco and E. Guitter, *Electron. J. Combin.* **11**, Research Paper 69, 27 pp. (electronic) (2004).

THE SCALING LIMIT OF THE MINIMAL SPANNING TREE — A PRELIMINARY REPORT

GÁBOR PETE

*Department of Mathematics, University of Toronto, Canada
<http://www.math.toronto.edu/~gabor>*

joint work with CHRISTOPHE GARBAN* and ODED SCHRAMM**

**ENS Lyon*

***Microsoft Research (died on Sept. 1, 2008)*

This is a brief description of how the recent proof of the existence and conformal covariance of the scaling limits of dynamical and near-critical planar percolation implies the existence and several topological properties of the scaling limit of the Minimal Spanning Tree, and that it is invariant under scalings, rotations and translations. However, we do not expect conformal invariance: we explain why not and what is missing for a proof.

1. Introduction

Critical planar percolation, both the discrete process and the continuum scaling limit, have become central objects of probability theory and statistical mechanics; see [1] for the classical results and [2] for a great course on our present knowledge, using conformal invariance [3] and SLE [4]. In the past few years, there has also been a lot of progress on dynamical percolation, which is not only the natural time evolution with critical percolation as the stationary measure, and the natural framework to study how noise effects the system and how it produces exceptional events, but also provides tools to understand the near-critical regime and related objects like Invasion Percolation and the Minimal Spanning Tree (MST). See the survey [5], which discusses not only the recent [6], but also some work in preparation [7, 8]. In this note, we explain the applications to the MST, partly to encourage others to work on the main remaining open problem: conformal non-invariance of the MST scaling limit, which is certainly interesting given the translational, rotational and scale invariance that we can now prove.

We thank David Wilson for conversations and one of the pictures. The work of GP was supported by an NSERC Discovery Grant.

2. Dynamical and near-critical percolation

In a series of papers [7], we show that, in dynamical percolation, if each site of the triangular grid with mesh η has a Poisson clock with rate $r(\eta) = \eta^2 \alpha_4(\eta, 1) =$

$\eta^{3/4+o(1)}$ switching between black and white, then the $\eta \downarrow 0$ limit of this system exists as a Markov process. If the clocks always switch from white to black when they ring, then started from critical percolation at time 0, at time t we have near-critical percolation with density roughly $1/2 + tr(\eta)$ for black, for $t \in (-\infty, \infty)$. We show that this near-critical ensemble also has a scaling limit (called NCSL). We should briefly mention here what the topological space is where these limit processes live: following the description of [9] for the scaling limit of static critical percolation, this is a compact metrizable space that encodes all macroscopic crossing events of “quads” (conformal rectangles), i.e., an element of the space tells which quads are crossed by the percolation configuration. Then, partly following the suggestion of [10], we build the two limits from critical percolation, in two main steps:

(1) The **normalized counting measure** on the sites that are pivotal for the left-right crossing of any given quad Q converge to a finite measure μ^Q that is measurable in the scaling limit of critical percolation. We also show that this measure is **conformally covariant**: if the domain is changed by $\phi(z)$, then we get $\mu^{\phi(Q)}$ from μ^Q by scaling locally by $|\phi'(z)|^{3/4}$: there are more pivotals for a larger domain. Finally, we show that the collection of these pivotal measures μ^Q can be used to understand the “importance measures” μ^ϵ : the amount of sites that have the alternating 4-arm event to macroscopic distance at least ϵ .

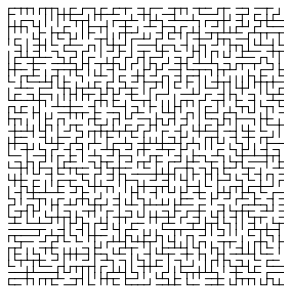
(2) **Stability**: Fix a quad Q , and let the set of sites switched in $[0, t]$ be W_t . Then the probability that a configuration ω_η can be changed on W_t into $\omega'_\eta, \omega''_\eta$ such that they agree on any site that is at least ϵ -important in ω_η , but Q is crossed by ω'_η while not crossed by ω''_η , is small if ϵ is small, uniformly in the mesh η . Note that this is a strengthening of Kesten’s theorem [11] that the 4-arm probabilities remain comparable in the entire near-critical regime.

By (1), using only macroscopic information, we can tell for all $\epsilon > 0$ how many microscopic ϵ -important sites there are in any region, and hence we know the rate with which important switches start happening when we start the dynamics. Then part (2) says that by following the switches of all these initially ϵ -important sites we can predict well (as $\epsilon \downarrow 0$) the state of any quad crossing event at any later time. Hence we get a well-defined Markov process in the scaling limit, both in the dynamical and the near-critical cases. We also get that these scaling limits are **conformally covariant**: time is scaled locally by $|\phi'(z)|^{3/4}$.

These results have (or may have) several applications. Similarly to the measurable measure on pivotals, we can construct the limit of the length measure on a percolation interface, giving the first **physical time-parametrization** for the SLE_6 curve (as opposed to the conformal capacity parametrization). We hope to describe **near-critical interfaces** with a **massive SLE_6** , involving a self-interacting drift term. The rotational invariance of the NCSL seems to help prove the asymptotic circularity of the percolation **Wulff crystal**, as $p \downarrow p_c$. We are also planning to study the dynamical and near-critical **FK-Ising models**. However, there are limitations to our methods: we do not have any **near-critical Cardy’s formula**, and do not have a guess for the **dimension** of Minimal Spanning Tree paths.

3. The Minimal Spanning Tree

For each edge of a finite graph, say $e \in E(\mathbb{Z}_n^2)$, let $U(e)$ be an i.i.d. $\text{Unif}[0, 1]$ label. The **Minimal Spanning Tree** is the spanning tree T for which $\sum_{e \in T} U(e)$ is minimal. This is well-known to be the same as the collection of lowest level paths between all pairs of vertices (i.e., the path between the two points for which the maximum label on the path is minimal). Or, delete from each cycle the edge with the highest label U . This also shows that T depends only on the ordering of the labels, not the values themselves.

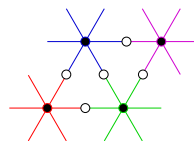


We can also use $\text{Unif}[0, 1]$ labels to get a coupling of percolation for all densities p , and use the same labels to get the MST. This way we get a **coupling** between the MST and the **percolation ensemble**. Moreover, the macroscopic structure of the MST is basically determined by the labels in the near-critical regime, as follows.

Consider the λ -clusters in NCE, for $\lambda \in (-\infty, \infty)$, i.e., the connected components given by the labels at most $1/2 + \lambda r(\eta)$. Contract each component into a single vertex, resulting in the “cluster graph”. It is easy to verify that making these contractions on the MST we get exactly the MST on the cluster graph. We denote this **cluster tree** by T_λ . Since the largest λ -clusters for $\lambda \ll 0$ are of very small macroscopic size, the tree T_λ will tell us the macroscopic structure of the MST. On the other hand, for $\lambda \gg 0$, most sites are in a few large λ -clusters, with only few T_λ edges between them. For $\lambda_1 < \lambda_2$, we get the tree T_{λ_2} from T_{λ_1} by contracting the edges with labels in $(\lambda_1, \lambda_2]$. Thus, if we have the collection of λ -clusters for all $\lambda \in (-\infty, \infty)$, then we can reconstruct the trees T_λ , and hence the macroscopic structure of the NCE seems to determine that of the MST. However, this is only an intuitive description: we ignored that there are a lot of small λ -clusters for any λ , hence it is not at all clear that the NCESL still determines an object that can be the scaling limit of the MST. We will see in the next section how one can build an actual proof.



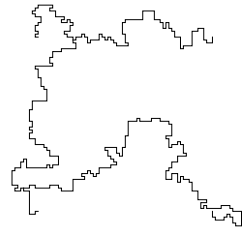
Since we have a proof of the existence and properties of the NCESL only for site percolation on the triangular lattice Δ , if we want to use this to build the MST scaling limit, we will need a version of the MST that uses $\text{Unif}[0, 1]$ vertex labels $\{V(x)\}$ on Δ . So, replace each edge of Δ by two in series, and for each new edge e , denote its endpoint that was originally a vertex of Δ by e^* . Then, let $U(e) := V(e^*)$. The MST using these edge labels $\{U(e)\}$ will inherit the right connectivity properties from the percolation ensemble. Our strongest results will apply to this model, but some of them will also hold for subsequential limits of the usual MST on $\mathbb{Z}^2$, known to exist by [12].



We note that the MST is also the union of the invasion trees of **Invasion Percolation**, see [12–14] and the references there.

4. How do we see the MST in the scaling limit?

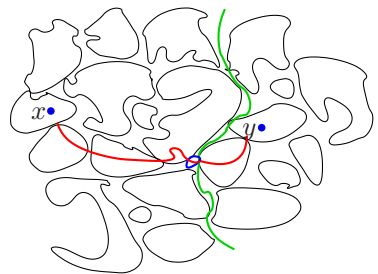
The MST scaling limit (MSTSL) should be a random spanning tree on all the points of the real plane. This, of course, cannot be a random subset of the plane, even though the discrete MST is a subgraph of the lattice. Rather, for any pair of points, x, y , we want the MSTSL path $P(x, y)$ connecting them. (And this collection of paths should satisfy some compatibility relations: the symmetric difference of $P(x, y)$ and $P(x, z)$ should be $P(y, z)$, and more generally, the subtree of the MSTSL connecting a finite set of given points should be part of the subtree connecting a larger finite set.) In fact, we fix a countable dense set of points Z in the plane, and the MSTSL will be a collection of paths connecting all pairs of Z . This way, almost sure results for the path between a fixed pair will hold for all the pairs in Z simultaneously. Then, for arbitrary points x, y , we can take sequences $x_n \rightarrow x$ and $y_n \rightarrow y$ with $x_n, y_n \in Z$, and take the limit of the paths $P(x_n, y_n)$ (with the metric given by the infimum of L^∞ -distances over all possible parameterizations of the paths). There will be pairs x, y for which there are at least two different limiting paths; in this sense, the MSTSL is not exactly a tree. Nevertheless, this seems to be the best possible notion of the MST scaling limit. See [12] for the precise description and basic topological properties of the subsequential scaling limits of the MST.



We now show that the MST path joining $x, y \in \mathbb{R}^2$ is determined by the NCE in such a way that the MSTSL will be a measurable function of the NCESL.

Fix some $\lambda_1 \ll 0$, so that even the outermost λ_1 -clusters are small. As explained in Section 3, it is enough to find the path between the outermost clusters of x, y in the cluster tree T_{λ_1} in the scaling limit, then $\lambda_1 \downarrow -\infty$ will give the MSTSL.

The basic difficulty now is that there are infinitely many outermost λ_1 -clusters even in a bounded region, so the MST on this cluster-graph may not be well-defined. So, take only the outermost λ_1 -clusters with diameter $\geq \epsilon$.

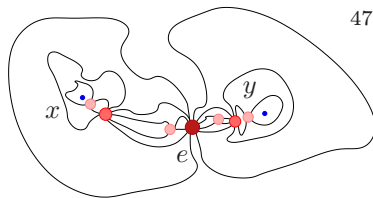


Using RSW arguments, one can show that if ϵ is small enough, then there is a finite path between x and y in this cluster-graph. The resulting MST path uses labels $\leq \lambda_2$ only, since we need to open only finitely many cut-edges.

Now take $\delta \ll \epsilon$, and the corresponding new cluster graph. We claim that if ϵ was small enough, then the MST path in the new δ -cutoff cluster graph is the same.

Since the old path is still available in the δ -cutoff cluster graph, the new path also has labels $\leq \lambda_2$. If the path goes through an outermost λ_1 -cluster that has diameter δ , then there is a λ_1 -cut-edge e on the path that is only δ -important. But then, moving all the cut-edge labels from at most λ_2 to λ_1 , this e becomes very important at level λ_1 , although it was very little important before. This contradicts the stability result (2) of Section 2. Therefore, the T_{λ_1} path does not go through very small λ_1 -clusters, so it is visible also in the scaling limit, and we are done.

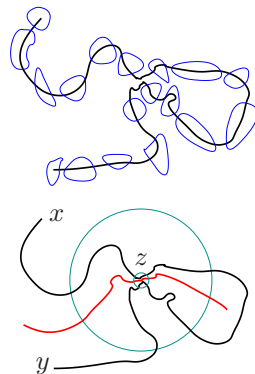
Alternatively, one may try to find the minimal level where x and y become connected, at the cut-edge $e_{x,y}$, then repeat this between $e_{x,y}$ and x and y dyadically, until the entire path is recovered. However, when the mesh $\eta \downarrow 0$, the labels on the path near x and y blow up, hence $e_{x,y}$ in the scaling limit “simultaneously coincides” with both x and y , so the procedure does not make sense. Hence we again need some macroscopic cut-off $\epsilon \downarrow 0$, for which proving the convergence seems harder than above.



5. Topology of the MST scaling limit

In [12] it was proved for any subsequential limit that almost all vertices are leafs and that there is a uniform bound on the vertex degrees. We can now prove stronger results: *In either lattice, there are no degree ≥ 5 points. For any two points in the plane, the MSTSL path joining them is a.s. unique and simple (not even a figure of 6).* However, we do not know, for instance, if there are degree 4 points.

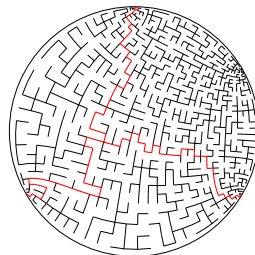
Here is a proof sketch of the simple path claim. Consider a generic nearly non-simple path between x and y . Take λ_1 very negative, and consider the part of $P(x, y)$ between the outermost λ_1 -clusters surrounding x and y . This part will be entirely below some finite level λ_2 . On the other hand, since the λ_1 -clusters all have small diameters, on the path there are labels above λ_1 “all over the place”. But this implies that there must exist two macroscopic dual arms with labels all above λ_1 that force $P(x, y)$ go around the almost touch-point z . Altogether, we have 4 primal arms below level λ_2 , and 2 dual arms with labels above λ_1 , i.e., a six arm event around z within a W_λ -modification, as in Claim (2) of Section 2. Since it is known even on $\mathbb{Z}^2$ that the six-arm event does not happen at criticality, touch-points are ruled out.



6. Conformal non-invariance?

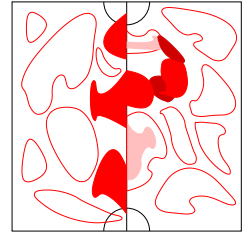
Scaling and rotational invariance of MSTSL follows easily from the same properties of NCSL, but the conformal covariance of the latter suggests **conformal non-invariance of the MSTSL**. This is also supported by careful simulations [15]: the law of the trifurcation point on the first picture (a conformal image of a discrete square) is not invariant under rotation by $2\pi/3$. Here is a simplified version of what should be proved, leaving as an exercise to figure out the exact connection to the problem:

Take an $n \times n$ square, with $\text{Unif}([0, 1/5] \cup [4/5, 1])$ labels on the left half, and $\text{Unif}[2/5, 3/5]$ labels on right. Take the MST path between the endpoints x, y of the vertical half-line, and consider its segment between the ϵn -neighbourhoods of x and



y (with high probability, there is only one). Does this segment feel the asymmetry between the two halves, as $n \rightarrow \infty$? It would be enough to prove, e.g., that the probability that it is contained in this or that side has different limits as $n \rightarrow \infty$. (In fact, my guess is that these two limit probabilities converge to 0 as $\epsilon \downarrow 0$ with different exponents in ϵ .) It would also suffice to show that the limit law of this path is different from the law in a symmetric $n \times n$ square.

Now, what is the “obvious asymmetry”? The edges open at level $1/2$ form critical percolation, hence, for small ϵ , it is unlikely that there is a $1/2$ -cluster connecting the two neighbourhoods. But then, the invasion tree from x to y will certainly use edges on the right side to travel between the $1/2$ -clusters, since all the labels larger than $1/2$ are smaller on the right than on the left. On the other hand, the invasion tree will explore the entire $1/2$ -cluster on the left once having entered one, since all the labels smaller than $1/2$ are smaller on the left than on the right, while it will explore only parts of the $1/2$ -clusters on the right. So, the MST path has more options on the right side for long distances, while more options on the left for short distances. The effects of this competition have to be understood well in order to produce an actual proof, but simulations suggest that the path spends more time on the right than on the left, and seems to intersect the midline less than in the symmetric situation.



This conformal non-invariance proof might be easier for Invasion Percolation, using known differences between invasion and critical percolation clusters [16].

References

- [1] G. Grimmett, *Percolation*, second edn. (Springer-Verlag, Berlin, 1999). Grundlehren der mathematischen Wissenschaften 321.
- [2] W. Werner, *Lectures on two-dimensional critical percolation* (2007), IAS Park City Graduate Summer School. [arXiv:0710.0856](#)
- [3] S. Smirnov, *C. R. Acad. Sci. Paris Sér. I Math.* **333**, 239 (2001).
- [4] O. Schramm, *Israel J. Math.* **118**, 221 (2000).
- [5] J. E. Steif, A survey of dynamical percolation (2009). [arXiv:0901.4760](#)
- [6] C. Garban, G. Pete and O. Schramm, The Fourier spectrum of critical percolation (2008), *Acta Math.*, to appear. [arXiv:0803.3750](#)
- [7] C. Garban, G. Pete and O. Schramm, Papers in preparation.
- [8] A. Hammond, G. Pete and O. Schramm, Local time for dynamical percolation, and the Incipient Infinite Cluster. In preparation.
- [9] O. Schramm and S. Smirnov, In preparation.
- [10] F. Camia, L. R. Fontes and C. Newman, *Bull. Braz. Math. Soc. (N.S.)* **37**, 537 (2006).
- [11] H. Kesten, *Comm. Math. Phys.* **109**, 109 (1987).
- [12] M. Aizenman, A. Burchard, C. M. Newman and D. B. Wilson, *Random Structures Algorithms* **15**, 319 (1999).
- [13] K. S. Alexander, *Ann. Probab.* **23**, 87 (1995).
- [14] R. Lyons, Y. Peres and O. Schramm, *Ann. Probab.* **34**, 1665 (2006).
- [15] D. B. Wilson, *Physical Review E* **69** (2004), 037105.
- [16] M. Damron, A. Sapozhnikov and B. Vágvolgyi, Relations between invasion percolation and critical percolation in two dimensions (2008). [arXiv:0806.2425](#)

SCALING LIMITS OF DYSON'S β -ENSEMBLE

BENEDEK VALKÓ

*Department of Mathematics, University of Wisconsin Madison
480 Lincoln Drive, Madison WI 53706
E-mail: valko@math.wisc.edu
www.math.wisc.edu/~valko*

Dyson's β -ensemble gives a one parameter generalization for the joint eigenvalue distribution of the Gaussian orthogonal and unitary ensemble. We review some recent results on the scaling limits of this ensemble and also discuss some large deviation results.

Keywords: random matrices, eigenvalue distributions

1. Dyson's β -ensemble

The study of the asymptotic behavior of eigenvalues of large random Hermitian matrices has yielded many applications to several fields of mathematics and physics. The simplest of such random models can be defined by taking an $n \times n$ matrix M with independent standard complex (resp., real) Gaussian random variables and averaging it with its adjoint: $A = \frac{M+M^*}{\sqrt{2}}$. The distribution of the resulting random matrix is called the Gaussian unitary ensemble or GUE (resp., Gaussian orthogonal ensemble or GOE). The joint density of the eigenvalues $\lambda_1 < \lambda_2 < \dots < \lambda_n$ is given by ([1], [2])

$$\frac{1}{Z_{n,\beta}} e^{-\beta \sum_{k=1}^n \lambda_k^2/4} \prod_{j < k} |\lambda_j - \lambda_k|^\beta, \quad (1)$$

where $\beta = 2$ (resp., $\beta = 1$) for the GUE (resp., GOE) and $Z_{n,\beta}$ is an explicitly computable constant.

It is natural to consider the density (1) for a general $\beta > 0$, the resulting one parameter family is called Dyson's β -ensemble. (Sometimes the names Gaussian or Hermite β -ensemble are used.) On a physical level (1) describes a one dimensional Coulomb gas with two dimensional interaction in a Gaussian potential at inverse temperature β . It was first studied in detail by Dyson (see e.g. [3]). It is also connected to the Calogero-Sutherland model ([4], [5]). The $\beta = 4$ case can also be obtained from a classical random matrix model called the Gaussian Symplectic Ensemble (GSE).

In [6] Dumitriu and Edelman showed that for every $\beta > 0$ the density (1) can be represented using certain random tridiagonal matrices. They proved that if

$X_1, X_2, \dots, Y_1, Y_2, \dots$ are independent random variables where $\frac{1}{\sqrt{2}}X_k$ is a standard Gaussian and Y_k has distribution $\chi_{\beta k}$ then the joint eigenvalue density of following $n \times n$ tridiagonal matrix is given by (1).

$$M(n) = \frac{1}{\sqrt{\beta}} \begin{bmatrix} X_n & Y_{n-1} & & \\ Y_{n-1} & X_{n-1} & Y_{n-2} & \\ & Y_{n-2} & X_{n-2} & \ddots \\ & & \ddots & \ddots \end{bmatrix} \quad (2)$$

Note that χ_d is just the square root of a $\Gamma(d/2, 1)$ random variable. If $d \in \mathbb{Z}$ this has the same distribution as the length of a d -dimensional standard Gaussian vector.

2. Scaling limits

It has been known since Wigner's work from the 1950s [7] that if we rescale the eigenvalues of an $n \times n$ GUE (or GOE) random matrix by $n^{-1/2}$ then their empirical distribution converges to the semicircle distribution given by the density $\frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{\{|x|\leq 2\}}$. This implies that 'most' of the eigenvalues will be in the interval $[-2\sqrt{n}, 2\sqrt{n}]$. It can also be shown that the same scaling limit holds in general for the β -ensemble, for any $\beta > 0$ (e.g. [2]).

In order to understand the finer asymptotic properties we could try to rescale the ensemble in a way that the average density becomes $\mathcal{O}(1)$ and prove that there exists a point process limit. The order of the scaling parameters can be guessed from the semicircle law. If we are interested in the asymptotic behavior in the 'bulk' (i.e. near $c\sqrt{n}$ with $|c| < 2$) then the order of the scaling factor should be $n^{1/2}$ and if we are interested in the edge behavior (near $\pm 2\sqrt{n}$) then the scaling factor should be $n^{1/6}$.

Some of the most celebrated results of random matrix theory deal with the bulk and edge scaling limits of the $\beta = 1, 2, 4$ cases. The limiting point processes have determinantal (for $\beta = 2$) or Pfaffian ($\beta = 1, 4$) structure. (See [1] or [2] for a complete treatment of these results.)

The tridiagonal representation (2) provides a tool to analyze the scaling limits for general values of β . In [8] and [9] the authors argued that applying the 'edge'-scaling to the tridiagonal matrix $M(n)$ in (2) one gets the following random Schrödinger operator in the limit:

$$\mathcal{H}_\beta = -\frac{d^2}{dx^2} + x + \frac{2}{\sqrt{\beta}}\dot{W}(x) \quad (3)$$

where $\dot{W}(x)$ is standard white noise. In [10] Ramírez, Rider and Virág gave a rigorous definition of the operator $\mathcal{H}_\beta$ and proved that the point process limit of the β -ensemble at the edge is given by the eigenvalue process of $\mathcal{H}_\beta$. They also give an alternative characterization for the limiting point process using the passage times of a certain diffusion process.

In a joint work with B. Virág in [11] we treat the bulk scaling limit of the general β -ensembles. The main result of that paper is the following:

Let Λ_n denote the point process given by density (1), and let $|c| < 2$. Then

$$\sqrt{4 - c^2}(\sqrt{n}\Lambda_n - cn) \Rightarrow \text{Sine}_\beta,$$

where Sine_β is a translation invariant discrete point process given by the Brownian carousel with speed function $f(t) = \frac{\beta}{4}e^{-\frac{\beta}{4}t}$.

In order to define the Brownian carousel we need to introduce an object called the hyperbolic carousel. We need the following ingredients: (i) a continuous path $b(t)$ in the 2-dimensional hyperbolic plane, (ii) a point z on the boundary of the hyperbolic plane, and (iii) an integrable function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ (the speed function). To these three objects, the hyperbolic carousel associates a multi-set of points on the real line defined via its counting function $N(\lambda)$ taking values in $\mathbb{Z} \cup \{-\infty, \infty\}$. As time increases from 0 to ∞ , the boundary point z is rotated about the center $b(t)$ at angular speed $\lambda f(t)$. $N(\lambda)$ is defined as the integer-valued total winding number

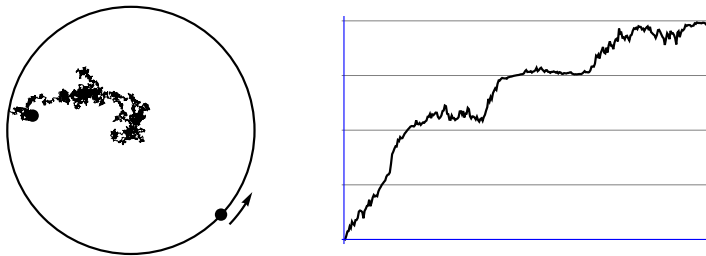


Fig. 1. The Brownian carousel and the winding angle α_λ

of the point about the moving center of rotation. Since $N(\lambda)$ is increasing in λ (as $f \geq 0$), this will indeed give a counting function.

The Brownian carousel is defined as a hyperbolic carousel with $b(t)$ being a standard Brownian motion in the hyperbolic plane. This geometric representation of Sine_β also gives a nice analytic tool to study the point process. If we denote the winding angle by $\alpha_\lambda(t)$ for the point corresponding to λ moving on the boundary in the carousel at time t then Itô's formula gives the the following family of stochastic differential equations:

$$d\alpha_\lambda = \lambda f dt + \Re((e^{-i\alpha_\lambda} - 1)dZ), \quad \alpha_\lambda(0) = 0, \quad (4)$$

driven by a two-dimensional standard Brownian motion Z . For a single λ , this reduces to the one-dimensional SDE

$$d\alpha_\lambda = \lambda f dt + 2 \sin(\alpha_\lambda/2)dW, \quad \alpha_\lambda(0) = 0. \quad (5)$$

As $t \rightarrow \infty$ the solution of this SDE converges to $\alpha_\lambda(\infty)$, an integer multiple of 2π . The counting function $N(\lambda)$ of the point process will be given by $\frac{1}{2\pi}\alpha_\lambda(\infty)$.

A description similar to (4) also appears in [12] in relation to the point process limits of the so-called circular ensembles.

3. Large gaps in the β -ensemble

One way to describe the interaction between the eigenvalues of GUE or in more general the random elements of the β -ensemble is to study the probability of having a large gap. To be more precise: we examine the asymptotic probability of having no element in a fixed interval of size λ (where λ is large) as $n \rightarrow \infty$ while the ensemble is rescaled to have an average spacing 2π . This is the same as the probability p_λ that point process Sine_β does not have any points in the interval $[0, \lambda]$.

This problem was first considered by Wigner for $\beta = 1$ [13], he gave a heuristic argument that p_λ should decay as $e^{-c\lambda^2}$. This is in contrast to the exponential decay one would get in case of a Poisson point process. Dyson in [3] improved this estimate to

$$p_\lambda = (\kappa_\beta + o(1))\lambda^{\gamma_\beta} \exp\left(-\frac{\beta}{64}\lambda^2 + \left(\frac{\beta}{8} - \frac{1}{4}\right)\lambda\right). \quad (6)$$

Dyson's computation of the exponent γ_β , namely $\frac{1}{4}(\frac{\beta}{2} + \frac{2}{\beta} + 6)$, was shown to be slightly incorrect. Indeed, [14] gave more substantiated predictions that γ_β is equal to $-1/8, -1/4$ and $-1/8$ for values $\beta = 1, 2$ and 4 , respectively. For these classical cases the gap probability can be computed using Fredholm determinants. Thus the problem can be attacked using various advanced analytical methods which resulted in precise proofs of (6) (see [15], [16]), together with the identification of the constant κ_β and higher order asymptotics (see e.g. [17], [18], [19]).

In [20] we prove the validity of the asymptotics (6) for all values of β with the corrected exponent $\gamma_\beta = \frac{1}{4}\left(\frac{\beta}{2} + \frac{2}{\beta} - 3\right)$ and an undetermined (but nonzero) constant κ_β . Using the SDE description of the Sine_β process we have that p_λ is equal to the probability that the solution of the SDE (5) converges to 0 as $t \rightarrow \infty$. In order to compute this probability we construct a diffusion process which approximates the conditional distribution of the diffusion α_λ given that $\alpha_\lambda(\infty) = 0$. Then the probability in question can be approximated by taking the expectation of the Radon-Nikodym derivative of the new process (using the Cameron-Martin-Girsanov formula) with respect to the distribution of the unconditioned process α_λ .

A possible analogue of this problem in the edge scaling is to analyze the tail probabilities of the largest point in the scaling limit of the β -ensemble. In the $\beta = 1, 2, 4$ cases the exact scaling limit of the largest eigenvalue has been computed by Tracy and Widom [21], [22]. The asymptotic tail probabilities for the

limiting random variable $TW(\beta)$, $\beta = 1, 2, 4$ is given by

$$P(TV(\beta) > \lambda) = \exp\left(-\frac{2}{3}\beta\lambda^{3/2}(1 + o(1))\right), \quad (7)$$

$$P(TV(\beta) < -\lambda) = \exp\left(-\frac{1}{24}\beta\lambda^3(1 + o(1))\right). \quad (8)$$

as $\lambda \rightarrow \infty$. In [10] the authors extend these results for all values of β . Their proofs rely on the various representations of the limiting point process: for certain parts they work with the random operator (3) and for other parts they use the diffusion representation. Recently [23] obtained the same asymptotics for the upper tail estimate (7) with a better (logarithmic) error term. Her proof builds on the ideas used for the bulk case [20].

References

- [1] M. L. Mehta, *Random matrices* (Elsevier/Academic Press, Amsterdam, 2004).
- [2] P. Forrester, *Log-gases and Random matrices* Book in preparation
www.ms.unimelb.edu.au/~matpjf/matpjf.html.
- [3] F. Dyson, *J. Math. Phys.* **3**, 157 (1962).
- [4] F. Calogero, *Journal of Mathematical Physics* **10**, 2197 (1969).
- [5] B. Sutherland, *Journal of Mathematical Physics* **12**, 246 (1971).
- [6] I. Dumitriu and A. Edelman, *J. Math. Phys.* **43**, 5830 (2002).
- [7] E. P. Wigner, *Ann. of Math. (2)* **62**, 548 (1955).
- [8] B. D. Sutton, The stochastic operator approach to random matrix theory (2005), Ph.D. thesis, MIT, Department of Mathematics.
- [9] A. Edelman and B. D. Sutton, *J. Stat. Phys.* **127**, 1121 (2007).
- [10] J. Ramirez, B. Rider and B. Virág, Beta ensembles, stochastic Airy spectrum, and a diffusion (2007), math/0607331.
- [11] B. Valkó and B. Virág, *Inventiones Mathematicae* **177**, 463 (2009).
- [12] R. Killip and M. Stoiciu, *Duke Math. J.* **146**, 361 (2009).
- [13] E. P. Wigner, *Gatlinberg Conference on Neutron Physics, Oak Ridge National Laboratory Report ORNL 2309*, p. 59 (1957).
- [14] J. des Cloizeaux and M. L. Mehta, *J. Mathematical Phys.* **14**, 1648 (1973).
- [15] H. Widom, *J. Approx. Theory* **77**, 51 (1996).
- [16] P. Deift, A. Its and X. Zhou, *Ann. Math.* **146**, 149 (1997).
- [17] I. V. Krasovsky, *Int. Math. Res. Not.*, 1249 (2004).
- [18] T. Ehrhardt, *Comm. Math. Phys.* **262**, 317 (2006).
- [19] P. Deift, A. Its, I. Krasovsky and X. Zhou, *J. Comput. Appl. Math.* **202**, 26 (2007).
- [20] B. Valkó and B. Virág, Large gaps between random eigenvalues (2008), arXiv:0811.0007.
- [21] C. A. Tracy and H. Widom, *Comm. Math. Phys.* **159**, 151 (1994).
- [22] C. A. Tracy and H. Widom, *Comm. Math. Phys.* **177**, 727 (1996).
- [23] L. Dumaz, The tracy-widom right tail (2009), M. Sc. thesis, University of Toronto, Department of Mathematics.

OTHER TALKS AND POSTERS

INVITED TALKS

INTERLACING PARTICLE SYSTEMS

ALEXEI BORODIN

California Institute of Technology

E-mail: borodin@caltech.edu

The goal of the talk is to survey recent results on interlacing particle systems that arise in representation theory, random matrices, and dimer models.

PATH TRANSFORMATIONS OF BROWNIAN MOTION CONNECTED WITH RANDOM MATRICES AND REPRESENTATION THEORY

PHILIPPE BIANE, PHILIPPE BOUGEROL, NEIL O'CONNELL*

** University of Warwick*

E-mail: n.m.o-connell@warwick.ac.uk

I will describe some recent work in which we introduce certain path-transformations of Brownian motion which have remarkable properties from both algebraic and probabilistic points of view. These are closely related to random matrices and representation theory, and have applications to directed percolation. I will also discuss a 'finite-temperature' version of this framework which is related to the quantum Toda lattice and has applications to a directed polymer model in $1+1$ dimensions.

CONTRIBUTED TALKS

MAXIMUM LOCAL ROUGHNESS IN 2D BROWNIAN AND PERCOLATION MODELS

ALLAN HAMMOND*, YUVAL PERES

** New York University*

E-mail: alanmh@cims.nyu.edu

As a model of the boundary curve between two phases, consider the boundary of a finite cluster in supercritical percolation, conditioned to entrap a large area; or a low-temperature Ising model in a finite region, where an excess of plus signs is conditioned to be present, the excess typically collecting in a single large droplet; or a planar Brownian motion conditioned to enclose a large area, where the droplet is asymptotically circular. We discuss several techniques that address the fluctuation behaviour of such boundary curves. There are several ways to measure this fluctuation, including the length of the longest line segment in the convex boundary of the curve, and the maximum local roughness, which measures the inward deviation of the curve from this boundary. We will explain a rough probabilistic equivalence between the first of these quantities and the "area excess" enclosed by the curve, for some of the models in question.

ON ESTIMATES OF THE MOMENTS OF WIGNER RANDOM MATRICES

OLEKSII KHORUNZHIY

Université de Versailles – Saint-Quentin

E-mail: khorunzhiy@math.uvsq.fr

Unifying two different approaches proposed by Z. Füredi and J. Komlós from one side and Ya. Sinai and A. Soshnikov from another side, we develop a method to estimate the high moments of large hermitian random matrices whose entries are given by jointly independent centered random variables with given variance. This allows us to generalize the results obtained by A. Soshnikov on the universality conjecture of the edge eigenvalue distribution in the Wigner ensemble of random matrices.

RANDOM MATRICES WITH EXTERNAL SOURCE

PAVEL M. BLEHER, ARNO B.J. KUIJLAARS*

**Katholieke Universiteit Leuven, Belgium*

E-mail: arno.kuijlaars@wis.kuleuven.be

We consider the random matrix model with external source

$$\frac{1}{Z_n} \exp(-n(\text{Tr} V(M) - AM)) dM$$

where the external source A has two distinct eigenvalues of equal multiplicities. In the Gaussian case the model is equivalent to a model of non-intersecting Brownian motions, and a phase transition occurs that is described by Pearcey integrals. For a general even polynomial potential V we present a vector equilibrium problem for the asymptotic density of eigenvalues. The equilibrium problem indicates that the Pearcey phase transition is non-generic.

POSTERS

KHAIRIA EL-SAID EL-NADI (Alexandria University, Faculty of Science): **ON SOME STOCHASTIC DYNAMICAL SYSTEMS AND CANCER**

JOSEP FREIXAS (Technical University of Catalonia), with MONTSERRAT PONS: **ON THE INFLUENCE OF THE NODE CRITICALITY RELATION ON CIRCUIT AND MECHANICS SYSTEMS**

TARO NAGAO (Graduate School of Mathematics, Nagoya University): **PERIODIC ORBIT THEORY AND PARAMETRIC RANDOM MATRICES**

MARIA ALBINA PUENTE (Technical University of Catalonia), with FRANCESC CARRERAS and JOSEP FREIXAS: **RELIABILITY OF SELF-DUAL K-OUT-OF-N SYSTEMS**

EVELINA SHAMAROVA (University of Porto), with ANA BELA CRUZEIRO: **NAVIER-STOKES EQUATIONS AND FORWARD-BACKWARD SDES ON THE GROUP OF DIFFEOMORPHISMS OF A TORUS**

Operator Algebras

(including noncommutative geometry)

Session organizers:

Detlev Buchholz
Masaki Izumi

DEFORMATIONS OF OPERATOR ALGEBRAS AND THE CONSTRUCTION OF QUANTUM FIELD THEORIES

GANDALF LECHNER

*Vienna University, Faculty of Physics,
Boltzmannngasse 5, 1090 Vienna, Austria
E-mail: gandalf.lechner@univie.ac.at*

This note outlines a novel approach to the construction of quantum field theories on four-dimensional Minkowski space, which is based on operator-algebraic techniques. It is explained how Rieffel deformations can be used to deform quantum field theories in a manner which is compatible with Poincaré covariance and locality.

Keywords: Deformations of algebras, Rieffel deformation, deformation quantization, non-commutative geometry

1. Introduction: Wedge triples and local nets

In relativistic quantum field theory, the rigorous construction of interacting models in four space-time dimensions is still an open, challenging problem for mathematical physics. Despite the insights we have gained from constructive quantum field theory [1] and other methods [2], even the existence of interacting quantum field theories satisfying standard assumptions has not been proved until today, and it seems that new ideas are needed to improve this situation.

The aim of this contribution is to outline a recent development which uses operator-algebraic deformation techniques as a new tool in the construction of QFT models. This method is adapted to the framework of algebraic QFT [3, 4], where models are specified in terms of their local observable algebras: For each spacetime region $O \subset \mathbb{R}^4$ (in four-dimensional Minkowski space), one considers the algebra $\mathcal{A}(O)$ generated by all observables localized in O . Basic properties like Einstein causality and relativistic covariance can be formulated naturally in this algebraic language, and nowadays many tools exist to extract important physical quantities, such as the particle content, S-matrix, charge structure, thermal equilibrium states and local field content, from the *local net* $O \mapsto \mathcal{A}(O)$ [3].

In all algebraic construction procedures developed so far, a particular kind of unbounded region in Minkowski space plays a special role, the so-called *right wedge*

$$W_R := \{x = (x_0, x_1, x_2, x_3) \in \mathbb{R}^4 : x_1 > |x_0|\}.$$

The following algebraic structure is associated with this region.

Definition 1.1. A *wedge triple* $(\mathcal{A}, \mathcal{B}, \alpha)$ consists of an inclusion $\mathcal{A} \subset \mathcal{B}$ of C^* -algebras and a strongly continuous automorphic action of the Poincaré group $\mathcal{P}$ on $\mathcal{B}$, such that

- (1) $\alpha_g(\mathcal{A}) \subset \mathcal{A}$ for all $g \in \mathcal{P}$ with $gW_R \subset W_R$,
- (2) $\alpha_{g'}(\mathcal{A}) \subset \mathcal{A}' \cap \mathcal{B}$ for all $g' \in \mathcal{P}$ with $g'W_R \subset -W_R$,

where $\mathcal{A}' \cap \mathcal{B}$ denotes the relative commutant of $\mathcal{A}$ in $\mathcal{B}$.

The significance of wedge triples for QFT relies on the following facts [5]. On the one hand, any QFT defines a wedge triple: Take $\mathcal{B}$ as the C^* -algebra of all observables of the theory, localized anywhere in Minkowski space, and $\mathcal{A}$ as the C^* -algebra of all observables localized in the wedge W_R . Since we are considering relativistic field theories, $\mathcal{B}$ carries an automorphic action of the Poincaré group, and the two *consistency conditions* in the above definition hold because of the locality and covariance properties of the theory.

On the other hand, any wedge triple gives rise to an associated QFT. Namely, starting from a triple $(\mathcal{A}, \mathcal{B}, \alpha)$, one can define the “wedge algebras”

$$\mathcal{A}(\Lambda W_R + x) := \alpha_{x, \Lambda}(\mathcal{A}), \quad (1)$$

and the consistency conditions imply that this assignment of wedges (Poincaré transforms of W_R) to C^* -algebras, $W \mapsto \mathcal{A}(W)$, complies with the usual isotony, covariance and locality assumptions of QFT. Moreover, given a bounded spacetime region O , one can unambiguously construct the maximal algebra $\mathcal{A}(O)$ of all observables localized in O . These local algebras inherit the basic covariance and locality properties from the wedge triple. Since the construction of such $\mathcal{A}(O)$ involves intersections of subalgebras of $\mathcal{B}$, one has to check however that these algebras $\mathcal{A}(O)$ are non-trivial, i.e. that the QFT associated to the triple $(\mathcal{A}, \mathcal{B}, \alpha)$ contains strictly localized observables.

This two-sided relation between QFT models and wedge triples opens up the possibility to realize new QFTs by exploring examples of wedge triples. The construction then consists of first finding an appropriate triple $(\mathcal{A}, \mathcal{B}, \alpha)$, and then working out the physics of the QFT associated to this triple.

As explained in [5], for the construction of QFTs in their vacuum representation, one can work with a particular concrete form of wedge triples: Consider a Fock space $\mathcal{H}$ with its second quantized unitary strongly continuous representation U of $\mathcal{P}$, and let $\mathcal{B} := \mathcal{B}(\mathcal{H})$, $\alpha := \text{ad } U$. Then the task is to find $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ such that the two conditions of Def. 1.1 are satisfied. In this setting, the structure of $\mathcal{H}$ and U is dictated by the particle spectrum of the QFT to be described, and the choice of $\mathcal{A}$ encodes the interaction.

In two dimensions, the constructive program making use of wedge triples has already been carried through for a large family of interacting models, where the algebra $\mathcal{A}$ is set up with the help of a factorizing S-matrix in the spirit of inverse scattering theory [6–9]. Also in the four-dimensional case, a number of authors have used

wedge-localized objects for constructive purposes [10–13]. First wedge triples corresponding to interacting QFTs have been constructed in [14] and then generalized in [5] and [15]. These examples will be reviewed from a novel point of view in the following.

As non-trivial wedge triples are hard to construct from scratch, we will in the following consider the question how wedge triples can be *deformed*. That is, we will assume a triple $T := (\mathcal{A}, \mathcal{B}, \alpha)$ complying with Def. 1.1 to be given (for example by an interaction-free theory) and then “perturb” this structure in the family of wedge triples to some T_θ . Here θ is a “coupling constant” such that $T_0 = T$, and the QFT corresponding to the deformed triple T_θ exhibits non-trivial interaction for $\theta \neq 0$.

2. Rieffel deformations of wedge triples

The particular deformation of wedge triples which we want to discuss here relies on Rieffel’s work on deformations of C^* -algebras, and we briefly recall some aspects of this analysis [16]. Rieffel considers a unital C^* -algebra $\mathcal{B}$ equipped with a strongly continuous action α of $\mathbb{R}^d$ by automorphisms. In our application to wedge triples, $\mathcal{B}$ will be the larger algebra of a wedge triple $(\mathcal{A}, \mathcal{B}, \alpha)$, and the action α is given by restricting the Poincaré action of this triple to the translation subgroup $\mathbb{R}^4$.

Rieffel’s analysis is based on a deformed product $\times_\theta$ which can be introduced on the dense subalgebra $\mathcal{B}^\infty \subset \mathcal{B}$ of elements $B \in \mathcal{B}$ for which $x \mapsto \alpha_x(B)$ is smooth. To define it in our context, we equip $\mathbb{R}^4$ with the Minkowski inner product and consider as deformation parameter a real (4×4) -matrix θ which is antisymmetric with respect to the Minkowski product. Then the deformed product $\times_\theta$ on $\mathcal{B}^\infty$ is given by the integral formula

$$A \times_\theta B := (2\pi)^{-d} \int_{\mathbb{R}^d} dp \int_{\mathbb{R}^d} dx e^{-ipx} \alpha_{\theta p}(A) \alpha_x(B), \quad A, B \in \mathcal{B}^\infty. \quad (2)$$

Note that the integral (2) has to be defined in an oscillatory sense.

Some results established in the general context of Rieffel deformations are the following: $\times_\theta$ carries $\mathcal{B}^\infty$ into itself and is jointly continuous in the natural Fréchet topology of this algebra, it is an associative product and reproduces the undeformed product in $\mathcal{B}$ for $\theta = 0$, i.e. $A \times_0 B = AB$. Furthermore, $\times_\theta$ is compatible with the identity and star involution of $\mathcal{B}$.

We now consider a wedge triple $(\mathcal{A}, \mathcal{B}, \alpha)$, and want to use Rieffel’s procedure to deform it. For the purposes of this note, we restrict ourselves to consider only a realization associated with a $\mathcal{P}$ -invariant state. That is, in the following we put $\mathcal{B} := \mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$ which carries a strongly continuous unitary representation U of the Poincaré group with a U -invariant vector Ω . The action α is the adjoint action of U on $\mathcal{B}(\mathcal{H})$, and $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ a C^* -subalgebra.

To deform $\mathcal{A}$, we consider the left multipliers $B \mapsto (A \times_\theta B)$ w.r.t. the deformed

product (2). Since Ω is U -invariant, these products evaluate on Ω to

$$A_\theta : B\Omega \longmapsto (A \times_\theta B)\Omega = (2\pi)^{-4} \int dp \int dx e^{-ipx} U(\theta p) A U(x - \theta p) B\Omega, \quad (3)$$

and define a new, deformed, operator A_θ , if the integral is interpreted in the same oscillatory sense as before. This operator can also be represented as a “warped convolution” of A by the spectral measure of U [5], and this point of view shows in particular that the p -integration in (3) runs only over the spectrum $S \subset \mathbb{R}^4$ of U , while the x -integration runs over all of $\mathbb{R}^4$. That is, we have

$$A_\theta = (2\pi)^{-4} \int_S dp \int dx e^{-ipx} U(\theta p) A U(x - \theta p), \quad (4)$$

and define the deformed wedge algebra $\mathcal{A}_\theta$ as the C^* -algebra generated by all A_θ , where $A \in \mathcal{A}$ is smooth.

We thus consider the deformed triple $(\mathcal{A}_\theta, \mathcal{B}(\mathcal{H}), \text{ad } U)$, and turn to the crucial question under which conditions these data still define a wedge triple, such that it can be used to build a QFT model.

Because of the form of the conditions in Def. 1.1, one has to study the effect of Lorentz transformations on the operators A_θ to answer this question. A proof cannot be given in this short contribution, but we will at least sketch what the main mechanisms are which are relevant for making the Rieffel deformation preserve the wedge triple structure.

One first computes that general Poincaré transformations (a, Λ) consisting of a translation $a \in \mathbb{R}^4$ and a Lorentz transformation Λ act on the Rieffel product $\times_\theta$ according to

$$\alpha_{a,\Lambda}(A \times_\theta B) = \alpha_{a,\Lambda}(A) \times_{\Lambda\theta\Lambda^{-1}} \alpha_{a,\Lambda}(B). \quad (5)$$

We thus have to compare Rieffel deformations with different deformation parameters, and first choose a suitable θ adapted to the geometry of W_R . Let

$$\theta := \begin{pmatrix} 0 & \kappa_1 & 0 & 0 \\ \kappa_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa_2 \\ 0 & 0 & -\kappa_2 & 0 \end{pmatrix}, \quad \kappa_1, \kappa_2 \in \mathbb{R}, \quad (6)$$

which is antisymmetric w.r.t. the Minkowski inner product. Then the transformations (a, Λ) appearing in Def. 1.1 can be characterized as follows [14]:

- $\Lambda W_R + a \subset W_R \Leftrightarrow \Lambda\theta\Lambda^{-1} = \theta$ and $a \in \overline{W_R}$,
- $\Lambda W_R + a \subset -W_R \Leftrightarrow \Lambda\theta\Lambda^{-1} = -\theta$ and $a \in -\overline{W_R}$.

So the wedge-preserving transformations appearing in the first condition in Def. 1.1 preserve θ , and the wedge-reflecting transformations appearing in the second condition map θ to $-\theta$.

With this choice of θ , it is straightforward to verify that the deformed algebra $\mathcal{A}_\theta$ satisfies the first consistency condition for any $\kappa_1, \kappa_2 \in \mathbb{R}$.

The second condition, related to the locality of the deformed theory, is more involved. Since the transformations mapping W_R into its causal complement correspond to $-\theta$, we must study the relation between the Rieffel deformations with parameters θ and $-\theta$. The second condition is valid if for all smooth $A \in \mathcal{A}$, $A' \in \mathcal{A}'$, $B \in \mathcal{B}(\mathcal{H})$,

$$0 = (A \times_{\theta} (A' \times_{-\theta} B) - A' \times_{-\theta} (A \times_{\theta} B))\Omega = (A_{\theta}A'_{-\theta} - A'_{-\theta}A_{\theta})B\Omega.$$

Using Rieffel's formula (2), this commutator can be computed as the integral

$$[A_{\theta}, A'_{-\theta}] = (2\pi)^{-4} \int_S dp \int dx e^{-ipx} U(\frac{x}{2}) [\alpha_{\theta p}(A), \alpha_{-\theta p}(A')] U(\frac{x}{2}), \quad (7)$$

where as before, S denotes the joint spectrum of the generators of the representation of the translations.

Since $\mathcal{A}$ is stable under translations in the direction of $\overline{W_R}$, the commutant $\mathcal{A}'$ is stable under translations in the opposite direction. Hence if $\theta p \in \overline{W_R}$ for all p in the above integral, the commutator vanishes. As observed by Buchholz and Summers [5], this situation is realized if one considers a vacuum representation where S is contained in the forward lightcone, and the parameter κ_1 is non-negative.

Theorem 2.1. *Let θ be of the form (6), and let the spectrum of $U|_{\mathbb{R}^4}$ be contained in the forward lightcone. Then the Rieffel-deformed triple $(\mathcal{A}_{\theta}, \mathcal{B}(\mathcal{H}), \text{ad}U)$ is a wedge triple for any $\kappa_1 \geq 0$, $\kappa_2 \in \mathbb{R}$.*

This theorem has been obtained in the context of warped convolution deformations [5]. If it is applied to an interaction-free field theory, this wedge triple defines a new QFT model [14].

The discussion of the properties of this field theory goes beyond the scope of this note. We only mention here that the two-particle S-matrix changes under the deformation [5, 14], which proves in particular that the deformed wedge triple is not equivalent to the undeformed one. In application to Wightman quantum field theories, the deformed theory is governed by θ -dependent fields ($\tilde{\phi}^{\theta}(p) = \tilde{\phi}(p)U(\theta p)$ in the scalar case), and can thus be interpreted as a field theory on non-commutative Minkowski space [15]. This point of view also explains why the QFTs obtained by Rieffel deformations have strong non-local features and do not contain strictly localized observables [5].

The details of the construction outlined here will soon be published in a joint paper with D. Buchholz and S. J. Summers. Beyond the example of Rieffel deformations, an infinite class of other deformations of wedge triples has been found which give rise to similar field theories^a. It seems that the family of wedge triples which lead to new QFT models is very large, and it would be interesting to systematically investigate which models can be obtained from a free field theory by such deformation procedures.

^aG. Lechner, work in progress

References

- [1] J. Glimm and A. Jaffe, *Quantum Physics. A Functional Integral Point Of View*, 2 edn. (Springer, New York, 1987). New York, Usa: Springer (1987) 535p.
- [2] H. M. Babujian and M. Karowski, *Int. J. Mod. Phys.* **A19S2**, 34 (2004).
- [3] R. Haag, *Local Quantum Physics - Fields, Particles, Algebras* 1996.
- [4] H. Araki, *Mathematical Theory of Quantum Fields* Int. Series of Monographs on Physics, Int. Series of Monographs on Physics (Oxford University Press, Oxford, 1999).
- [5] D. Buchholz and S. J. Summers, Warped Convolutions: A Novel Tool in the Construction of Quantum Field Theories, in *Quantum Field Theory and Beyond: Essays in Honor of Wolfhart Zimmermann*, eds. E. Seiler and K. Sibold (World Scientific, 2008).
- [6] B. Schroer and H. W. Wiesbrock, *Rev. Math. Phys.* **12**, 301 (2000).
- [7] G. Lechner, *Lett. Math. Phys.* **64**, 137 (2003).
- [8] D. Buchholz and G. Lechner, *Annales Henri Poincaré* **5**, 1065 (2004).
- [9] G. Lechner, *Commun. Math. Phys.* **277**, 821 (2008).
- [10] R. Brunetti, D. Guido and R. Longo, *Rev. Math. Phys.* **14**, 759 (2002).
- [11] R. Longo and K.-H. Rehren, *Rev. Math. Phys.* **16**, p. 909 (2004).
- [12] J. Mund, B. Schroer and J. Yngvason, *Commun. Math. Phys.* **268**, 621 (2006).
- [13] D. Buchholz and S. J. Summers, *J. Phys.* **A40**, 2147 (2007).
- [14] H. Grosse and G. Lechner, *JHEP* **11**, p. 012 (2007).
- [15] H. Grosse and G. Lechner, *JHEP* **09**, p. 131 (2008).
- [16] M. A. Rieffel, *Deformation Quantization for Actions of R^d* , Memoirs of the American Mathematical Society, Vol. 106 (American Mathematical Society, Providence, Rhode Island, 1992).

ON SUPERSELECTION THEORY OF QUANTUM FIELDS IN LOW DIMENSIONS*

MICHAEL MÜGER

*Institute of Mathematics, Astrophysics and Particle Physics
Radboud University, Nijmegen, The Netherlands
E-mail: mueger@math.ru.nl*

We discuss finite local extensions of quantum field theories in low space time dimensions in connection with categorical structures and the question of modular invariants in conformal field theory, also touching upon purely mathematical ramifications.

1. Introduction: Superselection theory in Local Quantum Theory

The aim of this contribution is to briefly review some aspects of the ‘quantum symmetry’ question in low dimensional (typically conformal) quantum field theory (QFT), from the point of view of algebraic quantum field theory. The latter (also called ‘local quantum physics’) arguably is the most successful approach to the axiomatic study of QFT in Minkowski space (or on more general pseudo-Riemannian manifolds). Due to space constraints we must refer to [16] or [17] for typical lists of axioms and recall only that a QFT is a map $\mathcal{O} \mapsto \mathcal{A}(\mathcal{O})$ assigning to every decent spacetime region $\mathcal{O}$ an algebra of operators $\mathcal{A}(\mathcal{O})$ satisfying requirements like spacelike commutativity, covariance w.r.t. the group of spacetime symmetries, irreducibility, etc. Under these assumptions, augmented by Haag duality, it was shown in [9] that the category $\text{DHR}(\mathcal{A})$ of ‘localized representations’ (with finite statistical dimension) of a QFT $\mathcal{A}$ in Minkowski space of dimension $\geq 2+1$ is a rigid semisimple symmetric tensor $*$ -category with simple unit. Later it was proven [11] that every such category is the representation category of a unique compact group G . Furthermore [12], there is a (graded local) QFT $\mathcal{F}$ with the following properties: (1) It is acted upon by G and $\mathcal{F}^G \cong \mathcal{A}$. (2) The category $\text{DHR}(\mathcal{F})$ is trivial [4]. (This must be modified suitably when $\mathcal{A}$ has fermionic representations). (3) Upon restriction to $\mathcal{F}^G \cong \mathcal{A}$, the representation space of $\mathcal{F}$ decomposes into a direct sum of irreps, which are precisely the localized representations of $\mathcal{A}$. (4) Every local extension $\mathcal{B} \supset \mathcal{A}$ is isomorphic to $\mathcal{F}^H$ for some closed subgroup $H \subset G$. As argued in [17, Section 10.6], the above amounts to a very satisfactory Galois theory for QFTs. [17] also provides an alternative and more transparent approach to the

*Partially based on joint work with A. Kitaev and with A. Davydov, D. Nikshych and V. Ostrik

construction of $\mathcal{F}$, based on unpublished work by Roberts and Deligne's independent approach [7] to proving $\mathcal{C} \simeq \text{Rep } G$ for symmetric categories, cf. also [26].

2. What changes in low dimensional spacetimes?

When one considers QFTs in Minkowski space of dimension $1 + 1$ or just on the light ray $\mathbb{R}$, the category $\text{DHR}(\mathcal{A})$ can only be proven to be braided [14, I]. In [19] this was strengthened considerably, in that it was shown that $\text{DHR}(\mathcal{A})$ is maximally non-symmetric, to wit a modular category in the sense of Turaev, for the important class of 'completely rational' models (a rigorous axiomatization of rational chiral conformal QFTs). In this situation it is clear that $\text{DHR}(\mathcal{A})$ cannot be equivalent to a representation category of a group. Another advantage of the alternative construction of the field net [17] is that it also applies in low dimensions *provided* a fiber functor from $\text{DHR}(\mathcal{A})$ to the category of Hilbert spaces is given: Roberts' construction of the field net as described in [17] goes through essentially unchanged, and appealing to a version of Woronowicz's Tannakian theorem [29] one obtains a discrete algebraic quantum group acting on $\mathcal{F}$ and an R-matrix describing the spacelike commutation relations of $\mathcal{F}$ (which will be far from local), cf. [28].

However, in general a fiber functor for $\text{DHR}(\mathcal{A})$ will not exist (as can be proven e.g. when the category is finite and contains objects of non-integer dimension). At least in the case of a finite category, one can always get around this problem obtaining [30] a weak Hopf algebra, acting on the reduced field bundle of [14, II], but it is not clear that this is useful: The weak Hopf algebra is not uniquely determined by $\text{DHR}(\mathcal{A})$ so that it does not have an intrinsic physical meaning. (Furthermore, the inclusion $\mathcal{A} \subset \mathcal{F}$ has the undesirable property of being reducible.) For many purposes it seems better to resign oneself to considering the category $\text{DHR}(\mathcal{A})$ itself as the fundamental structure. The rest of this note is an attempt to convince the reader of the feasibility, even elegance, of such an approach. Cf. [28] for details.

3. Categorical analysis of local extensions

The first question we address is the classification of extensions of a QFT, in particular local ones and their representation categories. By an extension $\mathcal{B} \supset \mathcal{A}$ of a QFT we mean an inclusion preserving assignment $\mathcal{O} \mapsto \mathcal{B}(\mathcal{O}) \supset \mathcal{A}(\mathcal{O})$ satisfying covariance and irreducibility, but not necessarily locality. If $\mathcal{B}$ is local we call it a local extension. In [23] it was shown that there is a bijection between (unitary equivalence classes of) extensions $\mathcal{B} \supset \mathcal{A}$ that are finite (i.e. $[\mathcal{B}(\mathcal{O}) : \mathcal{A}(\mathcal{O})]$ is independent of $\mathcal{O}$ and finite) and (equivalence classes of) Q-systems in the category $\text{DHR}(\mathcal{A})$. (A Q-system in a $*$ -category is essentially a monoid/algebra (Γ, m, η) such that $(\Gamma, m, \eta, m^*, \eta^*)$ is a Frobenius algebra.) Furthermore, $\mathcal{B}$ is local iff (Γ, m, η) is commutative, i.e. $m \circ c(\Gamma, \Gamma) = m$, where c is the braiding.^a When this is the case, it

^aFrom now on we will simply write 'commutative algebra' with the understanding that $\dim \text{Hom}(\mathbf{1}, \Gamma) = 1$ and that the algebra is 'strongly separable symmetric Frobenius' or 'étale' [5].

is natural to ask for a determination of the category $\mathrm{DHR}(\mathcal{B})$ in terms of $\mathrm{DHR}(\mathcal{A})$ and (Γ, m, η) . The answer is given by:

Theorem 3.1. [28] *Let $B \supset \mathcal{A}$ be the extension of the completely rational theory $\mathcal{A}$ corresponding to the commutative algebra (Γ, m, η) in $\mathrm{DHR}(\mathcal{A})$. Then there is an equivalence*

$$\mathrm{DHR}(B) \simeq \Gamma - \mathrm{Mod}_{\mathrm{DHR}(\mathcal{A})}^0$$

of braided tensor categories. In particular, $\dim(\mathrm{DHR}(B)) = \dim(\mathrm{DHR}(\mathcal{A}))/d(\mathcal{A})^2$.

Remark 3.1. 1. Here, $\Gamma - \mathrm{Mod}_{\mathrm{DHR}(\mathcal{A})}^0$ is the full subcategory of the category $\Gamma - \mathrm{Mod}_{\mathrm{DHR}(\mathcal{A})}$ of left Γ -modules consisting of the objects (X, μ) satisfying $\mu \circ c(X, \Gamma) \circ c(\Gamma, X) = \mu$. These ‘dyslexic’ or ‘local’ modules form a braided tensor category in a canonical way, cf. [31]. While the proof of Theorem 3.1 is too long to be exhibited here, it can be given in a few pages including the prerequisites. It uses results of [1] on α -induction to obtain a full and faithful braided tensor functor $\Gamma - \mathrm{Mod}_{\mathrm{DHR}(\mathcal{A})}^0 \rightarrow \mathrm{DHR}(\mathcal{B})$ and [19] together with the result $\dim \Gamma - \mathrm{Mod}_{\mathcal{C}}^0 = \dim \mathcal{C}/d(\Gamma)^2$ from [20] to conclude essential surjectivity.

2. The analogous result in the context of vertex operator algebras appears in [20], but many hard VOA technicalities have not yet been worked out.

3. The above result also holds in $\geq 2 + 1$ dimensions where the representation categories are symmetric and every Γ -module is local. But in this case, $\mathrm{DHR}(\mathcal{A})$ is equivalent to $\mathrm{Rep} G$ and commutative algebras in $\mathrm{Rep} G$ correspond to subgroups of $H \subset G$, cf. [30]. Thus we recover (at least for finite extensions) the Galois correspondence mentioned in the introduction.

4. In view of the above result, almost all questions concerning finite local extensions of QFTs and their representation categories can be reduced to purely categorical, i.e. algebraic considerations. An obvious exception to this is the inverse problem, to wit the question which categories are realized in some QFT model. (The analogous question in the group theoretic situation $d \geq 2 + 1$ was settled in [10].)

Recall that in $d \geq 2 + 1$ dimensions, the extension $\mathcal{F}$ has trivial representation category. In low dimensions it is *not* true that a local extension with this property always exists, since in view of Theorem 3.1, we have:

Corollary 3.1. *A completely rational theory $\mathcal{A}$ admits a local extension with trivial representation category iff $\mathcal{C} = \mathrm{DHR}(\mathcal{A})$ contains a commutative algebra Γ such that $\Gamma - \mathrm{Mod}_{\mathcal{C}}^0$ is trivial, equivalently $d(\Gamma)^2 = \dim \mathcal{C}$.*

(Notice that every commutative algebra Γ in a modular category $\mathcal{C}$ satisfies $d(\Gamma)^2 \leq \dim \mathcal{C}$, cf. [20].) The question when such a commutative algebra exists is answered by the following remarkable

Theorem 3.2. *A commutative algebra Γ in a modular category gives rise to a braided equivalence $Z(\Gamma - \mathrm{Mod}_{\mathcal{C}}) \simeq \widehat{\mathcal{C} \boxtimes \Gamma - \mathrm{Mod}_{\mathcal{C}}^0}$. Thus if $\Gamma - \mathrm{Mod}_{\mathcal{C}}^0$ is trivial,*

equivalently $d(\Gamma)^2 = \dim \mathcal{C}$, then $\mathcal{C} \simeq Z(\mathcal{D})$, where $\mathcal{D} = \Gamma - \text{Mod}_{\mathcal{C}}$ is a fusion category. Conversely, if $\mathcal{D}$ is a fusion category, then $Z(\mathcal{D})$ contains a commutative algebra Γ with $d(\Gamma)^2 = (\dim \mathcal{D})^2 = \dim Z(\mathcal{D})$.

Here $Z(\cdot)$ is the categorical center of a tensor category due to Drinfeld, Majid, Joyal and Street which was proven [24] to be modular for every fusion category $\mathcal{D}$ with $\dim \mathcal{D} \neq 0$. Theorem 3.2 was obtained by Kitaev and the author [21]. The first half was discovered independently in [13], and results related to the second half are obtained in [3]. Combining Corollary 3.1 and Theorem 3.2 we have:

Corollary 3.2. *A completely rational theory $\mathcal{A}$ admits a local extension with trivial representation category iff the modular category $\text{DHR}(\mathcal{A})$ is a Drinfeld center.*

The $\Rightarrow$ direction actually seems to be more useful. E.g., one can use it to prove:

Corollary 3.3. *Let $\mathcal{B}$ be a completely rational theory with trivial representation category acted upon freely by a finite group G , and let $\mathcal{A} = \mathcal{B}^G$ be the ‘orbifold’ theory. Then there is a unique class $[\omega] \in H^3(G, \mathbb{T})$ such that $\text{DHR}(\mathcal{A}) \simeq D^\omega(G) - \text{Mod}$, where $D^\omega(G)$ is the twisted quantum double of G [8].*

Remark 3.2. A more general analysis of orbifold QFTs $\mathcal{A} = \mathcal{B}^G$ without the triviality assumption on $\text{DHR}(\mathcal{B})$ was given in [25]. However, this did not yield the above corollary since a certain coherence theorem for braided crossed G -categories was missing. But see [27].

Even when $\mathcal{A}$ admits no local extension $\mathcal{B} \supset \mathcal{A}$ with trivial $\text{DHR}(\mathcal{B})$, it follows easily from the above results that maximal local extensions always exist and that every local extension embeds into a maximal one. Contrary to the situation for $d \geq 2 + 1$, it is not true that there is a unique (up to unitary equivalence) maximal local extension. (E.g., if $\mathcal{A}$ is completely rational with $\text{DHR}(\mathcal{A}) \simeq D(G) - \text{Mod}$, then there exist at least two non-isomorphic commutative algebras $\Gamma, \hat{\Gamma}$ such that $d(\Gamma) = d(\hat{\Gamma}) = |G|$, giving rise to local extensions that are not unitarily equivalent.)

However, by a result of [5], cf. Section 5, all maximal local extensions have braided equivalent representation categories. In particular, if a local extension with trivial representation category exists, then every local extension embeds into one with trivial representation category.

4. Connections with Rehren’s approach to modular invariants

It is well known that a modular category $\mathcal{C}$ gives rise to a finite dimensional representation of the modular group $SL(2, \mathbb{Z})$, acting on the complexified Grothendieck group of $\mathcal{C}$. Given two modular categories $\mathcal{C}_L, \mathcal{C}_R$, a *modular invariant* used to be defined as a $\mathbb{Z}_{\geq 0}$ -valued matrix (Z_{ij}) , indexed by the respective sets I_L, I_R of simple objects, that satisfies $Z_{00} = 1$ and intertwines the associated representations of $SL(2, \mathbb{Z})$, i.e. $Z\pi_L(g) = \pi_R(g)Z$. This definition turned out to be insufficient, and from now on ‘modular invariant’ will mean the following for us:

Definition 4.1. A modular invariant for a pair $(\mathcal{C}_L, \mathcal{C}_R)$ of modular categories is a triple (Γ_L, Γ_R, E) , where Γ_L, Γ_R are commutative algebras in $\mathcal{C}_L, \mathcal{C}_R$, respectively, and E is a braided monoidal equivalence $\Gamma_L - \text{Mod}_{\mathcal{C}_L}^0 \rightarrow \Gamma_R - \text{Mod}_{\mathcal{C}_R}^0$.

The question immediately arises whether two given categories $\mathcal{C}_L, \mathcal{C}_R$ admit a modular invariant. In the left-right symmetric case $\mathcal{C}_L = \mathcal{C}_R = \mathcal{C}$ at least one modular invariant always exists, to wit the triple $(\mathbf{1}, \mathbf{1}, \text{id}_{\mathcal{C}})$ (giving rise to the diagonal matrix $Z = \mathbf{1}$). A very ingenious subfactor theoretic approach to finding non-trivial modular invariants in this left-right symmetric situation was initiated in [1] and pursued further in [2]. This approach turned out [15, 30] to be of essentially categorical nature, revolving around algebras in $\mathcal{C}$ that are not necessarily commutative and thus do not have an interpretation in terms of local extensions. A more physically motivated approach, applicable also when $\mathcal{C}_L \not\cong \mathcal{C}_R$, was proposed in [32], where it was shown that, given two chiral CQFTs A_L, A_R and considering their product $\mathcal{A} = A_L \otimes \overline{A_R}$ as a QFT on $1 + 1$ -dimensional Minkowski space, every finite local extension $\mathcal{B} \supset \mathcal{A}$ gives rise to a matrix Z satisfying $Z_{00} = 1$ and intertwining the T -matrices of the modular categories $\mathcal{C}_L, \mathcal{C}_R$. Furthermore, it was conjectured that $ZS_L = S_R Z$ iff $\text{DHR}(\mathcal{B})$ is trivial. Proofs of this were found by the author and by R. Longo and Y. Kawahigashi (both remained unpublished). In [32] it was also shown that, given such a local extension $\mathcal{B} \supset A_L \otimes \overline{A_R}$, there are chiral local extensions $\widehat{A}_L \supset A_L, \widehat{A}_R \supset A_R$ that are maximal w.r.t. the property $A_L \otimes \overline{A_R} \subset \widehat{A}_L \otimes \overline{\widehat{A}_R} \subset \mathcal{B}$ and have isomorphic fusion rings. In [18] it was even shown that one has an equivalence $E : \text{DHR}(\widehat{A}_L) \rightarrow \text{DHR}(\widehat{A}_R)$ of braided tensor categories. Recalling the correspondence [23] between local extensions $\mathcal{B} \supset A$ and commutative algebras in $\text{DHR}(A)$ as well as Theorem 3.1, we conclude that a commutative algebra $\Gamma \in \text{DHR}(A_L \otimes \overline{A_R}) \simeq \text{DHR}(A_L) \boxtimes \widetilde{\text{DHR}(\widehat{A}_R)}$ of maximal dimension gives rise to a modular invariant for $\mathcal{C}_L, \mathcal{C}_R$. The converse follows from the fact, resulting from [23] and [19], that $A \otimes \overline{A}$ admits a local extension with trivial representation category. At this point, two questions arise:

- (1) What is the significance of the above reasoning for the construction of $d = 2$ CQFTs? Perhaps the requirement that $\mathcal{B}$ have trivial representation category amounts to the absence of an obstruction to existence of a ‘Wick rotated’ Euclidean CQFT that can be defined on arbitrary Riemann surfaces? (The fact that the DHR category has a cohomological interpretation in terms of Roberts’ local cohomology may be taken as further support for this speculation.)
- (2) Do the above results hold irrespective of whether the categories $\mathcal{C}_{L/R}$ arise from chiral CFTs $A_{L/R}$? The answer is yes:

Theorem 4.1. *Given two modular categories $\mathcal{C}_L, \mathcal{C}_R$, a modular invariant (Γ_L, Γ_R, E) gives rise to a commutative algebra $\Gamma \in \mathcal{C}_L \boxtimes \widetilde{\mathcal{C}_R}$ of maximal dimension (i.e. $d(\Gamma) = (\dim \mathcal{C}_L \cdot \dim \mathcal{C}_R)^{1/2}$). Conversely, every such algebra arises from a modular invariant [5], which in fact is unique [6].*

(In the case $\mathcal{C}_L = \mathcal{C}_R$ this was proven in [22].) The proof of the first half requires little more than ideas already contained in [23] and [20]. The converse implication follows from facts outlined in the final section.

5. On the classification of modular categories

The circle of ideas of the preceding section has important pure mathematical ramifications, cf. [5]. Namely, given modular categories $\mathcal{C}_L, \mathcal{C}_R$, it is not hard to see that existence of a fusion category $\mathcal{D}$ such that $\mathcal{C}_L \boxtimes \widetilde{\mathcal{C}_R} \simeq Z(\mathcal{D})$ is equivalent to existence of fusion categories $\mathcal{E}_1, \mathcal{E}_2$ such that $\mathcal{C}_L \boxtimes Z(\mathcal{E}_1) \simeq \mathcal{C}_R \boxtimes Z(\mathcal{E}_2)$. In this case we write $\mathcal{C}_L \sim \mathcal{C}_R$, and one easily sees that $\sim$ is an equivalence relation, called Witt equivalence in analogy with the classical theory of quadratic forms. The set W_M of Witt equivalence classes of modular categories acquires an abelian monoid structure by $[\mathcal{C}_1] \cdot [\mathcal{C}_2] := [\mathcal{C}_1 \boxtimes \mathcal{C}_2]$ and $\mathbf{1}_{W_M} = [\text{Vect}]$. In view of $\mathcal{C} \boxtimes \widetilde{\mathcal{C}} \simeq Z(\mathcal{C})$ (by [24]) and $Z(\mathcal{C}) \sim \mathbf{1}$ (by definition), W_M is a group with inverse operation $[\mathcal{C}]^{-1} = [\widetilde{\mathcal{C}}]$.

In view of Theorem 3.2, if Γ is a commutative algebra in $\mathcal{C}$ then $[\mathcal{C}] = [\Gamma - \text{Mod}_{\mathcal{C}}^0]$. (Thus if $A \subset B$ are completely rational CFTs then $[\text{DHR}(A)] = [\text{DHR}(B)]$.) Calling a modular category *completely anisotropic* if every commutative algebra in it is isomorphic to $\mathbf{1}$, one can prove that Witt-equivalent completely anisotropic categories are actually braided equivalent. (In other words: every Witt equivalence class contains one completely anisotropic category up to braided equivalence). This is the reason for the uniqueness of the representation category of maximal local extensions and of the converse statement in Theorem 4.1: By Theorem 3.2, existence of a maximal commutative algebra in $\mathcal{C}_L \boxtimes \widetilde{\mathcal{C}_R}$ implies $\mathcal{C}_L \boxtimes \widetilde{\mathcal{C}_R} \simeq Z(\mathcal{D})$ and thus $[\mathcal{C}_L] = [\mathcal{C}_R]$. Choosing commutative algebras $\Gamma_{L/R} \in \mathcal{C}_{L/R}$ such that the respective local module categories are completely anisotropic, we have $[\Gamma_L - \text{Mod}_{\mathcal{C}_L}^0] = [\mathcal{C}_L] = [\mathcal{C}_R] = [\Gamma_R - \text{Mod}_{\mathcal{C}_R}^0]$ and thus a braided equivalence $\Gamma_L - \text{Mod}_{\mathcal{C}_L}^0 \simeq \Gamma_R - \text{Mod}_{\mathcal{C}_R}^0$.

Finally, the Witt group holds great promise for the classification of modular categories. The point is that $Z(\mathcal{C})$ is modular for every fusion category, of which there are far too many to hope for a classification. (The fact that inequivalent fusion categories can have equivalent centers [24] does not help much.) Passing to the Witt group not only kills those ‘trivial’ modular categories but has the nice effect of yielding an abelian group W_M . Many generators (from quantum groups at roots of unity) and relations (from conformal extensions and cosets) for W_M are already known, and one may hope that W_M can be determined completely. This would seem to be a rigorous implementation to the idea from CQFT folklore that “all modular categories” arise from the chiral WZW models via local extensions and coset constructions.

References

- [1] J. Böckenhauer, D. E. Evans: Modular invariants, graphs and α -induction for nets of subfactors I-III. *Commun. Math. Phys.* **197**, 361-386 (1998); **200**, 57-103 (1999); **205**, 183-228 (1999).

- [2] J. Böckenhauer, D. E. Evans, Y. Kawahigashi: Commun. Math. Phys. **208**, 429-487 (1999); **210**, 733-784 (2000); Publ. Res. Inst. Math. Sci. **37**, 1-35 (2001).
- [3] A. Bruguières, A. Virelizier: The Double of a Hopf monad. [arXiv:0812.2443](#).
- [4] R. Conti, S. Doplicher, J. E. Roberts: Superselection theory for subsystems. Commun. Math. Phys. **218**, 263-281 (2001).
- [5] A. Davydov, M. Müger, D. Nikshych, V. Ostrik: In preparation
- [6] A. Davydov, M. Müger: Work in progress.
- [7] P. Deligne: Catégories tannakiennes. In: P. Cartier et al. (eds.): *Grothendieck Festschrift*, vol. II, pp. 111-195. Birkhäuser Verlag, 1991.
- [8] R. Dijkgraaf, V. Pasquier, P. Roche: Quasi Hopf algebras, group cohomology and orbifold models. Nucl. Phys. B (Proc. Suppl.) **18B**, 60-72 (1990).
- [9] S. Doplicher, R. Haag, J. E. Roberts: Local observables and particle statistics I & II. Commun. Math. Phys. **23**, 199-230 (1971); **35**, 49-85 (1974).
- [10] S. Doplicher, G. Piacitelli: Any compact group is a gauge group. Rev. Math. Phys. **14**, 873-885 (2002).
- [11] S. Doplicher, J. E. Roberts: A new duality theory for compact groups. Invent. Math. **98**, 157-218 (1989).
- [12] S. Doplicher, J. E. Roberts: Why there is a field algebra with a compact gauge group describing the superselection structure in particle physics. Commun. Math. Phys. **131**, 51-107 (1990).
- [13] V. Drinfeld, S. Gelaki, D. Nikshych, V. Ostrik: Group-theoretical properties of nilpotent modular categories. [math.QA/0704.1095](#).
- [14] K. Fredenhagen, K.-H. Rehren, B. Schroer: Superselection sectors with braid group statistics and exchange algebras I & II. Commun. Math. Phys. **125**, 201-226 (1989), Rev. Math. Phys., Special Issue 1992, 113-157.
- [15] J. Fröhlich, J. Fuchs, I. Runkel, C. Schweigert: Correspondences of ribbon categories. Adv. Math. **199**, 192-329 (2006).
- [16] R. Haag: *Local Quantum Physics*. 2nd ed., Springer, 1996
- [17] H. Halvorson: Algebraic quantum field theory. In: J. Butterfield & J. Earman (eds.): *Handbook of the Philosophy of Physics*. Elsevier, 2006.
- [18] Y. Kawahigashi, R. Longo: Classification of two-dimensional local conformal nets with $c < 1$ and 2-cohomology vanishing for tensor categories. Commun. Math. Phys. **244**, 63-97 (2004).
- [19] Y. Kawahigashi, R. Longo, M. Müger: Multi-interval subfactors and modularity of representations in conformal field theory. Commun. Math. Phys. **219**, 631-669 (2001).
- [20] A. Kirillov Jr., V. Ostrik: On q-analog of McKay correspondence and ADE classification of $sl(2)$ conformal field theories. Adv. Math. **171**, 183-227 (2002).
- [21] A. Kitaev, M. Müger: Unpublished work 2006-2008.
- [22] L. Kong, I. Runkel: Cardy algebras and sewing constraints I. [arXiv:0807.3356](#).
- [23] R. Longo, K.-H. Rehren: Nets of subfactors. Rev. Math. Phys. **7**, 567-597 (1995).
- [24] M. Müger: From subfactors to categories and topology II. The quantum double of tensor categories and subfactors. J. Pure Appl. Alg. **180**, 159-219 (2003).
- [25] M. Müger: Conformal orbifold theories and braided crossed G-categories. Commun. Math. Phys. **260**, 727-762 (2005).
- [26] M. Müger: Abstract duality theory for symmetric tensor $*$ -categories. App. to [17].
- [27] M. Müger: On braided crossed G-categories. App. to V. Turaev: *Homotopy Quantum Field Theory*. EMS, to appear.
- [28] M. Müger: Local extensions and modular invariants of quantum fields. In preparation.
- [29] M. Müger, J. E. Roberts, L. Tuset: Representations of algebraic quantum groups and reconstruction theorems for tensor categories. Alg. Repres. Theor. **7**, 517-573 (2004).

- [30] V. Ostrik: Module categories, weak Hopf algebras and modular invariants. *Transf. Groups* **8**, 177-206 (2003).
- [31] B. Pareigis: On braiding and dyslexia. *J. Algebra* **171**, 413-425 (1995).
- [32] K.-H. Rehren: Locality and modular invariance in 2D conformal QFT. *Fields Inst. Commun.* **20**, 297-319 (2001).

QUANTIZATIONS OF POISSON LIE GROUPS AS NONCOMMUTATIVE MANIFOLDS

SERGEY NESHVEYEV

*Department of Mathematics, University of Oslo,
P.O. Box 1053 Blindern, NO-0316 Oslo, Norway
E-mail: sergeyn@math.uio.no*

LARS TUSET

*Faculty of Engineering, Oslo University College,
P.O. Box 4 St. Olavs plass, NO-0130 Oslo, Norway
E-mail: Lars.Tuset@iu.hio.no*

On any q -deformation of a simply connected simple compact Poisson Lie group we construct an equivariant spectral triple which is an isospectral deformation of that defined by the Dirac operator on the original group. Our quantum Dirac operator is defined using a Drinfeld twist which relates the q -deformed compact quantum group to the original group, and thus a priori depends on the choice of the twist, but it turns out that the spectral triple is nevertheless unique up to unitary equivalence.

Introduction

By work of Drinfeld and Belavin all simply connected simple compact Poisson Lie groups have been classified. Drinfeld further showed that all of these do admit quantization by q -deformation. The quantization of the standard Poisson bracket on G is the compact quantum group G_q one usually encounters. The quantizations corresponding to the other brackets are obtained by twisting the coproduct of G_q or G with a 2-cocycle q^{iu} , where u is a self-adjoint element of $\mathfrak{h} \wedge \mathfrak{h}$.

In [1] we constructed a quantum Dirac operator D_q on G_q that defines a biequivariant spectral triple which is an isospectral deformation of that defined by the Dirac operator D on G . To do this we relied on the existence of a special element $\mathcal{F}$ in the group von Neumann algebra $W^*(G \times G)$, and an isomorphism $\varphi: W^*(G_q) \rightarrow W^*(G)$ satisfying certain properties. The existence of the pair $(\varphi, \mathcal{F})$ follows from work by Kazhdan and Lusztig [2, 3] and is an analytic version of a result by Drinfeld [4, 5]. From the outset D_q and the associated spectral triple depend on the choice of $(\varphi, \mathcal{F})$, but by a uniqueness result for Drinfeld twists, which we established in [6], we showed that the spectral triple is unique up to unitary equivalence.

In this note we briefly describe the above results and show that one can construct Dirac operators providing equivariant spectral triples for all the twisted versions as well.

1. Quantum groups

Let G be a compact connected simply connected simple Lie group, $\mathfrak{g}$ its complexified Lie algebra. Let $\mathfrak{h} \subset \mathfrak{g}$ be the Cartan subalgebra defined by a maximal torus T in G . Fix a system $\{\alpha_1, \dots, \alpha_r\}$ of simple roots. Let $(a_{ij})_{1 \leq i, j \leq r}$ be the Cartan matrix and $d_1, \dots, d_r$ be the coprime positive integers such that $(d_i a_{ij})_{i, j}$ is symmetric. Let $h_i \in \mathfrak{h}$ be such that $\alpha_j(h_i) = a_{ij}$. Denote by $\mathfrak{h}_{\mathbb{R}}$ the $\mathbb{R}$ -linear span of h_i , $i = 1, \dots, r$. Define a bilinear form on $\mathfrak{h}^*$ by $(\alpha_i, \alpha_j) = d_i a_{ij}$. The dual form on $\mathfrak{h}$ extends to a symmetric invariant form on $\mathfrak{g}$. Denote by $t \in \mathfrak{g} \otimes \mathfrak{g}$ the corresponding $\mathfrak{g}$ -invariant element. Consider the decomposition $\mathfrak{g} = \mathfrak{n}_+ \oplus \mathfrak{h} \oplus \mathfrak{n}_-$ and write $t = t_{+-} + t_0 + t_{-+}$ with $t_{+-} \in \mathfrak{n}_+ \otimes \mathfrak{n}_-$, $t_{-+} \in \mathfrak{n}_- \otimes \mathfrak{n}_+$ and $t_0 \in \mathfrak{h} \otimes \mathfrak{h}$. Define $r_0 \in \mathfrak{g} \otimes \mathfrak{g}$ by

$$r_0 = t_0 + 2t_{+-}.$$

It is known that any self-adjoint solution of the modified classical Yang-Baxter equation can be written as $ar_0 + u$ for some $a \in \mathbb{R}$ and $u \in \wedge^2 \mathfrak{h}_{\mathbb{R}}$. This means that any Poisson Lie group structure on G is given by the bracket $\{\cdot, \cdot\}: \mathbb{C}[G] \times \mathbb{C}[G] \rightarrow \mathbb{C}[G]$ defined by

$$\{f, g\} = (f \otimes g)([ar_0 + u, \hat{\Delta}(\cdot)]),$$

up to inner automorphisms of G and a complex rescaling of the bracket, where $\hat{\Delta}$ is the comultiplication on $U\mathfrak{g}$. They all admit quantization. For $a = 1$ and $u = 0$ the quantization is the standard q -deformation, and that for $a \neq 1$ (and $u = 0$) just means a reparametrization, i.e. change of q , of the standard one. We shall focus on the q -deformations for $0 < q < 1$ associated to $a = 1$ and $u \in \wedge^2 \mathfrak{h}_{\mathbb{R}}$; the case $a = 0$ is analogous but easier, one essentially has to replace $U_q\mathfrak{g}$ everywhere by $U\mathfrak{g}$.

The quantized universal enveloping algebra $U_q\mathfrak{g}$ is generated by elements E_i, F_i, K_i , $1 \leq i \leq r$, satisfying the relations

$$K_i K_j = K_j K_i, \quad K_i E_j K_i^{-1} = q_i^{a_{ij}} E_j, \quad K_i F_j K_i^{-1} = q_i^{-a_{ij}} F_j,$$

$$E_i F_j - F_j E_i = \delta_{ij} \frac{K_i - K_i^{-1}}{q_i - q_i^{-1}},$$

$$\sum_{k=0}^{1-a_{ij}} (-1)^k \begin{bmatrix} 1-a_{ij} \\ k \end{bmatrix}_{q_i} E_i^k E_j E_i^{1-a_{ij}-k} = 0,$$

$$\sum_{k=0}^{1-a_{ij}} (-1)^k \begin{bmatrix} 1-a_{ij} \\ k \end{bmatrix}_{q_i} F_i^k F_j F_i^{1-a_{ij}-k} = 0,$$

where $\begin{bmatrix} m \\ k \end{bmatrix}_{q_i} = \frac{[m]_{q_i}!}{[k]_{q_i}! [m-k]_{q_i}!}$, $[m]_{q_i}! = [m]_{q_i} [m-1]_{q_i} \dots [1]_{q_i}$, $[n]_{q_i} = \frac{q_i^n - q_i^{-n}}{q_i - q_i^{-1}}$ and $q_i = q^{d_i}$.

This is a Hopf $*$ -algebra with coproduct $\hat{\Delta}_q$ defined by

$$\hat{\Delta}_q(K_i) = K_i \otimes K_i, \quad \hat{\Delta}_q(E_i) = E_i \otimes 1 + K_i \otimes E_i, \quad \hat{\Delta}_q(F_i) = F_i \otimes K_i^{-1} + 1 \otimes F_i$$

and involution

$$K_i^* = K_i, E_i^* = F_i K_i, F_i^* = K_i^{-1} E_i.$$

Let P be the lattice of integral weights. If V is a finite dimensional $U_q \mathfrak{g}$ -module and $\lambda \in P$, denote by $V(\lambda)$ the space of vectors $v \in V$ of weight λ , that is, $K_i v = q_i^{\lambda(h_i)} v$. The module V is called admissible if $V = \bigoplus_{\lambda \in P} V(\lambda)$.

The quantized algebra of regular functions $\mathbb{C}[G_q] \subset (U_q \mathfrak{g})^*$ is the Hopf $*$ -algebra of matrix coefficients of finite dimensional admissible $U_q \mathfrak{g}$ -modules.

Denote by $\mathcal{C}(\mathfrak{g}, q)$ the category of finite dimensional admissible $U_q \mathfrak{g}$ -modules. It is a semisimple tensor category, with simple objects V_λ indexed by dominant integral weights $\lambda \in P_+$. Therefore, if $F: \mathcal{C}(\mathfrak{g}, q) \rightarrow \mathcal{V}ec$ is the forgetful functor, then

$$\mathcal{U}(G_q) := \text{Nat}(F, F) \cong \prod_{\lambda \in P_+} \text{End}(V_\lambda).$$

The algebra $\mathcal{U}(G_q)$ is a completion of $U_q \mathfrak{g}$. It can also be identified with the algebra of closed densely defined operators affiliated with the von Neumann algebra $W^*(G_q)$ of G_q . The category $\mathcal{C}(\mathfrak{g}, q)$ is braided, with braiding $\sigma = \Sigma \mathcal{R}$, where Σ is the flip and $\mathcal{R} \in \mathcal{U}(G_q \times G_q) := \text{Nat}(F^{\otimes 2}, F^{\otimes 2})$ is the universal R -matrix.

The algebra $\mathcal{U}(T)$ embeds into $\mathcal{U}(G_q)$ by identifying h_i with a unique self-adjoint element $H_i \in \mathcal{U}(G_q)$ such that $K_i = q_i^{H_i}$. If $u \in \wedge^2 \mathfrak{h}_{\mathbb{R}}$ we can therefore consider the element q^{iu} as an element $\mathcal{H} \in \mathcal{U}(G_q \times G_q)$. Since $u^* = u$ and u is skew-symmetric, we have $\mathcal{H}^* = \mathcal{H}^{-1} = \mathcal{H}_{21}$. We also have $u \otimes 1 + (\hat{\Delta} \otimes \iota)u = 1 \otimes u + (\iota \otimes \hat{\Delta})u$, so that by commutativity of $\mathcal{U}(T)$ we get

$$(\mathcal{H} \otimes 1)(\hat{\Delta}_q \otimes \iota)(\mathcal{H}) = (1 \otimes \mathcal{H})(\iota \otimes \hat{\Delta}_q)(\mathcal{H}).$$

Therefore $\mathcal{H}$ is a unitary 2-cocycle for $(\mathcal{U}(G_q), \hat{\Delta}_q)$ with $\mathcal{H}^{-1} = \mathcal{H}_{21}$. Thus we can define a new coproduct $\hat{\Delta}_{u,q}: \mathcal{U}(G_q) \rightarrow \mathcal{U}(G_q \times G_q)$ by $\hat{\Delta}_{u,q} = \mathcal{H} \hat{\Delta}_q(\cdot) \mathcal{H}^{-1}$. It defines a new tensor structure on finite dimensional admissible $U_q \mathfrak{g}$ -modules. Denote by $\mathbb{C}[G_q^u]$ the corresponding Hopf $*$ -algebra of matrix coefficients. In other words, $\mathbb{C}[G_q^u]$ coincides with $\mathbb{C}[G_q]$ as a coalgebra, but has a new $*$ -algebra structure defined by the twist $\mathcal{H}$, see [7]. The compact quantum group G_q^u is the q -deformation of the Poisson Lie group G with the Poisson structure associated to u (and $a = 1$). The category of finite dimensional representations of G_q^u is braided with braiding defined by the R -matrix $\mathcal{R}^u = \mathcal{H}_{21} \mathcal{R} \mathcal{H}^{-1}$.

2. Drinfeld twist

Let $\hbar \in i\mathbb{R}$ be such that $q = e^{\pi i \hbar}$. Assume V_1, V_2, V_3 are finite dimensional $\mathfrak{g}$ -modules and put $V = V_1 \otimes V_2 \otimes V_3$. Consider $\text{End}(V)$ -valued solutions of the equation

$$w' = \hbar \left(\frac{t_{12}}{x} + \frac{t_{23}}{x-1} \right) w$$

on $(0, 1)$. There exist unique solutions G_0 and G_1 such that the functions $G_0(x)x^{-\hbar t_{12}}$ and $G_1(1-x)x^{-\hbar t_{23}}$ extend to holomorphic functions in the unit disc

with value 1 at $x = 0$. Hence there exists $\Phi_{KZ} \in \text{GL}(V)$ such that

$$G_0(x) = G_1(x)\Phi_{KZ} \quad \text{for all } x \in (0, 1).$$

The operators Φ_{KZ} for different V_1, V_2, V_3 define an element of $\mathcal{U}(G \times G \times G)$. It is a unitary 3-cocycle for $(\mathcal{U}(G), \hat{\Delta})$.

As irreducible representations of G and G_q are both parametrized by dominant integral weights, we have a canonical identification of the centers of $\mathcal{U}(G)$ and $\mathcal{U}(G_q)$. Since the dimensions of irreducible modules with the same highest weight do not depend on q , this identification extends to a $*$ -isomorphism $\varphi: \mathcal{U}(G_q) \rightarrow \mathcal{U}(G)$. Furthermore, the dimensions of the weight spaces do not depend on q either, which implies that we can arrange that $\varphi(K_i) = q_i^{h_i}$. In particular, $(\varphi \otimes \varphi)(\mathcal{H}) = q^{iu}$.

Theorem 2.1. *There exists a unitary $\mathcal{F} \in \mathcal{U}(G \times G)$ such that*

- (i) $(\varphi \otimes \varphi)\hat{\Delta}_{u,q} = \mathcal{F}\hat{\Delta}\varphi(\cdot)\mathcal{F}^{-1}$;
- (ii) $(\hat{\varepsilon} \otimes \iota)(\mathcal{F}) = (\iota \otimes \hat{\varepsilon})(\mathcal{F}) = 1$, where $\hat{\varepsilon}$ is the trivial representation of G ;
- (iii) $(\varphi \otimes \varphi)(\mathcal{R}^u) = \mathcal{F}_{21}q^t\mathcal{F}^{-1}$;
- (iv) $\Phi_{KZ} = (\iota \otimes \hat{\Delta})(\mathcal{F}^{-1})(1 \otimes \mathcal{F}^{-1})(\mathcal{F} \otimes 1)(\hat{\Delta} \otimes \iota)(\mathcal{F})$.

Proof. For $u = 0$ the existence of $\mathcal{F}$ follows by work of Kazhdan and Lusztig [2, 3], see [8]. Denote this element by $\mathcal{F}_0$. For general u the required element is $\mathcal{F} := q^{iu}\mathcal{F}_0$. Indeed, the only nontrivial property is (iv). Since the elements K_i are group-like, we have $\hat{\Delta} = (\varphi \otimes \varphi)\hat{\Delta}_q\varphi^{-1}$ on $\mathcal{U}(T)$. It follows that $\mathcal{F}_0$ commutes with elements of the form $\hat{\Delta}(w)$, $w \in \mathcal{U}(T)$. Then (iv) for $\mathcal{F}$ follows from the corresponding property of $\mathcal{F}_0$ and the fact that q^{iu} is a 2-cocycle for $(\mathcal{U}(T), \hat{\Delta})$. $\square$

We call $\mathcal{F}$ a unitary Drinfeld twist for G_q^u . We have the following uniqueness result.

Theorem 2.2. *Suppose $\mathcal{E}$ and $\mathcal{F}$ are two unitary Drinfeld twists for G_q^u for the same $*$ -isomorphism φ . Then there exists a unitary central element c in $\mathcal{U}(G)$ such that $\mathcal{E} = (c \otimes c)\mathcal{F}\hat{\Delta}(c)^{-1}$.*

Proof. The elements $q^{-iu}\mathcal{E}$ and $q^{-iu}\mathcal{F}$ are unitary Drinfeld twists for G_q . Therefore the result follows from [6, Theorem 5.2]. Briefly, the reason is that

$$\mathcal{G} := (\varphi^{-1} \otimes \varphi^{-1})(q^{-iu}\mathcal{E}\mathcal{F}^{-1}q^{iu})$$

is a symmetric invariant unitary 2-cocycle for $(\mathcal{U}(G_q), \hat{\Delta}_q)$, that is, $\mathcal{R}\mathcal{G} = \mathcal{G}_{21}\mathcal{R}$ and $\mathcal{G}$ commutes with elements of the form $\hat{\Delta}_q(w)$. Hence $\mathcal{G}$ is the coboundary of a central element in $\mathcal{U}(G_q)$ by [6, Theorem 2.1]. $\square$

3. Quantum Dirac operator

Denote by $\text{Cl}(\mathfrak{g})$ the complex Clifford algebra of $\mathfrak{g}$ and by $\gamma: \mathfrak{g} \rightarrow \text{Cl}(\mathfrak{g})$ the canonical embedding, so $\text{Cl}(\mathfrak{g})$ is generated by $\gamma(x)$, $x \in \mathfrak{g}$, and $\gamma(x)^2 = (x, x)1$. The adjoint

action of G on $\mathfrak{g}$ extends to an action of G on $\text{Cl}(\mathfrak{g})$ which lifts to a homomorphism $G \rightarrow \text{Spin}(\mathfrak{g})$. On the Lie algebra level it is given by

$$\mathfrak{g} \ni x \mapsto \widetilde{\text{ad}}(x) := -\frac{1}{4} \sum_i \gamma(x_i) \gamma([x, x^i]),$$

where $\{x_i\}_i$ is a basis in $\mathfrak{g}$ and $\{x^i\}_i$ is the dual basis. Fix a spin module, an irreducible $*$ -representation $s: \text{Cl}(\mathfrak{g}) \rightarrow \text{End}(\mathbb{S})$. Identifying the smooth sections of the spin bundle $S = G \times \mathbb{S}$ with $C^\infty(G) \otimes \mathbb{S}$, the Dirac operator $D: C^\infty(G) \otimes \mathbb{S} \rightarrow C^\infty(G) \otimes \mathbb{S}$ is given by

$$D = \sum_i \left(\partial(x_i) \otimes s\gamma(x^i) - \frac{1}{2} \otimes s(\gamma(x_i) \widetilde{\text{ad}}(x^i)) \right),$$

where ∂ is the representation of $U\mathfrak{g}$ by left-invariant differential operators. This can be written as $D = (\partial \otimes s)(\mathcal{D})$, where

$$\mathcal{D} = -(\iota \otimes \gamma)(t) - \sum_i \frac{1}{2} \otimes \gamma(x_i) \widetilde{\text{ad}}(x^i)$$

is an element of the non-commutative Weil algebra $U\mathfrak{g} \otimes \text{Cl}(\mathfrak{g})$.

Fix a $*$ -isomorphism $\varphi: \mathcal{U}(G_q^u) = \mathcal{U}(G_q) \rightarrow \mathcal{U}(G)$ as in the previous section and choose a unitary Drinfeld twist $\mathcal{F} \in \mathcal{U}(G \times G)$ for G_q^u . Define

$$\mathcal{D}_q^u = (\varphi^{-1} \otimes \iota)((\iota \otimes \widetilde{\text{ad}})(\mathcal{F})\mathcal{D}(\iota \otimes \widetilde{\text{ad}})(\mathcal{F}^{-1})) \in \mathcal{U}(G_q^u) \otimes \text{Cl}(\mathfrak{g}).$$

The quantum Dirac operator D_q^u is the unbounded operator on $L^2(G_q^u) \otimes \mathbb{S}$ defined by

$$D_q^u = (\partial_q^u \otimes s)(\mathcal{D}_q^u),$$

where ∂_q^u is the right regular representations of $\mathcal{U}(G_q^u)$ on $L^2(G_q^u)$.

Theorem 3.1. *The triple $(\mathbb{C}[G_q^u], L^2(G_q^u) \otimes \mathbb{S}, D_q^u)$ is a G_q^u -biequivariant spectral triple of the same parity as the dimension of G . It does not depend on the choice of φ and $\mathcal{F}$ up to unitary equivalence.*

The proof is identical to that of [1, Theorem 3.7] and [6, Theorem 6.1].

References

- [1] S. Neshveyev and L. Tuset, *arXiv:math/0703161v2 [math.OA]* (2007).
- [2] D. Kazhdan and G. Lusztig, *J. Amer. Math. Soc.* **7**, 335 (1994).
- [3] D. Kazhdan and G. Lusztig, *J. Amer. Math. Soc.* **7**, 383 (1994).
- [4] V. G. Drinfeld, *Leningrad Math. J.* **1**, 1419 (1990).
- [5] V. G. Drinfeld, *Leningrad Math. J.* **2**, 829 (1991).
- [6] S. Neshveyev and L. Tuset, *arXiv:0902.2365v1 [math.QA]* (2009).
- [7] S. Levendorskii and Y. Soibelman, *Commun. Math. Phys.* **139**, 141 (1991).
- [8] S. Neshveyev and L. Tuset, *arXiv:0711.4302v1 [math.QA]* (2007).

POSITIVITY IN RIEFFEL'S STRICT DEFORMATION QUANTIZATION

STEFAN WALDMANN

*Faculty for Mathematics and Physics
Physics Institute
Albert-Ludwigs-University Freiburg
Germany*

E-mail: Stefan.Waldmann@physik.uni-freiburg.de

We review a recent result on Rieffel's deformation quantization by actions of $\mathbb{R}^d$: it is shown that for every state ω_0 of the undeformed C^* -algebra $\mathcal{A}_0$ there is a continuous section of states $\omega(\hbar)$ through ω_0 . We outline the physical interpretation in terms of quantization.

Keywords: Rieffel deformation quantization; Positive deformation

1. Introduction: Continuous Fields of C^* -Algebras

The aim of this short note is to recall some recently found positivity properties [1] of Rieffel's deformation quantization by actions of $\mathbb{R}^d$.

Quantization can be formulated in many different settings. Here, we will focus on the following framework: starting with a C^* -algebra $\mathcal{A}$, which we view as the observable algebra of the classical system, we are interested in deformations of $\mathcal{A}$ depending on a quantization parameter $\hbar$. Though for quantization purposes $\mathcal{A}$ is commutative, this will not be necessary in the following. Moreover, in typical situations one has a Poisson bracket defined on a (dense) subalgebra of $\mathcal{A}$ which gives the “direction of deformation”.

To make sense out of the notion “deformation” in this C^* -algebraic framework we use continuous fields [2] of C^* -algebras to formulate things properly. We recall the definition: let T be a Hausdorff space which in our case will be the deformation parameter space. Suppose for every $\hbar \in T$ we have a C^* -algebra $\mathcal{A}(\hbar)$. The product $\prod_{\hbar \in T} \mathcal{A}(\hbar)$ is then equipped with the structure of a $*$ -algebra by pointwise operations in $\hbar$. Elements in $\prod_{\hbar \in T} \mathcal{A}(\hbar)$ will be called *sections* and are written as maps $a : \hbar \mapsto a(\hbar)$. A *continuous field structure* for $\{\mathcal{A}(\hbar)\}_{\hbar \in T}$ is now a collection of sections $\Gamma \subseteq \prod_{\hbar \in T} \mathcal{A}(\hbar)$ satisfying the following conditions:

- (1) Γ is a sub-vector space.
- (2) Γ is closed under pointwise products and $*$ -involution.
- (3) $\Gamma|_{\hbar} = \{a(\hbar) \mid a \in \Gamma\} \subseteq \mathcal{A}(\hbar)$ is dense for all $\hbar$.

- (4) $\hbar \mapsto \|a(\hbar)\|_{\hbar}$ is continuous for all $a \in \Gamma$
- (5) If $\tilde{a}$ is an arbitrary section which can locally in T be approximated uniformly by sections in Γ , i.e. for all $\hbar_0 \in T$ and all $\epsilon > 0$ there is a section $a \in \Gamma$ and an open neighbourhood $U \subseteq T$ of $\hbar_0$ such that for all $\hbar \in U$

$$\|\tilde{a}(\hbar) - a(\hbar)\|_{\hbar} \leq \epsilon,$$

then $\tilde{a} \in \Gamma$.

The idea is that we axiomatically describe how “continuous sections” should behave. For a given Γ , the sections $a \in \Gamma$ are called the continuous sections. Note that we have *not* specified a topology on the total space $\prod_{\hbar \in T} \mathcal{A}(\hbar)$ which would be more involved. Note also, that the “fibers” $\mathcal{A}(\hbar)$ will in general not be isomorphic for different $\hbar$. In particular, a continuous field is, in general, not locally trivial and hence not a “bundle” of C^* -algebras.

The idea for quantization is now that $\mathcal{A}(0)$ is the classical observable algebra and $\mathcal{A}(\hbar)$ for $\hbar \neq 0$ are the quantized ones. In this case, $T = [0, +\infty)$ will be just an interval (or a suitable subset with 0 as accumulation point). This point of view for quantization is developed in detail by Landsman [3] or Rieffel [4]. The more algebraic version of this deformation program with formal dependence in $\hbar$ originates from Bayen et. al. [5], see also the textbook [6] for a gentle introduction and more references.

2. Quantization of States

Having a deformation of the observables it is very natural to ask how the states behave under the deformation. Here the following definition turns out to be appropriate. Given a continuous field structure Γ for a collection $\{\mathcal{A}(\hbar)\}_{\hbar \in T}$ of C^* -algebras we say that a collection of states $\{\omega(\hbar)\}_{\hbar \in T}$ is a *continuous section of states* if the map $\hbar \mapsto \omega(\hbar)(a(\hbar))$ is continuous for all continuous sections $a \in \Gamma$, see [3, Def. 1.3.1].

We are now interested in the following situation. Let $\hbar_0 \in T$ be fixed (the classical limit) and let $\omega_{\hbar_0} : \mathcal{A}(\hbar_0) \rightarrow \mathbb{C}$ be a given state of the (classical limit) C^* -algebra $\mathcal{A}(\hbar_0)$. If for any given $\omega_{\hbar_0}$ it is possible to find a continuous section $\omega(\hbar)$ of states with $\omega(\hbar_0) = \omega_{\hbar_0}$ then we call the continuous field Γ *positive*.

The physical interpretation is the following: for a positive continuous field (which we view as a quantization of $\mathcal{A}(0)$) every classical state is the classical limit of quantum states. Clearly, this is very much desirable from a physical point of view as quantum theory is believed to be the more fundamental description of nature and hence should contain the classical description as appropriate limit, both for the observables and the states.

Note that in general the continuous section $\omega(\hbar)$ deforming $\omega_{\hbar_0}$ is far from being unique. There will be many quantum states yielding the same classical limit.

3. Rieffel's construction by actions of $\mathbb{R}^d$

Rieffel's construction [4] of a deformation of C^* -algebras by actions of $\mathbb{R}^d$ will give us particular continuous fields which are nevertheless omnipresent in quantization theory. In particular, canonical quantization of the classical phase space $\mathbb{R}^{2n}$ can be viewed as a Rieffel deformation. We outline the basic ideas of his construction.

Suppose a C^* -algebra $\mathcal{A} = \mathcal{A}(0)$ is endowed with a strongly continuous action α of $(\mathbb{R}^{2n}, +)$ by $*$ -automorphisms. Then consider the smooth vectors $\mathcal{A}^\infty$ of this action, i.e. those elements $a \in \mathcal{A}$ where $u \mapsto \alpha_u(a)$ is smooth. It is well-known that $\mathcal{A}^\infty$ is a dense $*$ -subalgebra endowed with a (finer) Fréchet topology. Next, choose a symplectic form θ on $\mathbb{R}^{2n}$ and let $\hbar > 0$. For $a, b \in \mathcal{A}^\infty$ the integral

$$a \star_{\hbar} b = \frac{1}{(\pi\hbar)^{2n}} \int \alpha_u(a) \alpha_v(b) e^{\frac{2i}{\hbar}\theta(u,v)} du dv \quad (1)$$

is well-defined as an oscillatory integral and endows $\mathcal{A}^\infty$ with a new associative product turning $\mathcal{A}^\infty$ into a Fréchet algebra. Moreover, the old $*$ -involution is still a $*$ -involution for $\star_{\hbar}$. Note that the integrand of (1) has *constant* norm since α_u and α_v are *isometric*. Thus a naive definition of the integral is not possible. One needs a more sophisticated oscillatory integral here.

The star product $\star_{\hbar}$ allows for an asymptotic expansion with respect to the $\mathcal{A}^\infty$ -topology, explicitly given by

$$a \star_{\hbar} b \xrightarrow{\hbar \rightarrow 0^+} \mu \circ e^{\frac{i\hbar}{2}\theta^{kl}\partial_k \otimes \partial_l}(a \otimes b), \quad (2)$$

where the partial derivatives are defined by means of the action as $\partial_k a = \frac{d}{dt}|_{t=0} \alpha_{te_k}(a)$ and $\mu(a \otimes b) = ab$ denotes the undeformed product. Here θ^{kl} are the components of the Poisson tensor associated to the symplectic two-form θ . We conclude that asymptotically, $\star_{\hbar}$ is the usual (formal) Weyl-Moyal star product. In case $\mathcal{A}$ is commutative, the first order commutator is just the canonical Poisson bracket determined by the partial derivatives and θ .

In a last step, Rieffel constructs also a C^* -norm $\|\cdot\|_{\hbar}$ for all the products $\star_{\hbar}$ turning $\mathcal{A}^\infty$ into a pre- C^* -algebra for all $\hbar$. Its completion to a C^* -algebra will be denoted by $\mathcal{A}(\hbar)$. It depends typically in a highly non-trivial way on $\hbar$. Since all the $\mathcal{A}(\hbar)$ contain $\mathcal{A}^\infty$ as a dense subspace it makes sense to speak of *constant* sections of this field of C^* -algebras. Then one shows that the constant sections determine a continuous field structure Γ by “completing” them with respect to the requirement (5) in the definition of a continuous field structure. This finally gives Rieffel's continuous field.

4. Positivity

We are now in the position to formulate the main result [1]: Rieffel's field is positive. Moreover, there is even a very explicit construction of the continuous section $\omega(\hbar)$ of states passing through a given state ω_0 of $\mathcal{A}(0)$.

We consider a fixed positive inner product g on $\mathbb{R}^{2n}$ which is *compatible* with θ , i.e. $g(u, v) = \theta(u, Jv)$ with a linear *complex structure* J . Recall that a complex

structure J is an endomorphism $J \in \text{End}(\mathbb{R}^{2n})$ with $J^2 = -\text{id}$. Note also, that such g always exist. Fixing one g , we consider the following convolution operator with the Gaussian determined by g . For $a \in \mathcal{A}$ we define

$$S_{\hbar}(a) = \frac{1}{(\pi\hbar)^n} \int e^{-\frac{g(u,u)}{\hbar}} \alpha_u(a) du. \quad (3)$$

Clearly, the above integral exists in the most naive way for all $a \in \mathcal{A}$. In fact, $S_{\hbar}(a) \in \mathcal{A}^\infty$ for all a .

The operator $S_{\hbar}$ depends now in a very nice way on $\hbar$. In fact, we have

$$\lim_{\hbar \rightarrow 0^+} S_{\hbar}(a) = a \quad (4)$$

for all $a \in \mathcal{A}$ with respect to the topology of $\mathcal{A}$ and, moreover, for $a \in \mathcal{A}^\infty$ with respect to the topology of $\mathcal{A}^\infty$. Moreover, in the $\mathcal{A}^\infty$ topology we have for $a \in \mathcal{A}^\infty$

$$\frac{d}{d\hbar} S_{\hbar}(a) = \frac{1}{4} S_{\hbar}(\Delta_g a), \quad (5)$$

where Δ_g is the Laplacian with respect to g and the partial derivatives coming from the action. From this, we immediately obtain the asymptotic expansion

$$S_{\hbar}(a) \xrightarrow{\hbar \rightarrow 0^+} e^{\frac{\hbar}{4} \Delta_g} a \quad (6)$$

in the topology of $\mathcal{A}^\infty$. This indicates that $S_{\hbar}$ plays the role of an equivalence transformation between the formal Weyl-Moyal star product and the formal Wick star product [7] determined by g . Since the Wick star product is known to have nicer positivity properties than the Weyl star product, one can use the operator $S_{\hbar}$ to correct the positive functionals ω_0 . In the formal setting this was done earlier both for the symplectic [8] and Poisson case [9]. In our present framework however, the operator $S_{\hbar}$ is far from being invertible. The image is in $\mathcal{A}^\infty$ and thus $S_{\hbar}$ is clearly not surjective. Thus there does not seem to be a Wick type deformation available directly.

Nevertheless, we have the following properties. For $a \in \mathcal{A}^\infty$ a simple computation gives

$$S_{\hbar}(a^* \star_{\hbar} a) = \frac{1}{(\pi\hbar)^n} \sum_{K \geq 0} \frac{1}{K!} \left(\frac{2}{\hbar}\right)^{|K|} a_K^* a_K, \quad (7)$$

where the series runs over all multiindices $K = (k_1, \dots, k_n)$ and

$$a_K = \int (z^1)^{k_1} e^{-\frac{|z^1|^2}{\hbar}} \dots (z^n)^{k_n} e^{-\frac{|z^n|^2}{\hbar}} \alpha_{z_1 e_1 + \dots + z_n e_n}(a) dz^1 \dots dz^n. \quad (8)$$

Here the $z_1, \dots, z_N$ are complex coordinates induced by J . The series in (7) converges in the topology of $\mathcal{A}^\infty$ and consists of *squares* of the undeformed product. But the $\mathcal{A}^\infty$ topology is finer than the original one. This allows to conclude that

$$S_{\hbar}(a^* \star_{\hbar} a) \in \mathcal{A}^+, \quad (9)$$

hence $S_{\hbar}$ maps squares to positive elements. In other words, $S_{\hbar}$ is a positive (and even completely positive) map from the deformed algebra to the undeformed.

Now there are two things to be checked: first, up to now the operator $S_{\hbar}$ is only defined on the pre- C^* -algebra $\mathcal{A}^\infty$ and not on $\mathcal{A}(\hbar)$. Nevertheless, one can show that $S_{\hbar}$ is continuous with respect to the deformed C^* -norm $\|\cdot\|_{\hbar}$ and thus extends to $\mathcal{A}(\hbar)$. Clearly, the positivity properties remain true under this completion. We arrive at a (completely) positive operator

$$S_{\hbar} : \mathcal{A}(\hbar) \longrightarrow \mathcal{A}. \quad (10)$$

Second, one checks that the $\hbar$ -dependence of $S_{\hbar}$ is well-behaved. In particular, on a constant and hence continuous section $a \in \mathcal{A}^\infty \subseteq \mathcal{A}(\hbar)$ we see that $\hbar \mapsto S_{\hbar}(a)$ is continuous with respect to the C^* -topology on the undeformed algebra $\mathcal{A}$. Therefore, if $\omega_0 : \mathcal{A} \longrightarrow \mathbb{C}$ is a state, it follows that $\hbar \mapsto \omega_0(S_{\hbar}(a))$ is on one hand a continuous map. Since the constant sections “generate” all continuous sections we conclude that $\hbar \mapsto \omega_0(S_{\hbar}(a(\hbar)))$ is continuous for all continuous sections $a \in \Gamma$, too. On the other hand, $\omega(\hbar) = \omega_0 \circ S_{\hbar}$ is a state of the deformed C^* -algebra $\mathcal{A}(\hbar)$ for all $\hbar \in [0, +\infty)$. Hence we proved the following theorem [1]:

Theorem 4.1 (Kaschek, Neumaier, Waldmann). *For every classical state $\omega_0 : \mathcal{A} \longrightarrow \mathbb{C}$*

$$\omega(\hbar) = \omega_0 \circ S_{\hbar} : \mathcal{A}(\hbar) \longrightarrow \mathbb{C} \quad (11)$$

defines a continuous section of states with $\omega(0) = \omega_0$.

There are several applications beyond the obvious ones in quantization theory. In particular, these deformed states have been used to explore the behaviour of the causal structure in non-commutative space-times [10, 11].

References

- [1] D. Kaschek, N. Neumaier and S. Waldmann, *J. Noncommut. Geom.* **3**, 361 (2009).
- [2] J. Dixmier, *C^* -Algebras* (North-Holland Publishing Co., Amsterdam, 1977). Translated from the French by Francis Jellett, North-Holland Mathematical Library, Vol. 15.
- [3] N. P. Landsman, *Mathematical Topics between Classical and Quantum Mechanics*, Springer Monographs in Mathematics (Springer-Verlag, Berlin, Heidelberg, New York, 1998).
- [4] M. A. Rieffel, *Mem. Amer. Math. Soc.* **106**, p. 93 pages (1993).
- [5] F. Bayen, M. Flato, C. Frönsdal, A. Lichnerowicz and D. Sternheimer, *Ann. Phys.* **111**, 61 (1978).
- [6] S. Waldmann, *Poisson-Geometrie und Deformationsquantisierung. Eine Einführung* (Springer-Verlag, Heidelberg, Berlin, New York, 2007).
- [7] M. Bordemann and S. Waldmann, *Lett. Math. Phys.* **41**, 243 (1997).
- [8] H. Bursztyn and S. Waldmann, On positive deformations of $*$ -algebras, in *Conférence Moshé Flato 1999. Quantization, Deformations, and Symmetries*, eds. G. Dito and D. Sternheimer (Mathematical Physics Studies(22) (Kluwer Academic Publishers, Dordrecht, Boston, London, 2000) pp. 69–80.
- [9] H. Bursztyn and S. Waldmann, *Lett. Math. Phys.* **72**, 143 (2005).
- [10] D. Bahns and S. Waldmann, *Rev. Math. Phys.* **19**, 273 (2007).
- [11] J. G. Heller, N. Neumaier and S. Waldmann, *Lett. Math. Phys.* **80**, 257 (2007).

OTHER TALKS AND POSTERS

INVITED TALKS

SYMMETRY BREAKING AND SPLIT PROPERTY IN QUANTUM SPIN CHAIN

TAKU MATSUI

Kyushu University, Fukuoka

E-mail: matsui@math.kyushu-u.ac.jp

Space-Time locality in quantum lattice models is quite different from that of relativistic quantum field theory as there is no constraint of the speed of light. Nevertheless, we have some results on lattice models analogous to those of local QFT. For example, Fredenhagen's cluster theorem claims that presence of the mass gap implies exponential decay of correlation and the theorem is valid for both local QFT satisfying standard axioms and quantum lattice models. In this talk, we explain the split property of states (weak statistical independence of two subsystems) implies non-trivial results on quantum spin chains. For example, split property automatically implies certain symmetry breaking for spin $1/2$ quantum spin chains. Haag duality plays an important role in our analysis. A part of results in this talk is joint work with M. Keyl, D. Schlingemann, R. Werner.

TOWARDS A SPECTRAL TRIPLE FOR LQG

RYSZARD NEST

Copenhagen University

E-mail: rnest@math.ku.dk

A semifinite spectral triple for an algebra canonically associated to canonical quantum gravity is constructed. The algebra is generated by based loops in a triangulation and its subdivisions. The underlying space can be seen as a gauge fixing of the unconstrained state space of Loop Quantum Gravity.

CONTRIBUTED TALKS

SPECTRAL CONDITIONS FOR POSITIVE MAPS IN MATRIX ALGEBRAS

DARIUSZ CHRUSCINSKI

Nicolaus Copernicus University, Toruń

E-mail: darch@fizyka.umk.pl

We provide partial classification of positive linear maps in matrix algebras which is based on a family of spectral conditions. This construction generalizes the celebrated Choi example of a map which is positive but not completely positive. It is shown how the spectral conditions enable one to construct linear maps on tensor products of matrix algebras which are positive but only on a convex subset of separable elements. Such maps provide basic tools to study quantum entanglement in multipartite systems.

PRO- C^* -DYNAMICAL SYSTEMS WITH CROSSED PRODUCTS C^* -ALGEBRAS

MARIA JOITA

University of Bucharest

E-mail: mjoita@fmi.unibuc.ro

A pro- C^* -algebra is a complete Hausdorff complex topological $*$ -algebra A whose topology is determined by its continuous C^* -seminorms in the sense that a net $\{a_i\}_{i \in I}$ converges to 0 in A if and only if the net $\{p(a_i)\}_{i \in I}$ converges to 0 for all continuous C^* -seminorms p on A . For a pro- C^* -algebra A , the set $S(A)$ of all continuous C^* -seminorms on A is directed. A pro- C^* -dynamical system is a triple (G, α, A) , where G is a locally compact group, A is a pro- C^* -algebra and α is a continuous action of G on A such that there is a cofinal subset of $S(A)$ consisting of the α -invariant continuous C^* -seminorms (that is, $p(\alpha_g(a)) = p(a)$ for any $a \in A$ and for all $g \in G$). The set $C_c(G, A)$ of all continuous functions from G to A with compact support is a $*$ -algebra with convolution of two functions as product and the involution defined by $f^\#(t) = \Delta(t)^{-1} \alpha_t(f(t^{-1})^*)$, where Δ is the modular function on G . For any $p \in S(A)$, the map $N_p: C_c(G, A) \rightarrow [0, \infty)$ defined by

$$N_p(f) = \int_G p(f(s)) ds$$

where ds denotes a left Haar measure on G , is a submultiplicative $*$ -seminorm on $C_c(G, A)$. The covariance algebra $L^1(G, \alpha, A)$ associated to (G, α, A) is an m^* -convex algebra with bounded approximate unit and it is obtained by the Hausdorff completion of $C_c(G, A)$ with respect to the topology defined by the family of submultiplicative $*$ -seminorms $\{N_p\}_{p \in S(A)}$. The full crossed product associated to (G, α, A) is the enveloping pro- C^* -algebra of the covariance algebra. In [S. J. Bhatt, D. J. Karia, Topological algebras with C^* -enveloping algebras, Proc. Indian Acad. Sci (Math. Soc.) 102 (1993), 201-215] are proved sufficient and necessary conditions under which an m^* -convex algebra with bounded approximate admits a C^* -algebra as enveloping pro- C^* -algebra. In this talk we discuss about pro- C^* -dynamical systems with crossed product C^* -algebras.

NUMERICAL RANGE AND QUASI-SECTORIAL CONTRACTIONS

YURI ARLINSKII, VALENTIN ZAGREBNOV*

* *Université Aix-Marseille II and Centre de Physique Théorique*

E-mail: zagrebnov@cpt.univ-mrs.fr

We apply a method developed by one of the authors (Yu Arlinskii) , to localize the numerical range of *quasi-sectorial* contractions semigroups. Our main theorem establishes a relation between the numerical range of quasi-sectorial contraction semigroups $\{\exp(-tS)\}_{t \geq 0}$, and the maximal sectorial generators S . We also give a new prove of the rate $O(1/n)$ for the operator-norm Euler formula approximation: $\exp(-tS) = \lim_{n \rightarrow \infty} (I + tS/n)^{-n}$, $t \geq 0$, for this class of semigroups.

POSTERS

BERNHARD BURGSTALLER (Czech Technical University): **G-EQUIVARIANT KK-THEORY FOR SEMIMULTIPLICATIVE SETS G**

PATRICK MOYLAN (Pennsylvania State Univ, Abington College): **POSITIVE ENERGY REPRESENTATIONS AND DEFORMATIONS OF CONFORMAL ALGEBRAS**

PETR NOVOTNÝ (Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University), with JAROSLAV HRIVNÁK: **GENERALIZED DERIVATIONS AND TWISTED COCYCLES OF LIE ALGEBRA**

JUAN NÚÑEZ (University of Seville), with MANUEL CEBALLOS GANZÁLEZ, ÁNGEL FRANCISCO TENORIO VILALLÓN: **MINIMAL LINEAR REPRESENTATION OF FILIFORM LIE ALGEBRAS**

ADAM PASZKIEWICZ (Łódź University), with PIOTR SIELSKI: **ON CONCEPTS OF SUFFICIENCY FOR QUANTUM STATISTICS**

NELSON VIEIRA (University of Aveiro): **L_p -DECOMPOSITION OF THE SPACE OF OF SOLUTIONS OF THE NON-STATIONARY SCHRÖDINGER EQUATION**

Nonrelativistic Quantum Mechanics

(including quantum information)

Session organizers:

Nilanjana Datta
Peter Kuchment

RADIATIONLESS TRANSITIONS THROUGH AVOIDED CROSSINGS

VOLKER BETZ

*Mathematics Institute, University of Warwick,
Coventry CV4 7AL, United Kingdom
E-mail: v.m.betz@warwick.ac.uk*

Understanding the quantum dynamics for transitions through actual or avoided crossings of electronic energy levels is one of the great open problems of quantum chemistry. Here we explain the context of transitions through avoided crossings, and present a novel method based on superadiabatic representations, leading to an explicit formula for a large class of one-dimensional models.

Keywords: Landau-Zener transitions, avoided crossings, superadiabatic representations

1. Radiationless transitions

Radiationless transitions are a peculiar but important reaction mechanism in nature: a photon is absorbed by a molecule, lifting the electronic configuration to an excited state. The nuclei of the molecule are then no longer near their minimal energy configuration, and start to move according to the time-dependent Born-Oppenheimer approximation (see below). At some point along the path of the nuclear configuration, the electronic energies of the ground state and the excited state, both of which are functions of the nuclear configuration, approach each other very closely. It is at this point that the electronic wave function can switch back into the ground state, without emitting a photon. The molecule then falls back into the ground state, or a metastable state, or it dissociates, and the surplus energy is dissipated via vibrations. Radiationless transitions are very fast, they happen on a femtosecond time scale and thus well before any spontaneous decay of the excited state to the ground state would occur. This makes them ideal candidates e.g. for the mechanism of reception of light in the retina, where the re-emission of a photon would be a most unwelcome event. And indeed, it is known that the cis-trans isomerisation of retinal rhodopsin is facilitated by a radiationless transition [1].

While the basic mechanism of radiationless transitions is well-understood, it is one of the great open problems of quantum chemistry to reliably predict the wave function of the molecule after such a transition. The reason is that the Born-Oppenheimer approximation, which is at the basis of almost all methods in quantum molecular dynamics, breaks down at a crossing of electronic energy levels. Thus

straightforward molecular dynamics fails. It can be improved by surface hopping algorithms of various degrees of sophistication [2, 3], but the obtained accuracy is still not quite satisfactory. Or one can use a priori methods that do not rely on the Born-Oppenheimer decoupling of dynamics, like MCTDH [4]. But this is numerically expensive, especially when the transition is much smaller than the original wave function, as is the case for avoided crossings. Here we describe another, analytical method that relies on superadiabatic representations [5, 6]. In the case of a diatomic molecule we obtain an explicit formula for radiationless transitions which agrees well with high precision a priori computations.

2. Adiabatic and superadiabatic representations

The Schrödinger equation of a diatomic molecule is given by

$$i\varepsilon\partial_t\psi(x,t) = H\psi(x,t), \quad (1)$$

with Hamiltonian $H = -\frac{\varepsilon^2}{2}\partial_x^2 + H_{\text{el}}(x)$. $x \in \mathbb{R}$ is the effective nuclear coordinate, and H_{el} is the electronic Hamiltonian, containing the kinetic energy of the electrons, and all interaction potentials. The time scale is chosen such that in a time interval of order one, the nuclei move by a distance of order one. H_{el} and thus its spectrum depend parametrically on x , and one statement of Born-Oppenheimer theory is that if a wave function is of the form $\psi(x,y) = \psi_0(x)\chi(x,y)$, with $y \mapsto \chi(x,y)$ an eigenfunction of $H_{\text{el}}(x)$ for each x with energy $E(x)$, then under the evolution (1) the wave function stays of the above form, to leading order in ε . Moreover, it is given by $\psi(x,t) = \psi_0(x,t)\chi(x,y)$ with $\psi_0(x,t)$ the solution of $i\varepsilon\partial_t\psi_0(x,t) = (-\varepsilon^2\partial_x^2/2 + E(x))\psi_0(x,t)$. In other words, the dynamics decouples according to the energy bands $E(x)$, which themselves act as effective potentials. While the theory is true in great generality [6], here we only need a very special case. Namely, we assume that H_{el} only acts on a two-dimensional Hilbert space, meaning that we assume that all energy bands save two are unimportant for the effect we want to study. While this assumption cannot as yet be rigorously justified, it is one that is almost always made in the context of radiationless transitions. H_{el} then becomes a 2×2 matrix V , and we have

$$H = -\frac{\varepsilon^2}{2}\partial_x^2\mathbf{I} + V(x) \quad \text{with} \quad V(x) = \begin{pmatrix} X(x) & Z(x) \\ Z(x) & -X(x) \end{pmatrix} + d(x)\mathbf{I} \quad (2)$$

$\mathbf{I}$ is the 2×2 unit matrix, and H acts on wave functions $\psi \in L^2(dx, \mathbb{C}^2)$, i.e. square integrable functions with values in $\mathbb{C}^2$. By the non-crossing rule [7], the energy levels generically will not cross for one degree of freedom. Therefore $\rho(x) = \sqrt{X^2 + Z^2} \geq \delta > 0$ for all $x \in \mathbb{R}$, and we assume that X and Z are analytic so that ρ is analytic in a strip containing the real axis.

We now transform (2) to a representation where H becomes approximately diagonal. It is implemented by the unitary transform U acting on $L^2(dx, \mathbb{C}^2)$ by with $Uf(x) = U_0(x)f(x)$, and such that U_0 diagonalizes V . Putting $\psi_a(x,t) =$

$U_0(x)\psi(x, t)$ we obtain the Schrödinger equation in the adiabatic representation,

$$i\varepsilon\partial_t\psi_a(x, t) = H_0\psi_a(x, t), \quad \psi_a(x, 0) = \psi_{a,0}(x), \quad (3)$$

where to leading order in ε ,

$$H_0 = U_0 H U_0^* = -\frac{\varepsilon^2}{2}\partial_x^2 \mathbf{I} + \begin{pmatrix} \rho(x) + d(x) & -\varepsilon\kappa_0(x)(\varepsilon\partial_x) \\ \varepsilon\kappa_0(x)(\varepsilon\partial_x) & -\rho(x) + d(x) \end{pmatrix}. \quad (4)$$

Here, $\kappa_0 = (Z'X - X'Z)/(Z^2 + X^2)$ is the adiabatic coupling element. As a consequence of the choice of time scale solutions of (3) oscillate with frequency $1/\varepsilon$, and thus the operator $\varepsilon\partial_x$ is actually of order one. The dynamics in the two components of ψ_a decouple to leading order in ε , and are Schrödinger evolutions with potentials given by the energy bands $\pm\rho(x)$; this is the Born-Oppenheimer approximation we described above.

A local minimum of ρ is called an avoided crossing. Let us assume that $x = 0$ is a global minimum, and that V is asymptotically constant for large $|x|$. A radiationless transition can then be obtained by starting the Schrödinger evolution with an initial wave packet $\psi_a(x, 0)$ such that its second component $\psi_a^-(x, t)$ is zero, and its first component $\psi_a^+(x, t)$ is localized at some $x_0 \ll 0$, and with sufficient momentum to propagate to positive values of x . $\psi_a^-(x, t)$ is called the transmitted wave packet. Up to errors of order ε^2 , it can be calculated straightforwardly by first order perturbation theory. Let us write $H^\pm = -\frac{\varepsilon^2}{2}\partial_x^2 \pm \rho(x) + d(x)$ for the Hamiltonian generating the uncoupled dynamics in the upper, resp. lower, band. With $\phi(x) = \psi_a^+(x, 0)$, we have

$$\psi_a^-(t) = -i\varepsilon \int_{-\infty}^t e^{-\frac{i}{\varepsilon}(t-s)H^-} K_0 e^{-\frac{i}{\varepsilon}sH^+} \phi ds. \quad (5)$$

Above, $K_0 = \kappa_0(x)(\varepsilon\partial_x)$ is the adiabatic coupling operator. From (5), we would expect radiationless transitions to be of order ε . But while this is true in the transition region, they are much smaller in the scattering regime. Indeed, if one were to do compute (5) numerically, one would find that the L^2 norm of ψ_a^- rises to order ε around $t = 0$, while decaying to an exponentially small value in ε later. This phenomenon is the reason why it is very difficult and time-consuming to obtain accurate quantitative predictions of the transmitted wave function through (5): one has to compute a highly oscillatory integral up to exponential accuracy in ε .

At this point, superadiabatic representations are helpful. They have been developed in the simpler Landau-Zener-Majorana model of adiabatic quantum transitions [8–10], where essentially the nuclear degree of freedom is replaced by a deterministic trajectory, and only electronic transitions are studied. The idea, first found by Berry in [5] and later proven to work rigorously in [11–13], is to find a change of coordinates in which the off-diagonal elements of the potential energy matrix are of order ε^n instead of just order ε . Such a representation is then called the n -th superadiabatic representation; it is asymptotic in n in that the coupling elements are really of order $\varepsilon^n n!$, and thus the limit $n \rightarrow \infty$ cannot be taken. Instead, by

optimizing n one obtains what is called the optimal superadiabatic representation, where transitions are uniformly exponentially small and have the universal shape of an error function.

To construct and asymptotically understand superadiabatic representations in the context of (2) is a very difficult and not yet fully solved problem. However, in the case of sufficiently high momentum of the wave function at the crossing point, it has been done in [14] in the case of $d(x) = 0$, and in [15] in the general case. Above, 'sufficiently high' means in theory at least $\sim \varepsilon^{-1/3}$, but in practice the results described below agree extremely well with high precision numerical solutions for a wide range of parameters. In [14] it is shown that for every $n \in \mathbb{N}$, there exists a unitary U_n , given by a pseudo-differential operator, acting in $L^2(\mathbb{R}, \mathbb{C}^2)$. With $H_n = U_n^{-1} H U_n$, to leading order we have

$$H_n = -\frac{\varepsilon^2}{2} \partial_x^2 \mathbf{I} + \begin{pmatrix} \rho(x) + d(x) & \varepsilon^{n+1} K_{n+1}^+ \\ \varepsilon^{n+1} K_{n+1}^- & -\rho(x) + d(x) \end{pmatrix}. \quad (6)$$

Therefore, with $\psi_n = U_n \psi$, (5) can be replaced by

$$\psi_n^-(t) = -i\varepsilon^n \int_{-\infty}^t e^{-\frac{i}{\varepsilon}(t-s)H^-} K_{n+1}^- e^{-\frac{i}{\varepsilon}sH^+} \phi \, ds. \quad (7)$$

Equation (7) describes the wave function in the n -th superadiabatic representation, which agrees with the adiabatic one in the scattering regime. The *superadiabatic coupling operator* K_n can be obtained by a recursive scheme detailed in [14].

3. The transmitted wave function

It is not immediately clear that (7) is useful for calculating the transmitted wave function: K_n is given by a complicated recursive scheme, and the propagators $H^\pm$ are not explicitly known for general potentials. However, a series of 'good fortunes' allows for significant simplifications. Firstly, by the Darboux principle of asymptotic theory, the operator K_n is determined by the complex singularities of the adiabatic coupling function $\kappa_0(x)$ that are closest to the real axis. We refer to [13] for discussion of the Darboux principle and a rigorous application of it. The transition operator K_n then becomes extremely localized, in that its application on a function differs significantly from zero only if that function has support near $x = 0$. As discussed in [15], this enables us to replace the full propagators $e^{-\frac{i}{\varepsilon}(t-s)H^\pm}$ by the linearized ones in (7). The result is a rather explicit, if somewhat lengthy, formula for the transmitted wave function, see equation (10) of [15]. It simplifies further when $d(x)$ is near constant at the transition point $x = 0$. In that case, and for times $t \gg \sqrt{\varepsilon}$, the transmitted wave packet is given by

$$\widehat{\psi}^\varepsilon(k, t) = e^{-\frac{i}{\varepsilon}tH^-} 1_{\{k^2 > 4\delta\}} \frac{v+k}{2|v|} e^{i\frac{\tau_\delta}{2\delta\varepsilon}|k-v|} \widehat{\phi}^\varepsilon(v), \quad (8)$$

where $v = v(k, \delta) = \text{sgn}(k)\sqrt{k^2 - 4\delta}$, and the constants τ_δ and δ can be calculated from ρ . For a function f , $\widehat{f}^\varepsilon(k) = \frac{1}{\sqrt{2\pi\varepsilon}} \int e^{-\frac{i}{\varepsilon}kq} f(q) \, dq$ denotes its semiclassical Fourier transform.

Formula (8) not only allows to compute the transmitted wave function to high accuracy with very little numerical effort, but also provides physical intuition about the transition mechanism. Namely, the portion of $\widehat{\psi}^\varepsilon$ with momentum k is obtained by choosing the portion of the upper band wave function at the crossing point having the correct momentum $v(k)$ according to energy conservation. This is then multiplied with a Landau-Zener factor that in general cannot be guessed, but for large incoming momenta converges to the one suggested by classical Landau-Zener theory [13]. One consequence of this is that the momentum of the outgoing wave packet is larger than could be expected from energy conservation alone: portions of the incoming wave function that are faster behave less adiabatically, and are more likely to make the transition.

Let us finally note a somewhat embarrassing feature of (8): it is most likely not asymptotically correct in ε , for finite incoming momenta. The reason is that our knowledge about the asymptotics of the operators K_n is not very good, and so we are confined to either large incoming momenta, or relatively large ε . While from a practitioner's point of view, (8) does very well for all cases that can be compared to ab initio calculations with reasonable effort, it is unsatisfactory from the point of view of exponential asymptotics. In [16], an alternative, asymptotically correct, formula is given for the transmitted wave function. But it is not nearly as easy to interpret or implement as (8). One would expect that there is an asymptotically exact formula that retains the general shape of (8), with the constants τ_δ replaced by different ones or possibly non-constant expressions. However, so far no such formula has been found, let alone proved.

References

- [1] R. Schoenlein, L. Petenau, R. Mathies and C. Shank, *Science* **254**, 412 (1991).
- [2] C. Lasser, T. Swart and S. Teufel, *Commun. Math. Sci.* **5**, 789 (2007).
- [3] J. Tully, *J Chem Phys* **93**, 1061 (1990).
- [4] G. A. Worth et al., *Int. Rev. Phys. Chem* **27(3)**, 569 (2008).
- [5] M. Berry, *Proc. Roy. Soc. A* **429**, 61 (1990).
- [6] S. Teufel, *Adiabatic Perturbation Theory in Quantum Dynamics*, Springer LNM 1821 (2003), p. 236.
- [7] J. von Neuman and E. Wigner, *Phys. Z.* **30**, 467 (1929).
- [8] L. Landau, *Physics of the Soviet Union* **2**, 46 (1932).
- [9] C. Zener, *Proc. Roy. Soc. A* **137**, 696 (1932).
- [10] E. Majorana, *Il Nuovo Cimento (1924-1942)* **9**, 43 (1932).
- [11] G. Hagedorn and A. Joye, *Comm. Math. Phys.* **250**, 393 (2004).
- [12] V. Betz and S. Teufel, *Ann. Henri Poincaré* **6**, 217 (2005).
- [13] V. Betz and S. Teufel, *Comm. Math. Phys.* **260**, 481 (2005).
- [14] V. Betz, B. D. Goddard and S. Teufel, *Proc. Roy. Soc. A* **465**, 3553 (2009).
- [15] V. Betz and B. Goddard, *Phys. Rev. Lett.* **103**, 213001 (2009).
- [16] G. A. Hagedorn and A. Joye, *Ann. Henri Poincaré* **6**, 937 (2005).

EQUIVALENCE OF SOBOLEV INEQUALITIES AND LIEB-THIRRING INEQUALITIES*

RUPERT L. FRANK

*Department of Mathematics, Princeton University,
Washington Road, Princeton, NJ 08544, USA
E-mail: rlf Frank@math.princeton.edu*

ELLIOTT H. LIEB

*Departments of Mathematics and Physics, Princeton University,
P. O. Box 708, Princeton, NJ 08542, USA
E-mail: lieb@princeton.edu*

ROBERT SEIRINGER

*Department of Physics, Princeton University, P. O. Box 708, Princeton, NJ 08542, USA
E-mail: rseiring@princeton.edu*

We show that, under very general definitions of a kinetic energy operator T , the Lieb–Thirring inequalities for sums of eigenvalues of $T - V$ can be derived from the Sobolev inequality appropriate to that choice of T .

Keywords: Schrödinger operator, Sobolev inequality, bound states, stability of matter

1. Introduction

The Sobolev and Lieb–Thirring (LT) inequalities seem to be very different. The former is a kind of uncertainty principle, which, effectively, states how large a negative potential $-V$ must be for the Schrödinger operator $H = -\Delta - V$ to have a bound state. The latter, which was originally introduced to prove the stability of matter [22], estimates the sum of all the negative eigenvalues of H and, apparently, is stronger. Our goal here is to summarize some recent work that says, surprisingly, that the latter can, nevertheless, be derived from the former.

This theme extends to other, more complicated operators than $T = -\Delta$. For example, to the barely positive “Hardy” operator $T = -\Delta - (d-2)^2/(4|x|^2)$ for dimensions $d \geq 3$. Another is the relativistic energy $T = \sqrt{-\Delta + m^2} - m$, or just $(-\Delta)^s$. Still another is the inclusion of a magnetic vector potential A and $T = -(\nabla + iA(x))^2$. In all cases there is a Sobolev type inequality $(\psi, T\psi) \geq C \|\psi\|^2$,

*This work was partially supported by U.S. NSF grants PHY-0652854 (R.L.F. and E.H.L.) and PHY-0652356 (R.S.).

where $\|\cdot\|$ is a suitable norm.

The LT inequalities [23] are of the form $\sum_j |\lambda_j|^\gamma \leq C' (\|V\|')^p$, where the λ_j are the negative eigenvalues of $H = T - V$, $\|V\|'$ is another norm, and the allowed range of exponents γ and p depends on the dimension d . A bound in the limiting case $\gamma = 0$, i.e., a bound on the number of bound states of $-\Delta - V$, which is valid for $d \geq 3$, is due to Cwikel [2], Lieb [20] and Rozenblum [27] and is called the CLR bound.

We shall explain, in very general terms, how LT and CLR inequalities can be derived from Sobolev inequalities (the converse being almost trivial). We shall also give several examples for the purpose of clarification. One of the most physically relevant of these is to the relativistic T , which in connection with earlier work [24] yields a proof of the stability of relativistic matter in arbitrary magnetic fields all the way up to the critical value of the allowed nuclear charge $Z = 2/\pi\alpha$, where α is the fine structure constant.

2. Main results

2.1. The setup

We start with an abstract setting, just to show how general the equivalence of Sobolev and LT is. Much of this can be skipped for practical applications.

Let X be a sigma-finite measure space. We consider the measure on X as fixed and denote integration with respect to this measure by dx . By $L^p(X)$ for $1 \leq p \leq \infty$ we denote the usual L^p space with respect to this measure. Moreover, if w is a non-negative function on X we write $L^p(X, w)$ for the L^p space with respect to the measure $w(x)dx$.

Let t be a non-negative quadratic form, with domain $\text{dom } t$, which is closed in the Hilbert space $L^2(X)$ and let T be the corresponding self-adjoint operator.

Throughout this paper we work under the following assumption which depends on a parameter $1 < \kappa < \infty$.

Assumption 2.1 (Generalized Beurling-Deny conditions).

- (1) if $u, v \in \text{dom } t$ are real-valued then $t[u + iv] = t[u] + t[v]$,
- (2) if $u \in \text{dom } t$ is real-valued then $|u| \in \text{dom } t$ and $t[|u|] \leq t[u]$.
- (3) there is a measurable, a.e. positive function ω such that if $u \in \text{dom } t$ is non-negative then $\min(u, \omega) \in \text{dom } t$ and $t[\min(u, \omega)] \leq t[u]$. Moreover, there is a form core $\mathcal{Q}$ of t such that $\omega^{-1}\mathcal{Q}$ is dense in $L^2(X, \omega^{2\kappa/(\kappa-1)})$.

For $\omega \equiv 1$, these are the usual Beurling-Deny conditions; see, e.g., [4, Sec. 1.3]. We note that in this case the assumption is independent of the value of κ . In our applications below it will be important to allow for $\omega \neq 1$. In those examples, X will have a differentiable structure and the density assumption will be satisfied for all κ because ω is sufficiently smooth. We refer to Section 3 for those examples.

2.2. Main results

Our first result concerns upper bounds on the number of negative eigenvalues of Schrödinger operators $T - V$ in terms of integrals of the potential $-V$. We denote by $N(-\tau, T - V)$ the number of eigenvalues less than $-\tau \leq 0$ of the operator $T - V$, taking multiplicities into account, and we abbreviate $N(T - V) := N(0, T - V)$. We shall prove

Theorem 2.1 (Equivalence of Sobolev and CLR Inequalities). *Under Assumption 2.1 for some $\kappa > 1$ the following are equivalent:*

- (1) *T satisfies a Sobolev inequality with exponent $q = 2\kappa/(\kappa - 1)$, that is, there is a constant $S > 0$ such that for all $u \in \text{dom } t$,*

$$t[u] \geq S \left(\int_X |u|^q dx \right)^{2/q}. \quad (1)$$

- (2) *T satisfies a CLR inequality with exponent κ , that is, there is a constant $L > 0$ such that for all $0 \leq V \in L^\kappa(X)$,*

$$N(0, T - V) \leq L \int_X V^\kappa dx. \quad (2)$$

The respective constants are related according to

$$S^{-\kappa} \leq L \leq e^{\kappa-1} S^{-\kappa}. \quad (3)$$

We emphasize that the statement of the theorem does not depend on ω in Assumption 2.1. Only its existence and not its form is relevant.

The implication $(2 \Rightarrow 1)$ is a simple consequence of Hölder's inequality and the variational characterization of the lowest eigenvalue and is valid *without* Assumption 2.1. The converse is much deeper.

Remark 2.1. Below we shall sketch two proofs of $(1 \Rightarrow 2)$ which are abstract versions of proofs by Lieb [20] and by Li and Yau [19]; see also [18] for the case $\omega \equiv 1$. The latter method gives the bound $L \leq e^{\kappa-1} S^{-\kappa}$ stated in (3). The method of [20] proceeds via the bound

$$\|\exp(-sT)\|_{1 \rightarrow \infty} \leq K s^{-\kappa}$$

which follows from the Sobolev inequality (1) with $K \leq (\kappa/S)^\kappa$. (For simplicity we consider the case $\omega \equiv 1$ here.) The method then gives the bound

$$L \leq \frac{K}{\kappa(\kappa-1)} \inf_{a>0} a^{-\kappa+1} e^a \left(1 - a \int_0^\infty e^{-\lambda} (\lambda + a)^{-1} d\lambda \right)^{-1}.$$

Although, by inserting $K \leq (\kappa/S)^\kappa$, this yields a slightly worse bound on L than the one in (3), in concrete applications one often has better bounds on K available. In particular, in the case $T = -\Delta$ in $d = 3$ one has $\kappa = 3/2$ and $K = (4\pi)^{-3/2}$. The upper and lower bounds on L derived this way then differ only by a factor 1.49.

Next, we turn to estimates on eigenvalue moments. Consider two sets of parameters $(\kappa, \gamma) \in (0, \infty) \times (0, \infty)$ and $(q, \theta) \in (2, \infty) \times (0, 1)$ related by

$$\gamma = \frac{q(1-\theta)}{q-2}, \quad \kappa = \frac{q\theta}{q-2}, \quad (4)$$

and

$$q = \frac{2(\gamma + \kappa)}{\gamma + \kappa - 1}, \quad \theta = \frac{\kappa}{\gamma + \kappa}. \quad (5)$$

Theorem 2.2 (Equivalence of Sobolev and weak LT Inequalities). *Let (κ, γ) and (q, θ) be as in (4) and (5) and assume that $\gamma + \kappa > 1$. Under Assumption 2.1 with κ replaced by $\gamma + \kappa$ the following are equivalent:*

- (1) *T satisfies a Sobolev interpolation inequality with exponent q , that is, there is a constant $S > 0$ such that for all $u \in \text{dom } t$,*

$$t[u]^\theta \|u\|^{2(1-\theta)} \geq S \left(\int_X |u|^q dx \right)^{2/q}. \quad (6)$$

- (2) *T satisfies a weak LT inequality with exponent κ , that is, there is a constant $L > 0$ such that for all $0 \leq V \in L^{\gamma+\kappa}(X)$ and all $\tau > 0$,*

$$N(-\tau, T - V) \leq L \tau^{-\gamma} \int_X V^{\gamma+\kappa} dx. \quad (7)$$

The respective sharp constants satisfy

$$(\theta^{-\theta}(1-\theta)^{-1+\theta}S)^{-\gamma-\kappa} \leq L \leq e^{\gamma+\kappa-1}(\theta^{-\theta}(1-\theta)^{-1+\theta}S)^{-\gamma-\kappa}. \quad (8)$$

Corollary 2.1 (LT Inequalities). *Let T satisfy the Sobolev interpolation inequality (6) for some $2 < q < \infty$ and $0 < \theta < 1$ and let Assumption 2.1 hold with κ replaced by $q/(q-2)$. Define $0 < \kappa < \infty$ and $0 < \gamma < \infty$ by (4). Then for all $\tilde{\gamma} > \gamma$ and for all $0 \leq V \in L^{\tilde{\gamma}+\kappa}(X)$ one has*

$$\text{Tr}(T - V)_-^{\tilde{\gamma}} \leq L_{\tilde{\gamma}} \int_X V^{\tilde{\gamma}+\kappa} dx \quad (9)$$

with

$$L_{\tilde{\gamma}} \leq \frac{\tilde{\gamma}^{\tilde{\gamma}+1}}{\gamma\tilde{\gamma}(\tilde{\gamma}-\gamma)^{\tilde{\gamma}-\gamma}} \frac{\Gamma(\gamma+\kappa+1)\Gamma(\tilde{\gamma}-\gamma)}{\Gamma(\tilde{\gamma}+\kappa+1)} L,$$

where L is the sharp constant in (7).

2.3. Inclusion of magnetic fields

The previous analysis can be extended to operators with magnetic fields, which do not satisfy Assumption 2.1. A judicious use of the diamagnetic inequality allows one to reduce the problem to the non-magnetic case. On the level of quadratic forms, the diamagnetic inequality means the following. Let t and t_A denote two closed, non-negative quadratic forms in $L^2(X)$. We say that t_A satisfies a diamagnetic

inequality with respect to t if for any $u \in \text{dom } t_A$ and $v \in \text{dom } t$ with $0 \leq v \leq |u|$ one has $|u| \in \text{dom } t$, $v \operatorname{sgn} u \in \text{dom } t_A$ and

$$t[v, |u|] \leq \operatorname{Re} t_A[v \operatorname{sgn} u, u]. \quad (10)$$

Here we use the definition $\operatorname{sgn} u(x) := u(x)/|u(x)|$ if $u(x) \neq 0$ and $\operatorname{sgn} u(x) := 0$ if $u(x) = 0$. Moreover, $t[\cdot, \cdot]$ denotes the sesqui-linear form associated to the quadratic form $t[\cdot]$ which is anti-linear in the first and linear in the second argument, and likewise for t_A .

Let T_A denote the operator corresponding to t_A .

Theorem 2.3 (Inequalities with magnetic fields). *Assume that t satisfies Assumption 2.1 as well as either (1) or (6) and assume that t_A satisfies a diamagnetic inequality with respect to t . Then the number of negative eigenvalues of $T_A - V$ satisfies the bounds (2) or (7) with the same upper bounds on the constants L as in (3) or (8).*

Corollary 2.1 has a similar extension to the magnetic case as well. For further results about the magnetic version of CLR and LT inequalities we refer to [8, 28].

3. Illustrative examples

3.1. The Laplacian

Let $T = -\Delta$ in $L^2(\mathbb{R}^d)$ corresponding to the quadratic form $t[u] := \int_{\mathbb{R}^d} |\nabla u|^2 dx$ with domain $H^1(\mathbb{R}^d)$. One easily checks that Assumption 2.1 is satisfied for $\omega \equiv 1$. Moreover, if $2 < q < \infty$ for $d = 1, 2$, or if $2 < q \leq 2d/(d-2)$ for $d \geq 3$ the following Sobolev inequalities

$$\left(\int_{\mathbb{R}^d} |\nabla u|^2 dx \right)^\theta \left(\int_{\mathbb{R}^d} |u|^2 dx \right)^{1-\theta} \geq S_{q,d} \left(\int_{\mathbb{R}^d} |u|^q dx \right)^{2/q} \quad (11)$$

are well-known; see [21, Sects. 8.3 and 8.5]. Here $\theta = d(\frac{1}{2} - \frac{1}{q})$. By Theorem 2.1 and Corollary 2.1 this implies the usual LT inequalities

$$\operatorname{Tr}(-\Delta - V)_-^\gamma \leq L_{\gamma,d} \int_{\mathbb{R}^d} V_+^{\gamma+d/2} dx \quad (12)$$

for $\gamma > (2-d)/2$ if $d = 1, 2$ and $\gamma \geq 0$ if $d \geq 3$. We note that the inequality in the case $\gamma = 1/2$, $d = 1$, though being valid [14, 33], cannot be obtained using the approach of the present paper. For a review of this topic, and remarks about the best constants, see [15, 17].

If $A \in L_{\text{loc}}^2(\mathbb{R}^d; \mathbb{R}^d)$ and $t_A[u] := \int_{\mathbb{R}^d} |(\nabla + iA)u|^2 dx$, then (10) holds [16, 31]. Hence by Theorem 2.3, inequality (12) holds with $-\Delta$ replaced by $-(\nabla + iA)^2$.

3.2. Fractional Laplacians

Let f be a non-negative, differentiable function on $(0, \infty)$ such that f' is completely monotone. We claim that the operator $T = f(-\Delta)$ satisfies Assumption 2.1 with

$\omega \equiv 1$. To verify this, we have to check by the Beurling–Deny theorem [4, Sec. 1.3] that $\exp(-tf(-\Delta))$ is positivity preserving and contractive in $L^\infty(\mathbb{R}^d)$. One easily checks that $E \mapsto e^{-tf(E)}$ is completely monotone and therefore by Bernstein’s theorem [5, Sec. I.5] it is of the form $e^{-tf(E)} = \int_0^\infty e^{-sE} d\mu_{f,t}(s)$ for some non-negative measure $\mu_{f,t}$. Hence $\exp(-tf(-\Delta)) = \int_0^\infty \exp(s\Delta) d\mu_{f,t}(s)$. Since $\exp(s\Delta)$ is positivity preserving and contractive in $L^\infty(\mathbb{R}^d)$ for any $s > 0$, so is $\exp(-tf(-\Delta))$. (For the contraction property we also use that $\int_0^\infty d\mu_{f,t}(s) = e^{-tf(0)} \leq 1$.)

In particular, the function $f(E) = E^s$ for $0 < s < 1$ is of the form described above. In this case, one has the Sobolev inequalities

$$\left\| (-\Delta)^{s/2} u \right\|^{2\theta} \left(\int_{\mathbb{R}^d} |u|^2 dx \right)^{1-\theta} \geq S_{d,s,q} \left(\int_{\mathbb{R}^d} |u|^q dx \right)^{2/q}, \quad \theta = \frac{d}{2s} \left(\frac{1}{2} - \frac{1}{q} \right),$$

for $2 < q < \infty$ if $d \leq 2s$ and for $2 < q \leq 2d/(d-2s)$ if $d > 2s$. Hence Theorem 2.1 and Corollary 2.1 yield the inequalities

$$\mathrm{Tr} \left((-\Delta)^s - V \right)_-^\gamma \leq L_{\gamma,s,d} \int_{\mathbb{R}^d} V_+^{\gamma+d/2s} dx \quad (13)$$

for $\gamma > (2s-d)/2s$ if $d \leq 2s$ and $\gamma \geq 0$ if $d > 2s$. These inequalities appeared first in [3].

3.3. Periodic Schrödinger operators

Let W be a $\mathbb{Z}^d$ -periodic function on $\mathbb{R}^d$ and consider the Schrödinger operator $-\Delta + W$ in $L^2(\mathbb{R}^d)$. With $E := \inf \mathrm{spec}(-\Delta + W)$ the quadratic form $t[u] := \int_{\mathbb{R}^d} (|\nabla u|^2 + (W-E)|u|^2) dx$ is non-negative. Under very weak conditions on W there is a periodic function ω satisfying $-\Delta\omega + W\omega = E\omega$, and ω is bounded (by elliptic regularity) and strictly positive (by Harnack’s inequality). The representation

$$t[u] = \int_{\mathbb{R}^d} |\nabla v|^2 \omega^2 dx, \quad u = \omega v,$$

together with the properties of ω implies that Assumption 2.1 is satisfied for any $\kappa > 1$. Moreover, by the same representation the Sobolev inequalities (11) hold with the constant $S_{q,d}$ replaced by $(\inf \omega / \sup \omega)^2 S_{q,d}$. Therefore Theorem 2.1 and Corollary 2.1 yield the CLR and LT inequalities

$$\mathrm{Tr} (-\Delta + W - V - E)_-^\gamma \leq L_{\gamma,d}(W) \int_{\mathbb{R}^d} V_+^{\gamma+d/2} dx$$

for the same values of γ as in Subsection 3.1. This was first shown in [11] by a different argument (which includes the case $\gamma = 1/2$ and $d = 1$).

3.4. Hardy–Lieb–Thirring inequalities

Let $d \geq 1$ and $0 < s \leq 1$ such that $d > 2s$, and denote by $\mathcal{C}_{s,d}$ the sharp constant in the Hardy inequality

$$\int_{\mathbb{R}^d} |\xi|^{2s} |\hat{u}(\xi)|^2 d\xi \geq \mathcal{C}_{s,d} \int_{\mathbb{R}^d} |x|^{-2s} |u(x)|^2 dx, \quad u \in C_0^\infty(\mathbb{R}^d).$$

Here $\hat{u}(\xi) = (2\pi)^{-d/2} \int e^{-i\xi \cdot x} u(x) dx$ is the Fourier transform of u . Explicitly (see [12]), one has

$$\mathcal{C}_{s,d} = 2^{2s} \frac{\Gamma((d+2s)/4)^2}{\Gamma((d-2s)/4)^2}.$$

Let t be the closure in $L^2(\mathbb{R}^d)$ of the non-negative quadratic form $\int_{\mathbb{R}^d} |\xi|^{2s} |\hat{u}(\xi)|^2 d\xi - \mathcal{C}_{s,d} \int_{\mathbb{R}^d} |x|^{-2s} |u(x)|^2 dx$ defined on $C_0^\infty(\mathbb{R}^d)$. In [10] we have derived the following ground state representation formula,

$$t[u] = a_{s,d} \iint_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|v(x) - v(y)|^2}{|x - y|^{d+2s}} \frac{dx}{|x|^{(d-2s)/2}} \frac{dy}{|y|^{(d-2s)/2}}, \quad u = \omega v,$$

where $\omega(x) = |x|^{-(d-2s)/2}$ and $a_{s,d}$ is a positive constant. This formula together with the fact that $C_0^\infty(\mathbb{R}^d \setminus \{0\})$ is a form core shows that Assumption 2.1 holds. Moreover, in [10] we have shown that for any $2 < q < 2d/(d-2s)$ there is an $\tilde{S}_{d,s,q} > 0$ such that

$$t[u]^\theta \left(\int_{\mathbb{R}^d} |u|^2 dx \right)^{1-\theta} \geq \tilde{S}_{d,s,q} \left(\int_{\mathbb{R}^d} |u|^q dx \right)^{2/q}, \quad \theta = \frac{d}{2s} \left(\frac{1}{2} - \frac{1}{q} \right).$$

In view of Corollary 2.1 we obtain

$$\mathrm{Tr} \left((-\Delta)^s - \mathcal{C}_{s,d} |x|^{-2s} - V \right)_-^\gamma \leq \tilde{L}_{\gamma,d,s} \int_{\mathbb{R}^d} V_+^{\gamma+d/2s} dx \quad (14)$$

for all $\gamma > 0$ and the values of s indicated above. This inequality for $s = 1$ was first proved in [6]. The proof sketched above is taken from [10]. For an alternative proof covering the cases $d \geq 3$ and $1 < s < d/2$ we refer to [7].

Using Theorem 2.3 one can show that inequality (14) holds also in the magnetic case, that is, with $(-\Delta)^s$ replaced by $|\nabla + iA|^{2s}$ for some $A \in L_{\mathrm{loc}}^2(\mathbb{R}^d; \mathbb{R}^d)$. This fact allowed us to prove stability of relativistic matter in magnetic fields up to the critical value of the nuclear charge; see [10] and also [9, 22].

4. Proofs of the main results

4.1. Proof of Theorem 2.1 for $\omega \equiv 1$

We begin by proving the easy implication ($2 \Rightarrow 1$). The CLR inequality and the variational principle imply that if $\int_X V^\kappa dx < L^{-1}$, then $t[u] \geq \int V |u|^2 dx$ for all $u \in \mathrm{dom} t$. Choosing $V = \alpha_\varepsilon |u|^{2/(\kappa-1)}$ with

$$\alpha_\varepsilon = (1 - \varepsilon) \left(L \int_X |u|^{2\kappa/(\kappa-1)} dx \right)^{-1/\kappa}$$

and letting $\varepsilon \rightarrow 0$ we obtain (1) with $S \geq L^{-1/\kappa}$.

Next, we shall present two proofs of the implication ($1 \Rightarrow 2$). The first one is an abstraction of the semi-group proof of the CLR inequality in [20]. It relies on the heat kernel bound

$$\|\exp(-sT)\|_{1 \rightarrow \infty} \leq K s^{-\kappa}. \quad (15)$$

We recall that by Varopoulos' theorem [32] (1) is equivalent to (15). An abstract version of an argument by Nash [25] allows us to derive (15) from (1) with constant

$$K \leq (\kappa/S)^\kappa. \quad (16)$$

Indeed, the Sobolev inequality (1) implies via Hölder the Nash inequality

$$t[u]^{\frac{q}{2(q-1)}} \left(\int_X |u| dx \right)^{\frac{q-2}{q-1}} \geq S^{\frac{q}{2(q-1)}} \int_X |u|^2 dx. \quad (17)$$

By Nash's argument (see [4, Sec. 2.4] or [21, Thm. 8.16]), using the contraction property in L^1 , this implies

$$\|\exp(-sT)\|_{1 \rightarrow 2}^2 \leq (\kappa/2S)^\kappa s^{-\kappa}, \quad (18)$$

which yields (15) and (16) by duality and the semi-group property.

With (15) at hand we can now follow the arguments in [20], replacing path integrals by Trotter's product formula (see also [29]). Defining for any non-negative, lower semi-continuous function f on $\mathbb{R}_+$ with $f(0) = 0$

$$F(\lambda) := \int_0^\infty f(\mu) e^{-\mu/\lambda} \mu^{-1} d\mu, \quad \lambda > 0, \quad (19)$$

one has the trace formula

$$\begin{aligned} \text{Tr } F(V^{1/2} T^{-1} V^{1/2}) &= \int_0^\infty \frac{ds}{s} \lim_{n \rightarrow \infty} \int_X \cdots \int_X dx_1 \cdots dx_n \\ &\quad \times \prod_{j=1}^n k\left(x_j, x_{j-1}, \frac{s}{n}\right) f\left(\frac{s}{n} \sum_{k=1}^n V(x_k)\right) \end{aligned} \quad (20)$$

with the convention that $x_0 = x_n$ and $k(x, y, s) = \exp(-sT)(x, y)$. By Assumption 2.1 and the Beurling–Deny theorem (see, e.g., [4, Sec. 1.3]) k is non-negative. If, in addition, f is convex then we can bound $f\left(\frac{s}{n} \sum V(x_k)\right) \leq \frac{1}{n} \sum f(sV(x_k))$ and obtain, using the semi-group property,

$$\text{Tr } F(V^{1/2} T^{-1} V^{1/2}) \leq \int_0^\infty \frac{ds}{s} \int_X dx k(x, x, s) f(sV(x)).$$

Now the heat kernel decay (15) implies

$$\text{Tr } F(V^{1/2} T^{-1} V^{1/2}) \leq K \int_0^\infty \frac{ds}{s^{\kappa+1}} \int_X dx f(sV(x)) = K \int_X V^\kappa dx \int_0^\infty f(\mu) \frac{d\mu}{\mu^{\kappa+1}}.$$

By the Birman-Schwinger principle, $N(0, T - V)$ coincides with the number of eigenvalues larger than one of the operator $V^{1/2} T^{-1} V^{1/2}$. Hence, since F is increasing,

$$N(0, T - V) \leq F(1)^{-1} \text{Tr } F(V^{1/2} T^{-1} V^{1/2}) \leq K \int_X V^\kappa dx F(1)^{-1} \int_0^\infty f(\mu) \frac{d\mu}{\mu^{\kappa+1}},$$

and the sought bound follows by choosing $f(\mu) = (\mu - a)_+$ and optimizing over $a > 0$.

The only place where part (3) of Assumption 2.1 entered in the proof is to obtain the heat kernel bound (15) from the Sobolev inequality (1). This part of

the assumption can thus be omitted if one is able to obtain such a bound by other means.

The second proof of the implication $(1 \Rightarrow 2)$ is an abstraction of Li and Yau's proof [19] of the CLR inequality and its improvement in [1]. By an approximation argument we may assume that $V \in L^1 \cap L^\infty$ and $V > 0$ a.e. Moreover, for the sake of simplicity we assume that the embedding of the completion of $\text{dom } t$ with respect to t into $L^q(X)$ is injective; see [18] for an additional argument in the general case. We consider the non-negative operator Υ in $L^2(X, V)$ given by the quadratic form $t[v]$. We shall prove that

$$\text{Tr} (2\Upsilon)^{-1} \exp(-2s\Upsilon) \leq (\kappa - 1)^{\kappa-1} (2S)^{-\kappa} \int_X V^\kappa dx s^{-\kappa+1}. \quad (21)$$

Since Υ in $L^2(X, V)$ is unitarily equivalent to the inverse of the Birman–Schwinger operator $V^{1/2}T^{-1}V^{1/2}$ in $L^2(X)$ one has $N(T - V) = N(1, \Upsilon^{-1})$. The inequality $N(1, \Upsilon^{-1}) \leq 2e^{2s} \text{Tr} (2\Upsilon)^{-1} \exp(-2s\Upsilon)$, together with (21) and optimization in s , will then imply (2) with the upper bound on L stated in (3).

In order to prove (21) we consider the operators $H_\beta(s) = (2\Upsilon)^{-\beta} \exp(-s\Upsilon)$ for $\beta \geq 0$. From the Sobolev inequality (1) and Assumption 2.1 one concludes, as in [18], that $H_0(s)$, and hence also $H_\beta(s)$, are integral operators with non-negative kernels $H_\beta(x, y, s)$. We abbreviate $h_\beta(s) := \text{Tr } H_\beta(2s)$ and estimate, using Hölder with $\frac{1}{q} + \frac{q-2}{q} + \frac{1}{q} = 1$,

$$\begin{aligned} h_\beta(s) &= \int_X dx V(x) \left(\int_X dy V(y) H_\beta(y, x, s) H_0(x, y, s) \right) \\ &\leq \int_X dx V(x) \left(\int_X dy H_\beta(y, x, s)^q \right)^{\frac{1}{q}} \left(\int_X dy H_0(x, y, s) V(y)^{\frac{q-1}{q-2}} \right)^{\frac{q-2}{q}} \\ &\quad \times \left(\int_X dy H_0(x, y, s)^2 V(y) \right)^{\frac{1}{q}}. \end{aligned}$$

Using Hölder once more with $\frac{1}{2} + \frac{q-2}{2q} + \frac{1}{q} = 1$ we obtain

$$\begin{aligned} h_\beta(t) &\leq \left(\int_X dx V(x) \left(\int_X dy H_\beta(y, x, s)^q \right)^{\frac{2}{q}} \right)^{\frac{1}{2}} \\ &\quad \times \left(\int_X dx V(x) \left(\int_X dy H_0(x, y, s) V(y)^{\frac{q-1}{q-2}} \right)^2 \right)^{\frac{q-2}{2q}} \\ &\quad \times \left(\int_X dx V(x) \int_X dy H_0(x, y, s)^2 V(y) \right)^{\frac{1}{q}} =: A^{\frac{1}{2}} B^{\frac{q-2}{2q}} C^{\frac{1}{q}}. \end{aligned}$$

We estimate A by the Sobolev inequality (1),

$$A \leq S^{-1} \int_X dx V(x) t[H_\beta(\cdot, x, s)] = (2S)^{-1} h_{2\beta-1}(s).$$

The contraction property of $\exp(-s\Upsilon)$ in $L^2(X, V)$ implies

$$B = \|\exp(-s\Upsilon)V^{\frac{1}{q-2}}\|_{L^2(X,V)}^2 \leq \|V^{\frac{1}{q-2}}\|_{L^2(X,V)}^2 = \int_X V^\kappa dx.$$

Moreover, $C = h_0(s)$. Hence, choosing $\beta = 1$ and using $h_0(s) = -h'_1(s)$, we have shown

$$h_1(s) \leq (2S)^{-1} \left(\int_X V^\kappa dx \right)^{\frac{1}{\kappa}} (-h'_1(s))^{\frac{2}{q}}$$

which implies (21) and completes the sketch of the proof.

Note that the only place where part (3) of Assumption 2.1 entered in the second proof is the existence of integral kernels for the operators $H_\beta(s)$. Hence this part of the assumption can, in principle, be omitted if this property can be shown by other means.

4.2. Proof of Theorem 2.1 for arbitrary ω

Since the proof of the implication $(2 \Rightarrow 1)$ in the previous subsection did not use Assumption 2.1 we are left with proving $(1 \Rightarrow 2)$. We will deduce this from the case $\omega \equiv 1$. We may assume that T is positive definite for otherwise we consider $T + \varepsilon$ and let $\varepsilon \rightarrow 0$ in the inequality obtained. The quadratic form $t_\omega[v] := t[\omega v]$ with domain $\omega^{-1}\mathcal{D}$ is closable in the Hilbert space $L^2(X, d\mu)$ with measure $d\mu := \omega^{2\kappa/(\kappa-1)}dx$. (Here we use that T is positive definite and that t is closed.) Let T_ω be the corresponding self-adjoint operator in $L^2(X, d\mu)$. We note that t_ω satisfies Assumption 2.1 with $\omega \equiv 1$ (it suffices to verify this assumption on a form core, see [4, Lem. 1.3.4]) and that the Sobolev inequality (1) for t can be written as

$$t_\omega[v] \geq S \left(\int_X |v|^q d\mu \right)^{2/q}.$$

Moreover, by the variational principle,

$$\begin{aligned} N(T - V) &= \sup \left\{ \dim M : M \subset \mathcal{D}, t[u] < \int_X V|u|^2 dx \text{ for all } 0 \neq u \in M \right\} \\ &= \sup \left\{ \dim \tilde{M} : \tilde{M} \subset \omega^{-1}\mathcal{D}, t_\omega[v] < \int_X \tilde{V}|v|^2 d\mu \text{ for all } 0 \neq v \in \tilde{M} \right\} \\ &= N(T_\omega - \tilde{V}) \end{aligned}$$

where $\tilde{V} := \omega^{-2/(\kappa-1)}V$. Since

$$\int_X \tilde{V}^\kappa d\mu = \int_X V^\kappa dx,$$

the assertion follows from the $\omega \equiv 1$ case of Theorem 2.1.

4.3. Proof of Theorem 2.2

We shall deduce the result of Theorem 2.2 for positive γ from that of Theorem 2.1 for $\gamma = 0$. To do so, we consider the operator $T_\tau := \tau^{-1+\theta}(T + \tau)$ and its quadratic form t_τ . Then condition (7) is equivalent to

$$N(T_\tau - V) \leq L \int_X V^{\gamma+\kappa} dx, \quad \tau > 0,$$

for all $0 \leq V \in L^{\gamma+\kappa}(X)$. Moreover, using that for $\alpha, \beta > 0$

$$\min_{\tau > 0} (\alpha \tau^{-1+\theta} + \beta \tau^\theta) = \theta^{-\theta} (1 - \theta)^{-1+\theta} \alpha^\theta \beta^{1-\theta},$$

condition (6) is equivalent to

$$t_\tau[u] \geq \theta^{-\theta} (1 - \theta)^{-1+\theta} S \left(\int_X |u|^q dx \right)^{2/q}, \quad \tau > 0.$$

Noting that T_τ satisfies Assumption 2.1, the assertion follows from Theorem 2.1.

4.4. Proof of Corollary 2.1

By Theorem 2.2 the Sobolev interpolation inequality (6) implies the weak LT inequality (7). We shall now use an interpolation argument from [23] in order to deduce the strong LT inequality for $\tilde{\gamma} > \gamma$ from a weak LT inequality for γ . For any fixed $0 < s < 1$ the variational principle implies

$$N(-\tau, T - V) \leq N(-(1-s)\tau, T - (V - s\tau)_+).$$

Hence the representation

$$\mathrm{Tr} (T - V)_-^{\tilde{\gamma}} = \tilde{\gamma} \int_0^\infty N(-\tau, T - V) \tau^{\tilde{\gamma}-1} d\tau$$

together with the weak LT inequality implies that

$$\begin{aligned} \mathrm{Tr} (T - V)_-^{\tilde{\gamma}} &\leq L \tilde{\gamma} (1-s)^{-\gamma} \int_0^\infty \int_X (V - s\tau)_+^{\gamma+\kappa} dx \tau^{\tilde{\gamma}-\gamma-1} d\tau \\ &= L \tilde{\gamma} (1-s)^{-\gamma} s^{-\tilde{\gamma}+\gamma} B(\gamma + \kappa + 1, \tilde{\gamma} - \gamma) \int_X V^{\tilde{\gamma}+\kappa} dx, \end{aligned}$$

with $B(\cdot, \cdot)$ the beta function. Minimizing in $s \in (0, 1)$ yields the claimed inequality with constant

$$L_{\tilde{\gamma}} \leq \frac{\tilde{\gamma}^{\tilde{\gamma}+1}}{\gamma^{\gamma}(\tilde{\gamma} - \gamma)^{\tilde{\gamma}-\gamma}} \frac{\Gamma(\gamma + \kappa + 1)\Gamma(\tilde{\gamma} - \gamma)}{\Gamma(\tilde{\gamma} + \kappa + 1)} L.$$

4.5. Proof of Theorem 2.3

From the diamagnetic inequality (10) for the quadratic forms one concludes that the heat kernel for the operator T_A is, in absolute value, pointwise bounded from above by the heat kernel for T . This was proved by Hess, Schrader, Uhlenbrock [13] and Simon [30]; see [26, Sec. 2] for a quadratic form version of this result. For $\omega \equiv 1$ this immediately implies that the first proof of $(1 \Rightarrow 2)$ in Theorem 2.1, using the method in [20], extends to the magnetic case with the same bound on the constant. For general ω one proceeds as in Subsection 4.3, noting that $t_{A,\omega}$ satisfies a diamagnetic inequality in the sense of (10) with respect to t_ω .

A similar argument shows that the operator Υ_A in the second proof of $(1 \Rightarrow 2)$ in Theorem 2.1, using the method in [19], satisfies a diamagnetic inequality with respect to Υ . Hence

$$\mathrm{Tr} (2\Upsilon_A)^{-1} \exp(-2t\Upsilon_A) \leq \mathrm{Tr} (2\Upsilon)^{-1} \exp(-2t\Upsilon).$$

Hence (21) leads to the same estimate in the magnetic case as in the non-magnetic case.

References

- [1] Ph. Blanchard, J. Rezende, J. Stubbe, *New estimates on the number of bound states of Schrödinger operators*. Lett. Math. Phys. **14** (1987), no. 3, 215–225.
- [2] M. Cwikel, *Weak type estimates for singular values and the number of bound states of Schrödinger operators*. Ann. Math. **106** (1977), 93–102.
- [3] I. Daubechies, *An uncertainty principle for fermions with generalized kinetic energy*. Comm. Math. Phys. **90** (1983), no. 4, 511–520.
- [4] E. B. Davies, *Heat kernels and spectral theory*. Cambridge Tracts in Mathematics **92**, Cambridge Univ. Press, Cambridge, 1990.
- [5] W. F. Donoghue, *Monotone matrix functions and analytic continuation*. Die Grundlehren der mathematischen Wissenschaften **207**, Springer, New York - Heidelberg, 1974.
- [6] T. Ekholm, R. L. Frank, *On Lieb-Thirring inequalities for Schrödinger operators with virtual level*. Comm. Math. Phys. **264** (2006), no. 3, 725–740.
- [7] R. L. Frank, *A simple proof of Hardy-Lieb-Thirring inequalities*. Comm. Math. Phys. **290** (2009), no. 2, 789–800.
- [8] R. L. Frank, *Remarks on eigenvalue estimates and semigroup domination*. In: Spectral and Scattering Theory for Quantum Magnetic Systems, P. Briet, et al. (eds.), 63–86, Contemp. Math. **500**, Amer. Math. Soc., Providence, RI, 2009.
- [9] R. L. Frank, E. H. Lieb, R. Seiringer, *Stability of relativistic matter with magnetic fields for nuclear charges up to the critical value*. Comm. Math. Phys. **275** (2007), no. 2, 479–489.
- [10] R. L. Frank, E. H. Lieb, R. Seiringer, *Hardy-Lieb-Thirring inequalities for fractional Schrödinger operators*. J. Amer. Math. Soc. **21** (2008), no. 4, 925–950.
- [11] R. L. Frank, B. Simon, T. Weidl, *Eigenvalue bounds for perturbations of Schrödinger operators and Jacobi matrices with regular ground states*. Comm. Math. Phys. **282** (2008), no. 1, 199–208.
- [12] I. W. Herbst, *Spectral theory of the operator $(p^2 + m^2)^{1/2} - Ze^2/r$* . Comm. Math. Phys. **53** (1977), no. 3, 285–294.

- [13] H. Hess, R. Schrader, D. A. Uhlenbrock, *Domination of semigroups and generalization of Kato's inequality*. Duke Math. J. **44** (1977), no. 4, 893–904.
- [14] D. Hundertmark, E. H. Lieb, L. E. Thomas, *A sharp bound for an eigenvalue moment of the one-dimensional Schrödinger operator*. Adv. Theor. Math. Phys. **2** (1998), no. 4, 719–731.
- [15] D. Hundertmark, *Some bound state problems in quantum mechanics*. In: Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon's 60th birthday, 463–496, Proc. Sympos. Pure Math. **76**, Amer. Math. Soc., Providence, RI, 2007.
- [16] T. Kato, *Schrödinger operators with singular potentials*. Israel J. Math. **13** (1972), 135–148.
- [17] A. Laptev, T. Weidl, *Recent results on Lieb–Thirring inequalities*. Journées “Équations aux Dérivées Partielles” (La Chapelle sur Erdre, 2000), Exp. No. XX, Univ. Nantes, Nantes, 2000.
- [18] D. Levin, M. Solomyak, *The Rozenblum–Lieb–Cwikel inequality for Markov generators*. J. Anal. Math. **71** (1997), 173–193.
- [19] P. Li, S.-T. Yau, *On the Schrödinger equation and the eigenvalue problem*. Comm. Math. Phys. **88** (1983), 309–318.
- [20] E. H. Lieb, *Bounds on the eigenvalues of the Laplace and Schrödinger operators*, Bull. Amer. Math. Soc. **82** (1976), 751–752. *The number of bound states of one body Schrödinger operators and the Weyl problem*. Proc. A.M.S. Symp. Pure Math. **36** (1980), 241–252.
- [21] E. H. Lieb, M. Loss, *Analysis. Second edition*. Graduate Studies in Mathematics **14**, American Mathematical Society, Providence, RI, 2001.
- [22] E. H. Lieb, R. Seiringer, *The Stability of Matter in Quantum Mechanics*. Cambridge Univ. Press, Cambridge, 2010.
- [23] E. H. Lieb, W. Thirring, *Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities*. Studies in Mathematical Physics, 269–303. Princeton Univ. Press, Princeton, NJ, 1976.
- [24] E. H. Lieb, H.-T. Yau, *The stability and instability of relativistic matter*. Comm. Math. Phys. **118** (1988), 177–213.
- [25] J. Nash, *Continuity of solutions of parabolic and elliptic equations*. Amer. J. Math. **80** (1958), 931–954.
- [26] E. M. Ouhabaz, *Analysis of heat equations on domains*. Princeton Univ. Press, Princeton, NJ, 2005.
- [27] G. V. Rozenblum, *Distribution of the discrete spectrum of singular differential operators*. Soviet Math. Dokl. **13** (1972), 245–249, and Soviet Math. (Iz. VUZ) **20** (1976), 63–71.
- [28] G. V. Rozenblum, *Domination of semigroups and estimates for eigenvalues*. St. Petersburg Math. J. **12** (2001), no. 5, 831–845.
- [29] G. Rozenblyum, M. Solomyak, *The Cwikel–Lieb–Rozenblyum estimator for generators of positive semigroups and semigroups dominated by positive semigroups*. St. Petersburg Math. J. **9** (1998), no. 6, 1195–1211.
- [30] B. Simon, *Kato's inequality and the comparison of semigroups*. J. Funct. Anal. **32** (1979), no. 1, 97–101.
- [31] B. Simon, *Maximal and minimal Schrödinger forms*. J. Operator Theory **1** (1979), no. 1, 37–47.
- [32] N. Th. Varopoulos, *Hardy–Littlewood theory for semigroups*. J. Funct. Anal. **63** (1985), 240–260.
- [33] T. Weidl, *On the Lieb–Thirring constants $L_{\gamma,1}$ for $\gamma \geq 1/2$* . Comm. Math. Phys. **178** (1996), no. 1, 135–146.

ENTANGLEMENT SPREAD AND CLEAN RESOURCE INEQUALITIES

ARAM W. HARROW

*Department of Mathematics, University of Bristol, Bristol, BS8 1TW, U.K.
E-mail: a.harrow@bris.ac.uk*

This talk will examine states that superpose different amounts of entanglement and protocols that generate or consume different amounts of entanglement, but run in superposition. In both cases we find a uniquely quantum difficulty: entanglement cannot be conditionally discarded without either using communication or causing decoherence. This problem can be addressed by introducing entanglement spread as an information-theoretic resource that can be manipulated using clean resource inequalities.

Keywords: quantum information; entanglement spread; resource inequality

1. Introduction

This talk will challenge the idea that, when it comes to entanglement, more is always better. While some resources in quantum information theory, like use of communication channels, can be safely discarded, entanglement cannot be kept in one branch of a superposition and discarded in another without causing decoherence. We begin by outlining two specific challenges that this fact poses to the traditional resource model of quantum information.

1.1. *Motivation: Coherent conditional execution of quantum communication protocols.*

A productive way of understanding quantum communication protocols can be to view quantum states and operations as *resources* and protocols that convert one resource into another as *resource inequalities* (RIs). So if $[q \rightarrow q]$ represents a noiseless qubit channel from Alice to Bob, $[qq]$ is an EPR pair and $[c \rightarrow c]$ is a noiseless classical bit channel (cbit) from Alice to Bob, then teleportation can be expressed as $2[c \rightarrow c] + [qq] \geq [q \rightarrow q]$. This picture was formalized in [1], which also proved many basic intuitive facts about RIs. For example, if $\alpha \geq \beta$ and $\beta \geq \gamma$ for some resources α, β, γ then $\alpha \geq \gamma$ as well.

However, there is another basic way of combining protocols which works trivially in the classical case but fails in the quantum case. Suppose Alice and Bob each know a common bit b and want to perform protocol $\mathcal{N}_b$ conditioned on the value of b . This may happen when communication protocols are embedded in larger applications

where the amount and type of communication is itself input-dependent. We call this conditional operation $b? \mathcal{N}_1 : \mathcal{N}_0$, following C notation. Classically it is trivial to show that if $\alpha \geq \langle \mathcal{N}_b \rangle$ for $b = 0, 1$ then $\alpha \geq \langle b? \mathcal{N}_1 : \mathcal{N}_0 \rangle$: Alice and Bob use α to perform either $\mathcal{N}_0$ or $\mathcal{N}_1$ depending on their shared value of b . However, in a quantum protocol this might leak information about b to the environment, because the environment may be able to distinguish $\mathcal{N}_0$ and $\mathcal{N}_1$ on some inputs. In Section 2.1 we will describe one possible solution to this problem, which will be called *clean resource inequalities*. The main idea will be to discard only (up to an asymptotically vanishing error) qubits in standard states, such as $|0\rangle$, so that the environment cannot learn which protocol is being run.

1.2. Motivation: Non-asymptotic analysis of entanglement

A second problem with the traditional resource framework arises in quantifying pure state entanglement. The entanglement of a pure state $|\psi\rangle^{AB}$ is usually said to be characterized by its *entropy of entanglement*, $E(\psi)$, which is defined as $E(\psi) := S(\psi^A) = S(\psi^B)$. Here $\psi := |\psi\rangle\langle\psi|$ is the density matrix corresponding to ψ , $\psi^A := \text{tr}_B \psi$ is Alice's reduced density matrix (and similarly for ψ^B), $S(\rho) := -\text{tr} \rho \log \rho$ is the von Neumann entropy, and the base of logs and exponentials will always be 2. Asymptotically, the entropy of entanglement characterizes the entanglement present in a state in the following sense: given $|\psi\rangle^{\otimes n}$ with $E := E(|\psi\rangle)$, *entanglement concentration* [2] can produce $nE - o(n)$ maximally entangled states (i.e. $|\Phi\rangle^{\otimes nE - o(n)}$, where $|\Phi\rangle^{AB} := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$) with $o(1)$ error, while *entanglement dilution* [2, 3] can map $|\Phi\rangle^{\otimes nE + o(n)}$ to $|\psi\rangle^{\otimes n}$ with $o(1)$ error and $o(n)$ bits of classical communication. Thus, to leading order in the number of copies, tensor powers of entangled pure states can be described by a single parameter: the entropy of entanglement.

In general, however, entangled pure states cannot be fully described by the entropy of entanglement, even given free local operations. Following [4], define the *entanglement spread* of a state $|\psi\rangle^{AB}$ to be

$$\Delta(\psi) := \log \text{rank } \psi^A + \log \|\psi^A\|_\infty. \quad (1)$$

The entanglement spread is never negative and $\Delta(\psi) = 0$ if and only if ψ has all non-zero Schmidt coefficients equal, meaning it is a product state or a maximally entangled state. If $\Delta(\psi) > 0$ then we say that $|\psi\rangle$ is *partially entangled* (since it is neither unentangled nor maximally entangled). Since $\Delta(\psi_1 \otimes \psi_2) = \Delta(\psi_1) + \Delta(\psi_2)$, it follows that if $\Delta(\psi) > 0$, then $\Delta(\psi^{\otimes n}) = \Theta(n)$. However, this value is not very robust: for any constant $\epsilon > 0$, we can perturb $|\psi\rangle^{\otimes n}$ by ϵ and reduce its spread to $\Theta(\sqrt{n})$. To capture this insight we will also use the ϵ -perturbed entanglement spread $\Delta_\epsilon(\psi)$, which is defined for any $\epsilon \geq 0$ to be (following [4])

$$\min \{ \log \text{tr } P + \log \|P\psi^A P\|_\infty : \text{tr } P\psi^A \geq 1 - \epsilon \},$$

where the minimization is over projectors P . Note that $\Delta_0(\psi) = \Delta(\psi)$. Now $\Delta_\epsilon(\psi^{\otimes n}) = \Theta(\sqrt{n})$ for any $\epsilon > 0$, but of course there are still states on $2n$ qubits,

such as the even superposition between $|\Phi\rangle^{\otimes n}$ and a product state, which have Δ_ϵ spread nearly equal to n .

The main application of entanglement spread is using the following result from [4] to produce lower bounds on communication:

Theorem 1.1 (Corollary 10 of [4]). *If $|\phi\rangle^{AB}$ is transformed using local operations and C bits of classical communication (in either direction) into a state that has fidelity $1 - \epsilon$ with $|\psi\rangle$ then*

$$C \geq \Delta_\delta(\psi) - \Delta_0(\phi) + 2\log(1 - \delta),$$

where $\delta = (4\epsilon)^{1/8}$.

In particular, if we begin with maximally entangled states, then preparing $|\psi\rangle$ to within a fidelity of $1 - \epsilon$ requires $\Delta_\delta(\psi) + 2\log(1 - \delta)$ bits of communication. This restriction holds even with an unlimited supply of EPR pairs, and so problematizes the idea that maximally entangled states are a good canonical form for the resource of pure state entanglement. In Section 2.2 we will discuss alternative ways to quantify entanglement as a resource.

2. Dealing with entanglement spread

We propose two solutions to the above problems. To build quantum protocols that can be run in superposition, we demand that they discard only (approximately) standard states to the environment, as we will describe in Section 2.1. Then in Section 2.2, we will propose measuring not just the maximum amount of entanglement that can be created by a protocol, but instead finding the range of entanglement that it can cleanly generate/consume. These approaches are not rigid rules, but rather illustrate principles that can be adapted to diverse situations.

2.1. Approach: clean resource inequalities

Allowing free local operations is standard practice in quantum information theory, but discarding to the environment can be dangerous when running different protocols in superposition. Following and extending [5], we will say that a *clean resource inequality* exists, and is denoted $\alpha \stackrel{\text{clean}}{\geq} \beta$, when α can be mapped to β using only (up to error ϵ_n that goes to 0 as $n \rightarrow \infty$)

- Local unitaries.
- Adding ancillas initialized in the $|0\rangle$ states.
- Discarding ancillas in the $|0\rangle$ state.
- Discarding messages that have been sent through classical channels.
- A dynamic resource ($\mathcal{N} : \omega$) (meaning, as defined in [1], an operation $\mathcal{N}$ constrained to act on average input ω) may be used as a consumed resource only in a way that sends a constant state to the environment. More formally, suppose we are given $(\mathcal{N}^{A_1 A_2 A_3 \rightarrow BE} : \omega^{A_1})$. Then we need to replace this resource with

$(\mathcal{N}^{A_1 A_2 A_3 \rightarrow B} : \tilde{\omega}^{A_1 A_2})$ such that $\tilde{\omega}^{A_1} = \omega^{A_1}$ and $\mathcal{N}(\sigma)^E$ is the same for all σ satisfying $\sigma^{A_1 A_2} = \tilde{\omega}^{A_1 A_2}$.

This last point says that noisy resources can be used only when they leak information to the environment that is independent of the inputs or the particular protocol being run. For example, if we are given $[c \rightarrow c]$ as part of the input resource α then we can use it only if we promise to input the same distribution of 0 and 1 regardless of which protocol we're using. On the other hand, $[q \rightarrow q]$ can be used with any input since it doesn't leak anything to the environment.

The primary application of clean protocols is the following general principle.

Lemma 2.1 (protocol superposition principle). *Suppose that Alice and Bob would like to execute m different operations, $\mathcal{P}_1, \dots, \mathcal{P}_m$ in superposition. This means that they would like to perform an operation $\mathcal{P}$ that satisfies*

$$\mathcal{P} \sum_{k=1}^m c_k |k\rangle_A |k\rangle_B |\psi\rangle_{AB} \approx_{\epsilon_n} \sum_{k=1}^m c_k |k\rangle_A |k\rangle_B \mathcal{P}_k |\psi\rangle_{AB}$$

for any coefficients $\{c_k\}$, and where $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

Let $\mathcal{R}_k$ denote the set of resources capable of simulating $\mathcal{P}_k$ cleanly: $\mathcal{R}_k := \left\{ \alpha : \alpha \stackrel{\text{clean}}{\geq} \langle \mathcal{P}_k \rangle \right\}$. Then

$$\mathcal{R} := \left\{ \alpha : \alpha \stackrel{\text{clean}}{\geq} \langle \mathcal{P} \rangle \right\} = \bigcap_k \mathcal{R}_k.$$

One direction of the proof is easy: $\mathcal{R} \subseteq \bigcap_k \mathcal{R}_k$, since $\mathcal{P} \stackrel{\text{clean}}{\geq} \mathcal{P}_k$ for each k . To prove $\mathcal{R} \supseteq \bigcap_k \mathcal{R}_k$, we start with $\alpha \in \bigcap_k \mathcal{R}_k$ and clean protocols for $\alpha \stackrel{\text{clean}}{\geq} \mathcal{R}_k$ for each k . Then have Alice and Bob run each protocol conditioned on their (shared) value of k . Since each resource inequality is clean, nothing is discarded that would break superpositions over different values of k .

2.2. Approach: entanglement capacity as an interval

Armed with the definition of clean RIs, we now examine the entangling capacities of various quantum operations. The most obvious restriction is that entanglement cannot be cleanly discarded, so that e.g. while $2[qq] \geq [qq]$, it does not hold that $2[qq] \stackrel{\text{clean}}{\geq} [qq]$. Moreover, eliminating entanglement cleanly is now a non-trivial resource. So $[c \rightarrow c : I/2] \stackrel{\text{clean}}{\geq} -[qq]$ via a protocol where Alice sends her half of a shared state $|\Phi\rangle$ through the classical channel and Bob performs a CNOT with the bit he receives as control and with his half of $|\Phi\rangle$ as target. Then he is left with a $|0\rangle$, which he discards. At the same time, $[c \rightarrow c : I/2] \stackrel{\text{clean}}{\geq} \emptyset$ (where $\emptyset$ is the null resource), since Alice can always send a random bit through a channel. This means that both protocols can be run in superposition and classical communication can

be used to generate superpositions of different amounts of entanglement. We can express this concisely by saying that the *entanglement capacity range* of $[c \rightarrow c]$ (alternately, its *spread capacity*) contains $[-1, 0]$. In fact, the spread capacity of $[c \rightarrow c]$ is exactly $[-1, 0]$ as can be seen from the non-increase of entanglement under LOCC for the upper bound and Theorem 1.1 for the lower bound.

We can apply this approach to other resources as well. The spread capacity of $[q \rightarrow q]$, for example, is $[-1, 1]$, since sending one qubit can coherently either create or destroy one ebit. The resource $[qq]$, on the other hand, has spread capacity equal to the single number $\{1\}$. An example of a more complicated resource is a bipartite unitary gate U . If the asymptotic entanglement capacity [6] of U is $E(U) := \max\{e : \langle U \rangle \geq e[qq]\}$, then the spread capacity of U is $[-E(U^\dagger), E(U)]$.

By time-sharing, the spread capacity of a resource α is completely characterized by an upper and lower bound (although there can always be tradeoffs between entanglement and other resources). And by Lemma 2.1 clean protocols using any amount of entanglement within that range can be performed in superposition using α . Thus, spread capacities can be a useful framework for determining when communication protocols can be combined in superposition. Examples of this are given in the arxiv version of this paper [7].

Acknowledgments: I want to thank Charlie Bennett, Patrick Hayden, Debbie Leung, Peter Shor and Andreas Winter for many useful conversations on this subject. My funding is from the Army Research Office under grant W9111NF-05-1-0294, the European Commission under Marie Curie grants ASTQIT (FP6-022194) and QAP (IST-2005-15848), and the U.K. Engineering and Physical Science Research Council through “QIP IRC.”

References

- [1] I. Devetak, A. W. Harrow and A. J. Winter, *IEEE Trans. Inf. Theory* **54**, 4587 (2008), arXiv:quant-ph/0512015.
- [2] C. H. Bennett, H. J. Bernstein, S. Popescu and B. Schumacher, *Phys. Rev. A* **53**, 2046 (1996), arXiv:quant-ph/9511030.
- [3] H.-K. Lo and S. Popescu, *Phys. Rev. Lett.* **83**, 1459 (1999).
- [4] P. Hayden and A. Winter, *pra* **67**, p. 012306 (2003), arXiv:quant-ph/0204092.
- [5] A. Harrow and P. Shor, Time reversal and exchange symmetries of unitary gate capacities (2005), arXiv:quant-ph/0511219.
- [6] C. H. Bennett, A. W. Harrow, D. W. Leung and J. A. Smolin, *IEEE Trans. Inf. Theory* **49**, 1895 (2003), arXiv:quant-ph/0205057.
- [7] A. W. Harrow, Entanglement spread and clean resource inequalities (2009), arXiv:0909.1557.

OPTIMAL DECOUPLING

RENATO RENNER

Institute for Theoretical Physics, ETH Zurich, Switzerland
E-mail: renner@phys.ethz.ch

Given a bipartite quantum system with parts A and R , we say that a mapping $\mathcal{E}$ applied to A decouples A from R if the outcome of $\mathcal{E}$ is uncorrelated to R . The notion of decoupling plays a crucial role in various information-theoretic arguments and is also used for foundational considerations in the context of statistical mechanics.

Here, we consider decoupling operations $\mathcal{E}$ which take the form of projective measurements. We review a recent result which shows that a randomly chosen projective measurement achieves decoupling if and only if a certain entropic quantity, called smooth entropy, is sufficiently large. Furthermore, the random choice is almost always optimal.

Keywords: Quantum information theory, state merging, smooth entropies, decoupling

1. Introduction

A central objective of information theory is the analysis of information-processing tasks. A traditional, and probably the most fundamental, example of such a task is *data compression*. Given a source emitting information A , one is interested in the minimum number of bits, ℓ , required to store the data such that A can later be recovered with almost certainty. In his seminal paper, Shannon showed that, under appropriate assumptions on the structure of the source, ℓ is asymptotically equal to the von Neumann entropy^a of the source, $H(A)$ [1]. More precisely, he considered an *iid source*^b, which emits n symbols $A_1, \dots, A_n$ distributed jointly according to $P_{A_1 \dots A_n} = P_A^{\times n}$ (where P_A is a probability distribution over a countable set $\mathcal{A}$). Let $\ell^\varepsilon(A^n)$ be the minimum number of bits from which $A^n = (A_1, \dots, A_n)$ can be reconstructed with a maximum error probability $\varepsilon \geq 0$. The *compression rate*, i.e., the number of storage bits required per symbol emitted by the source then satisfies

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{\ell^\varepsilon(A^n)}{n} = H(A) \quad (1)$$

where $H(A) := -\sum_{a \in \mathcal{A}} P_A(a) \log_2 P_A(a)$.

An important aspect of this result is that it reveals a connection between a quantity which is relevant in physics (the von Neumann entropy) on the one hand

^aIn the information theory literature, $H(A)$ is usually referred to as *Shannon entropy*, in honor of Shannon's contributions to the field.

^bThe term *iid* stands for *independent and identically distributed*.

and a pure information-theoretic quantity (the compression rate) on the other hand. Moreover, while one may argue that the compression rate is a slightly artificial quantity, it turns out that the von Neumann entropy plays a role as a measure for information in a variety of other information-theoretic considerations (e.g., related to channel capacities), too. In fact, it is fair to say that the von Neumann entropy is of almost universal significance in both classical and quantum information theory.

This raises some obvious questions: is there a deeper reason why the von Neumann entropy is so relevant in information theory? What is the common feature of information-theoretic questions that makes the von Neumann entropy so useful? Answering these questions in full would of course be far beyond the scope of this short article — and probably also beyond state-of-the-art knowledge in this research area. Here, we rather want to sketch some ideas that may point a direction towards possible answers.

To this purpose, we consider a specific information-theoretic primitive, called *decoupling* (Section 2). Its significance for the above discussion is due to two crucial properties. (i) Decoupling is quantitatively characterized by an entropy measure, called *smooth entropy* [2], which directly generalizes von Neumann entropy (Section 3). (ii) Decoupling is a building block for various information-processing tasks such as state merging, which can be seen as a generalization of data compression (Section 4).

2. Decoupling

Let ρ_{AR} be a density operator on a bipartite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_R$. Furthermore, let $\mathcal{E}$ be a completely positive map from the set of operators on $\mathcal{H}_A$ to the set of operators on a Hilbert space $\mathcal{H}_{A'} \otimes \mathcal{H}_C$. We assume that all involved Hilbert spaces are finite-dimensional. We denote by $\mathcal{I}_R$ the identity map on the set of operators on $\mathcal{H}_R$, and by $\tau_{A'} := \frac{\text{id}_{A'}}{\dim \mathcal{H}_{A'}}$ the fully mixed state on $\mathcal{H}_{A'}$. Also, we use subscripts to denote the reduced state obtained by taking the partial trace of a density operator defined on a larger space, e.g., $\rho_{CR} = \text{tr}_{A'}(\rho_{A'CR})$, for $\rho_{A'CR}$ defined on $\mathcal{H}_{A'} \otimes \mathcal{H}_C \otimes \mathcal{H}_R$.

Definition 2.1. For any $\varepsilon \geq 0$, we say that $\mathcal{E}$ ε -decouples the A -part of ρ_{AR} if $\rho_{A'CR} := (\mathcal{E} \otimes \mathcal{I}_R)(\rho_{AR})$ satisfies^c

$$\frac{1}{2} \|\rho_{A'CR} - \tau_{A'} \otimes \rho_{CR}\|_1 \leq \varepsilon .$$

There are simple examples of mappings $\mathcal{E}$ which perfectly decouple the A -part of any density operator ρ_{AR} , e.g., $\mathcal{E} : \sigma_A \mapsto \text{tr}(\sigma_A)\tau_{A'}$, where $\mathcal{H}_C$ is trivial (i.e., 1-dimensional).

In view of information-theoretic applications, one is typically interested in decoupling mappings $\mathcal{E}$ that have a specific structure. A particularly relevant example,

^cThe factor $\frac{1}{2}$ in front of the trace norm is a convention. The induced distance measure for operators is known as the *trace distance* and is widely used in quantum information theory.

on which we will focus in this article, are projective measurements. More precisely, we assume that $\mathcal{E}$ is obtained by concatenating the mapping $\mathcal{U} : \sigma_A \mapsto U \sigma_A U^\dagger$, for some isometry U from $\mathcal{H}_A$ to $\mathcal{H}_{A'} \otimes \mathcal{H}_C$, and a measurement on $\mathcal{H}_C$ with respect to an orthonormal basis $\{|e_j\rangle\}_j$. Formally, $\mathcal{E} = (\mathcal{I}_{A'} \otimes \mathcal{P}_C) \circ \mathcal{U}$, where

$$\mathcal{P}_C : \sigma_C \mapsto \sum_j |e_j\rangle\langle e_j| \sigma_C |e_j\rangle\langle e_j|. \quad (2)$$

In the next section, we will characterize the set of states ρ_{AR} for which mappings $\mathcal{E}$ of this type can decouple the A -part. For this, we make use of the notion of smooth entropies.

3. Smooth Entropies

Smooth entropies have been introduced in [2] in the context of classical information theory and later been generalized to quantum information theory [3]. In the following, we only state their definition and explain in which sense they generalize von Neumann entropy. For more details on smooth entropies and their properties we refer to [4, 5].

Definition 3.1. For any $\varepsilon \geq 0$, the ε -smooth conditional min-entropy of a positive operator ρ_{AR} on $\mathcal{H}_A \otimes \mathcal{H}_R$ is defined by

$$H_{\min}^\varepsilon(A|R)_{\rho_{AR}} := \sup\{\lambda : \bar{\rho}_{AR} \leq 2^{-\lambda} \cdot \text{id}_A \otimes \sigma_R\},$$

where σ_R is any density operator on $\mathcal{H}_R$ and where $\bar{\rho}_{AR}$ is any operator in $\mathcal{B}^\varepsilon(\rho_{AR})$, the ε -ball around ρ_{AR} with respect to the *purified distance* (see [6] for more details). Furthermore, the ε -smooth conditional max-entropy of a positive operator ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ is defined by

$$H_{\max}^\varepsilon(A|B)_{\rho_{AB}} := -H_{\min}^\varepsilon(A|R)_{\rho_{AR}},$$

where ρ_{AR} is the reduced state obtained by tracing out system $\mathcal{H}_B$ of any purification ρ_{ABR} of ρ_{AB} .

Smooth min- and max-entropies can be seen as generalizations of the von Neumann entropy [7]. More precisely, the conditional von Neumann entropy, $H(A|R)_{\rho_{AB}}$, of a state ρ_{AB} can be expressed asymptotically in terms of the smooth min- or max-entropy of the iid state $\rho_{A^n B^n} = \rho_{AB}^{\otimes n}$, i.e.,

$$H(A|B)_{\rho_{AB}} = \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} H_{\min}^\varepsilon(A^n | B^n)_{\rho_{AB}^{\otimes n}} = \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} H_{\max}^\varepsilon(A^n | B^n)_{\rho_{AB}^{\otimes n}}. \quad (3)$$

We are now ready to review the main results of [8, 9]. The following two theorems imply that the smooth min-entropy (almost) exactly characterizes the decoupling that can be achieved by projective measurements.

Theorem 3.1. *Let ρ_{AR} be a fixed density operator and let ε_U be the minimum value such that the mapping $\mathcal{E}$ defined before and by (2), for some fixed measurement basis*

$\{|e_j\rangle\}_j$ and for an arbitrary isometry U , ε_U -decouples the A -part of ρ_{AR} . Then, for any $\varepsilon \geq 0$ and $\ell = \log_2 \dim \mathcal{H}_{A'}$,

$$\int \varepsilon_U dU \leq 2^{-\frac{1}{2}(H_{\min}^\varepsilon(A|R)_{\rho_{AR}} - \ell)} + \varepsilon ,$$

where dU denotes the Haar measure on all isometries U .

In other words, decoupling is possible by projective measurements if $H_{\min}^\varepsilon(A|R)$ is (sufficiently) larger than the size, measured in terms of qubits, ℓ , of the remaining system $\mathcal{H}_{A'}$. Furthermore, the projective measurement can be chosen at random. The next theorem implies that such a random choice is essentially optimal for decoupling.

Theorem 3.2. *Let ρ_{AR} be a density operator and let $\mathcal{E}$ be any mapping of the form specified before and by (2). If $\mathcal{E}$ ε -decouples ρ_{AR} then*

$$\ell \leq H_{\min}^{\sqrt{2\varepsilon}}(A|R)_{\rho_{AR}}$$

where $\ell = \log_2 \dim \mathcal{H}_{A'}$.

4. Applications

Decoupling can be seen as a building block for various information-processing tasks. One particularly interesting example is *state merging* [10]. Consider two parties, *Alice* and *Bob*, who control quantum systems $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, and let ρ_{AB} be the density operator describing the initial state of their joint system, $\mathcal{H}_A \otimes \mathcal{H}_B$. Furthermore, assume that Alice and Bob can exchange classical messages and, in addition, share fully entangled qubits. Their goal is to transfer Alice's system $\mathcal{H}_A$ to Bob, so that Bob eventually holds both systems (therefore the name *merging*). This task should be achieved using only minimum resources, measured, for instance, in terms of the entanglement used up by the process. (In fact, it turns out that for initial states ρ_{AB} satisfying $H_{\max}^\varepsilon(A|B)_{\rho_{AB}} < 0$, entanglement can be generated.)

In [10], an explicit state merging protocol has been proposed. It essentially consists of a decoupling step by projective measurements as in (2). After this step, Alice's system is decoupled from any auxiliary system, denoted $\mathcal{H}_R$, which contains a purification of Alice and Bob's joint state. As shown in [10], the amount of entanglement generated by this protocol is equal to the size ℓ of the system obtained by this decoupling operation. Hence, by the results described in Section 3, we conclude that the amount of entanglement generated by a state merging protocol, starting from a state ρ_{AB} , is given by

$$\ell \approx H_{\min}^\varepsilon(A|R)_{\rho_{AR}} = -H_{\max}^\varepsilon(A|B)_{\rho_{AB}} .$$

(The approximation $\approx$ indicates that equality holds up to additive terms of the order $\log 1/\varepsilon$.) In particular, using (3), if Alice and Bob initially share an iid state $\rho_{AB}^{\otimes n}$, then the entanglement production rate (i.e., the total entanglement divided by n) is asymptotically equal to the negative von Neumann entropy, $-H(A|B)_{\rho_{AB}}$.

The notion of decoupling has various other information-theoretic applications. For example, it is used in quantum cryptography for randomness extraction (also known as privacy amplification) [11]. Furthermore, it has been applied for the characterization of correlation and entanglement between systems, erasure processes, as well as channel capacities (see, for instance, [12–14]).

Outside the area of information theory, the concept of decoupling could be employed in the context of statistical mechanics. In particular, the evolution of a thermodynamical system towards thermal equilibrium can be understood as a decoupling process, where the system under consideration decouples from the observer (somewhat analogous to the considerations in [15]). Recent work indeed suggests that there is a close relation between smooth entropies and quantities that are relevant in thermodynamics [16]. Similarly, black hole radiation may be analyzed from such a point of view [17].

References

- [1] C. Shannon, *Bell System Technical Journal* **27**, p. 379 (1948).
- [2] R. Renner and S. Wolf, Smooth Rényi entropy and applications, in *Proc. International Symposium on Information Theory*, (Chicago, IL, 2004).
- [3] R. Renner, Security of quantum key distribution, PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, (Zurich, Switzerland, 2005). Electronic version: arXiv:quant-ph/0512258.
- [4] N. Datta and R. Renner, *IEEE Transactions on Information Theory* **55**, p. 2807 (2009).
- [5] R. König, R. Renner and C. Schaffner, *IEEE Transactions on Information Theory* **55**, p. 4337 (2009).
- [6] M. Tomamichel, R. Colbeck and R. Renner, Duality between smooth min- and max-entropies, arXiv:0907.5238 (2009).
- [7] M. Tomamichel, R. Colbeck and R. Renner, to appear in *IEEE Transactions on Information Theory* (2009).
- [8] M. Berta, Single-shot quantum state merging, Master’s thesis, ETH Zurich (2008).
- [9] M. Berta, R. Renner, A. Winter and J. Wullschleger, One-shot source coding, in preparation.
- [10] M. Horodecki, J. Oppenheim and A. Winter, *Communications in Mathematical Physics* **269**, p. 107 (2007).
- [11] R. Renner and R. König, Universally composable privacy amplification against quantum adversaries, in *Second Theory of Cryptography Conference (TCC)*, ed. J. Kilian, Lecture Notes in Computer Science, Vol. 3378 (Springer, Cambridge, MA, 2005).
- [12] B. Groisman, S. Popescu and A. Winter, *Physical Review A* **72**, 032317 (2005).
- [13] F. Buscemi, Private quantum decoupling, arXiv:0901.4506 (2009).
- [14] P. Hayden, M. Horodecki, A. Winter and J. Yard, *Open Systems and Information Dynamics* **15**, p. 7 (2008).
- [15] N. Linden, S. Popescu, A. Short and A. Winter, *Phys. Rev. E* **79**, 061103 (2009).
- [16] O. Dahlsten, R. Renner, E. Rieper and V. Vedral, The work value of information, arXiv:0908.0424 (2009).
- [17] P. Hayden and J. Preskill, *Journal of High Energy Physics* **07**, p. 120 (2007).

ENTANGLEMENT AT FINITE TEMPERATURE IN THE ELECTRONIC TWO-PARTICLE INTERFEROMETER

PETER SAMUELSSON¹, IZHAR NEDER², and MARKUS BÜTTIKER³

¹*Division of Mathematical Physics, Lund University, Box 118, S-221 00 Lund, Sweden*

²*Physics Department, Harvard University, Cambridge, Massachusetts 02138, USA*

³*Département de Physique Théorique, Université de Genève, CH-1211 Genève 4, Switzerland*

In this work we discuss a theory for entanglement generation, characterization and detection in fermionic two-particle interferometers (2PI) at finite temperature. The motivation for our work is provided by the recent experiment by the Heiblum group, Neder *et al.*, Nature **448**, 333 (2007), realizing the 2PI proposed by Samuelsson, Sukhorukov, and Büttiker, Phys. Rev. Lett. **92**, 026805 (2004). The experiment displayed a clear two-particle Aharonov-Bohm effect, however with an amplitude suppressed due to finite temperature and dephasing. In our recent work Samuelsson, Neder, and Büttiker, Phys. Rev. Lett. **102**, 106804 (2009) we presented a general theory for finite temperature entanglement in mesoscopic conductors and applied it to the 2PI, here further discussed.

Keywords: Entanglement, mesoscopic, quantum state tomography.

1. Introduction

There is presently a strong interest in computation and information processing based on fundamental principles of quantum mechanics [1]. The prospect of scalability and integrability with conventional electronics makes solid state systems a likely future arena for quantum information processing. Of particular interest is the entanglement between the elementary charge carriers, quasiparticles, in meso- or nanoscopic solid state conductors. Due to controllable system properties and coherent transport conditions, conductors on the meso- and nanoscale constitute ideal systems for the generation and detection of quasiparticle entanglement.

To date quasiparticle entanglement has however remained experimentally elusive. A class of mesoscopic systems that appear promising for a successful entanglement experiment are conductors without direct interactions between the quasiparticles. It was shown by Beenakker *et al* [2] that fermions emitted from a thermal source can be entangled by scattering at a beam-splitter. This was originally discussed for electron-hole pairs [2] and shortly afterward for pairs of electrons [3, 4]. Since then there has been a large number of works on entanglement of non-interacting particles, see e.g. [5] for a review.

Several of the entanglement proposals have been based on electrical analogs of

optical interferometers and beam-splitter geometries. In this work we will focus on the electronic two-particle, or Hanbury Brown Twiss [6, 7], interferometer, proposed theoretically in Ref. [3]. Recently, an electronic 2PI was realized experimentally [8]. In perfect agreement with the theoretical predictions [3], the two-particle interference pattern was visible in the current correlations but not in the average current.

As discussed in Ref. [3], there is an intimate relation between two-particle interference and entanglement in the fermionic 2PI. Under ideal conditions, i.e. zero temperature and perfect coherence, two-particle interference implies that the two particle wave function is on the form

$$|\Psi_s\rangle = (1/\sqrt{2}) [|1\rangle_A |2\rangle_B - |2\rangle_A |1\rangle_B]. \quad (1)$$

Here 1, 2 denote the sources and A, B the sites of detection, as shown in Fig. 1. The wavefunction $|\Psi_s\rangle$ is maximally entangled, it is a singlet in the orbital [9], or pseudo spin, space $\{|1\rangle, |2\rangle\}$.

However, in the experiment [8], $\sim 25\%$ visibility of the current correlation oscillations was observed. This indicates that both decoherence and finite temperature is important. The experimental findings thus raised two important questions: are the electrons reaching the detectors at A and B entangled and if so, can this two-particle entanglement be unambiguously detected by measurements of currents and current correlators, the standard quantities accessible in electronic transport measurements? In our recent work [10] we provided a positive answer to both these questions. Here this is further discussed.

2. Fermionic two-particle interferometer

In the experiment [8] on the fermionic 2PI [3] it was possible to electrically tune the system between two individual Mach-Zehnder [11] interferometers and a 2PI. By measuring the Aharonov-Bohm (AB) effect in both the Mach-Zehnder interferometers and the 2PI, a two-particle AB-effect in the current correlations could be established beyond doubt.

In the experiment semitransparent beam splitters were used, $T_C = 1 - R_C = 1/2$, $T_D = 1 - R_D = 1/2$. For the current cross correlations, theory for finite temperature and dephasing, $\gamma < 1$, [13] predicts, for $A+, B+$,

$$S_{A+B+} = -[e^3 V / (4h)] H [1 - \gamma \sin \phi]. \quad (2)$$

with ϕ the AB-phase. The temperature dependence is fully contained in $H = \coth(eV/2kT) - 2kT/eV$, varying from unity for $kT \ll eV$ to zero for $kT \gg eV$. The effect of finite temperature is thus to suppress the overall amplitude of the current cross correlation oscillations.

3. Entanglement of projected and reduced density matrices

The two-particle entanglement at finite temperature is formally given by the entanglement of the density matrix $\rho_p(E)$ of the two-particle state projected out from

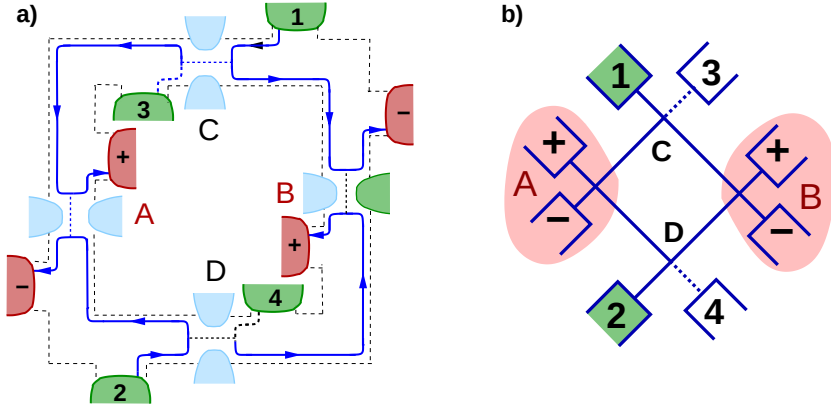


Fig. 1. a) Fermionic 2PI implemented in a conductor in the quantum Hall regime, from [3]. b) Schematic of the topologically equivalent two-particle interferometer (2PI) with beam splitters C,D and biased, active (grounded, inactive) source contacts 1,2 (3,4). Detector regions A and B (red shaded) contain beam splitters and grounded contacts $\pm$.

the emitted many body state. Following Ref. [5] we find for the 2PI

$$\rho_p(E) = (1 - f)^2 f_V^2 [\chi \rho_p^{diag} + (1 - \chi)^2 \rho^{int}] \quad (3)$$

where f and f_V are the Fermi distribution functions of the grounded and biased terminals respectively, $\chi = e^{-eV/kT}$ and the diagonal density matrix $\rho_p^{diag} = \chi \hat{1} \otimes \hat{1} + (1 - \chi)[\rho_A \otimes \hat{1} + \hat{1} \otimes \rho_B]$ with the zero temperature single particle density matrices $\rho_A = R_C|+\rangle\langle+| + R_D|-\rangle\langle-|$ and $\rho_B = T_C|+\rangle\langle+| + T_D|-\rangle\langle-|$. The density matrix $\rho^{int} = R_C T_D|+-\rangle\langle-+| + R_D T_C|+-\rangle\langle-+| - \gamma \sqrt{T_C R_C T_D R_D} [|+-\rangle\langle-+| + |-+\rangle\langle-+|]$ results from the two-particle interference. Here we used the shorthand notation $|+-\rangle \equiv |+\rangle_A |-\rangle_B$ with $|+\rangle_A$ denoting the state in the upper lead incident on splitter A etc and $\langle-+| = (|+-\rangle)^\dagger$ etc. Note that the effect of decoherence enters as a suppression of the off-diagonal components of ρ^{int} as $|\Psi^{int}\rangle\langle\Psi^{int}| \rightarrow \rho^{int}$, where $|\Psi^{int}\rangle = \sqrt{R_C T_D} |+-\rangle - \sqrt{T_C R_D} |-+\rangle$.

For the entanglement, following [5] we introduce σ_p and $w_p(E)$, the normalized density matrix and the emission probability of the emitted two-particle state respectively, defined from $\rho_p(E) = w_p(E)\sigma_p$, with $w_p(E) = \text{tr}[\rho_p(E)] = (1 - f)^2 f_V^2 [(R_C T_D + T_C R_D)(1 - \chi)^2 + 4\chi]$ where we note that σ_p is independent on energy. The concurrence production per unit energy is then $C_p(E) \equiv w_p(E)C(\sigma_p)$ and the total entanglement production during a time τ , $C_p = (\tau/h) \int dE C_p(E)$, is then ($\mathcal{N} = \tau eV/h$)

$$C_p = (\mathcal{N}H/2) \max \left\{ 4\gamma \sqrt{T_C R_C T_D R_D} - \sinh^{-2}(eV/2kT), 0 \right\}. \quad (4)$$

We denote this the projected entanglement. As shown in Fig. 2, C_p decreases monotonically as a function of T . It reaches zero at a critical temperature T_c^p . Inserting the parameter values from the experiment, we get $C_p \approx 0.1\mathcal{N}$ and $C(\sigma_p) \approx 0.3$, i.e. *the state emitted by the 2PI toward regions A and B was clearly entangled.*

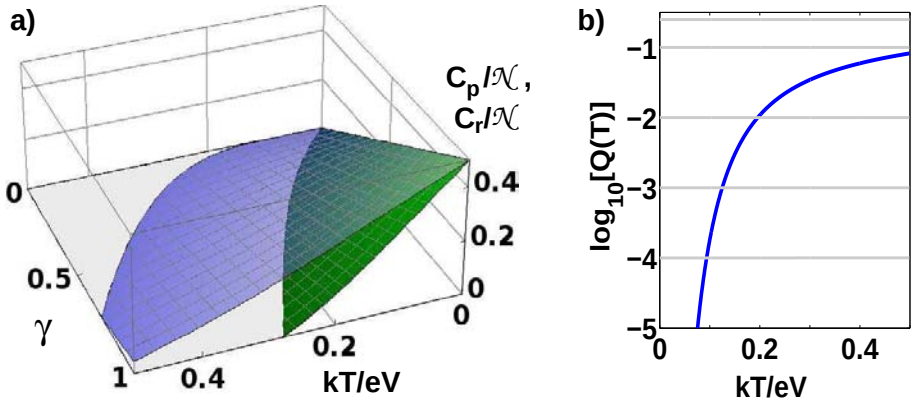


Fig. 2. a) Entanglement production C_p/N (blue, transparent) and C_r/N (green, opaque) as functions of temperature kT/eV and coherence γ for the semi-transparent 2PI. b) Parameter Q as a function of kT/eV (blue line). Values $0.25, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$ shown (gray lines). Figure reproduced from Ref. [10].

The entanglement of the projected density matrix is the entanglement one could access, had one been able to do arbitrary local operations and classical communication between A and B, i.e. fully energy and particle resolved measurements. Under realistic conditions this is not possible [14], the accessible physical quantities are currents and current cross correlators. Extending the zero temperature scheme for quantum state tomography in Ref. [12] it is natural to describe the two-particle properties of the emitted many-body state, available via current and current cross correlations, in terms of a reduced two-particle density matrix ρ_r . As discussed in Ref. [10] we have for finite temperature

$$\rho_r = (1 - H)[\rho_A \otimes \rho_B] + H\rho^{int}. \quad (5)$$

From this we see that the effect of increasing temperature is to monotonically increase the amplitude for the separable product state $\rho_A \otimes \rho_B$, while the amplitude of the interference component is suppressed. Dephasing $\gamma < 1$, in contrast only suppresses the off-diagonal components of ρ_{int} .

Turning to entanglement, introducing the normalized reduced density matrix σ_r we can write $\rho_r = w_r \sigma_r$, with $w_r = \text{tr}[\rho_r] = [R_C T_C + R_D T_D](1 - H) + R_C T_D + R_D T_C$. We then define the total entanglement production during a time τ as $C_r \equiv \mathcal{N} w_r C(\sigma_r)$. It is

$$C_r = 2\mathcal{N} \max\{\sqrt{T_C R_C T_D R_D} [H(1 + \gamma) - 1], 0\} \quad (6)$$

here called the reduced entanglement. As C_p , C_r decreases monotonically with increasing T . It reaches zero at a critical temperature T_c^r .

4. Entanglement bound and detection

Comparing Eqs. (4) and (6) quantitatively we find that $C_p \geq C_r$ for

$$Q(T) = H/[4(1 - H) \sinh^2(eV/2kT)] \leq \sqrt{T_C R_C T_D R_D}, \quad (7)$$

independent on γ (see Fig. 2). Consequently, for beam splitters away from the strongly asymmetrical (tunneling) limit, *the reduced entanglement constitutes a lower bound for the projected entanglement.*

Turning to the experiment [8], for the relevant parameters we have $Q(T) \approx 4 \times 10^{-4} \ll \sqrt{R_C T_C R_D T_D} \approx 0.25$, showing the validity of the bound. However, $C_r \approx 0.01\mathcal{N}$ and based on the measurement [8] no conclusive statement can be made about C_r .

5. Conclusions

In conclusion, we have investigate the effect of finite temperature on the entanglement production and detection in the fermionic two-particle interferometer, presenting an extended discussion of the results in Ref. [10]. We have shown that, based on parameters extracted from the 2PI experiment by Neder et al [8], the emitted two-particle state in the experiment was clearly entangled. However, an unambiguous demonstration of the entanglement via measurements of currents and current correlations requires a reduction of the dephasing and the temperature.

References

- [1] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000).
- [2] C.W.J. Beenakker, C. Emary, M. Kindermann, and J. L. van Velsen, *Phys. Rev. Lett.* **91**, 147901 (2003).
- [3] P. Samuelsson, E.V. Sukhorukov, and M. Büttiker, *Phys. Rev. Lett.* **92**, 026805 (2004).
- [4] C.W.J. Beenakker, M. Kindermann, C.M. Marcus, and A. Yacoby, in *Fundamental Problems of Mesoscopic Physics*, edited by I.V. Lerner, B.L. Altshuler, and Y. Gefen, NATO Science Series II. Vol. 154 (Kluwer, Dordrecht, 2004).
- [5] C.W.J. Beenakker, in *Proc. Int. School Phys. E. Fermi*, Vol. 162, *Quantum Computers, Algorithms and Chaos*, eds. G. Casati *et al* (IOS Press, Amsterdam, 2006).
- [6] R. Hanbury Brown and R. Q. Twiss, *Nature* **177**, 27, (1956).
- [7] R. Hanbury Brown and R. Q. Twiss, *Nature* **178**, 1046, (1956).
- [8] I. Neder, N. Ofek, Y. Chung, M. Heiblum, D. Mahalu, and V. Umansky, *Nature* **448**, 333 (2007).
- [9] P. Samuelsson, E.V. Sukhorukov, and M. Büttiker, *Phys. Rev. Lett.* **91**, 157002 (2003).
- [10] P. Samuelsson, I. Neder, and M. Büttiker, *Phys. Rev. Lett.* **102**, 106804 (2009).
- [11] Y. Ji, Y. Chung, D. Sprinzak, M. Heiblum, D. Mahalu, and H. Shtrikman, *Nature* **422**, 415 (2003).
- [12] P. Samuelsson, and M. Büttiker, *Phys. Rev. B* **73**, 041305 (2006).
- [13] V.S.-W. Chung, P. Samuelsson, and M. Büttiker, *Phys. Rev. B* **72**, 125320 (2005).
- [14] W.-R. Hannes, and M. Titov, *Phys. Rev. B* **77**, 115323 (2008).

SPECTRAL PROPERTIES OF DISCRETE ALLOY-TYPE MODELS

MARTIN TAUTENHAHN and IVAN VESELIĆ

*Emmy-Noether-Projekt Schrödingeroperatoren, Fakultät für Mathematik,
Technische Universität Chemnitz, 09127 Chemnitz, Germany
www.tu-chemnitz.de/mathematik/enp/*

We discuss recent results on spectral properties of discrete alloy-type random Schrödinger operators. They concern Wegner estimates and bounds on the fractional moments of the Green's function.

Keywords: random Schrödinger operators; discrete alloy-type model; single-site potential.

1. Introduction and model

A discrete alloy-type model is a family of operators $H_\omega = H_0 + V_\omega$ on $\ell^2(\mathbb{Z}^d)$ where H_0 is the negative discrete Laplacian given by $(H_0\psi)(x) = -\sum_{|e|=1} \psi(x+e)$, and V_ω a multiplication operator by the function

$$V_\omega(x) = \sum_{k \in \mathbb{Z}^d} \omega_k u(x-k). \quad (1)$$

Here $\omega_k, k \in \mathbb{Z}^d$, is an i.i.d. sequence of random variables and $u: \mathbb{Z}^d \rightarrow \mathbb{R}$ a so-called single-site potential. The minimal conditions which are assumed throughout this note are that the distribution of ω_0 has a density $\rho \in L^\infty(\mathbb{R})$ and that u belongs to $\ell^1(\mathbb{Z}^d)$. There are several well-studied relatives of the discrete alloy-type model:

- (i) The continuum alloy-type model $-\Delta + V_\omega$ on $L^2(\mathbb{R}^d)$ where V_ω has the same form as in (1).
- (ii) The Anderson model on $\ell^2(\mathbb{Z}^d)$: this is the special case of the discrete alloy-type model where $u = \delta_0$.
- (iii) The correlated Anderson model, which corresponds to $u = \delta_0$ and the case where the random coupling constants ω_k are no longer assumed independent.

Obviously, the discrete alloy-type model is a special case of case (iii). However, previous results on (iii)-type models require certain regularity conditions which are typically not satisfied for alloy-type potentials, cf. Section 3.

The main challenge of models of the type (1) is that u may change its sign which leads to negative correlations of the potential. This results in a non-monotone dependence on the random variables of certain spectral quantities. For this reason many established tools for the spectral analysis of random Schrödinger operators are not directly applicable in this model.

For literature concerning the discussion in this section we refer to the references of [1, 2].

2. Results

Our results concern averages of resolvents and projectors. The presented results are closely related to dynamical and spectral localisation. We use the symbol $\mathbb{E}$ to denote the average over the collection of random variables ω_k , $k \in \mathbb{Z}^d$. For $\Lambda \subset \mathbb{Z}^d$ we denote by $H_{\omega, \Lambda} : \ell^2(\Lambda) \rightarrow \ell^2(\Lambda)$ the natural restriction of H_ω to the set Λ .

2.1. Wegner estimate

We present here a selection of the results proven in [2]. For $L \in \mathbb{N}$ we denote the set $[0, L]^d \cap \mathbb{Z}^d$ by Λ_L . The number of points in the support of u is denoted by $\text{rank } u$.

Theorem 2.1. *Assume that the single-site potential u has support in Λ_n and ρ is of bounded variation. Then there exists a constant c_u depending only on u such that for any $L \in \mathbb{N}$, $E \in \mathbb{R}$ and $\epsilon > 0$ we have*

$$\mathbb{E} \left\{ \text{Tr} \left[\chi_{[E-\epsilon, E+\epsilon]}(H_{\omega, \Lambda_L}) \right] \right\} \leq c_u \|\rho\|_{\text{var}} \text{rank } u \epsilon (L+n)^{d \cdot (n+1)}.$$

Here $\|\rho\|_{\text{var}}$ denotes the total variation of ρ . By the assumption on the support of the single-site potential, $\text{rank } u \leq n^d$. Our bound is linear in the energy-interval length and polynomial in the volume of the cube. This implies that the Wegner bound can be used for a localisation proof via multiscale analysis, as soon as an appropriate initial length scale estimate is at disposal.

Theorem 2.2. *Assume $\bar{u} := \sum_{k \in \mathbb{Z}^d} u(k) \neq 0$ and that ρ has compact support. Let $m \in \mathbb{N}$ be such that $\sum_{\|k\| \geq m} |u(k)| \leq |\bar{u}|/2$. Then we have for any $L \in \mathbb{N}$, $E \in \mathbb{R}$ and $\epsilon > 0$*

$$\mathbb{E} \left\{ \text{Tr} \left[\chi_{[E-\epsilon, E+\epsilon]}(H_{\omega, \Lambda_L}) \right] \right\} \leq \frac{8}{\bar{u}} \|\rho\|_{\text{var}} \min(L^d, \text{rank } u) \epsilon (L+m)^d.$$

In the case that the support of u is compact we obtain a bound which is linear in the volume of the box and thus yields the Lipschitz continuity of the integrated density of states.

2.2. Boundedness of fractional moments of Green's function

In this subsection we assume that the function $\hat{u} : [0, 2\pi)^d \rightarrow \mathbb{C}$, defined by

$$\hat{u}(\theta) = \sum_{k \in \mathbb{Z}^d} u(k) e^{ik \cdot \theta},$$

does not vanish, $\text{supp } u$ is compact, and $\rho \in W^{1,1}(\mathbb{R})$. Let $\Lambda \subset \mathbb{Z}^d$ be finite. For $x, y \in \Lambda$ and $z \in \mathbb{C} \setminus \mathbb{R}$ we set $G_{\omega, \Lambda}(z; x, y) = \langle \delta_x, (H_{\omega, \Lambda} - z)^{-1} \delta_y \rangle$.

Theorem 2.3. *Let $s \in (0, 1)$ and $\Lambda \subset \mathbb{Z}^d$ be finite. Then there exists a constant C_u depending only on u , such that for all $z \in \mathbb{C} \setminus \mathbb{R}$ and all $x, y \in \Lambda$*

$$\mathbb{E}\left\{|G_{\omega, \Lambda}(z; x, y)|^s\right\} \leq (C_u \|\rho'\|_{L^1})^s \frac{2^{1+s} s^{-s}}{1-s}.$$

Theorem 2.3 is proven in Section 4. The proof is a combination of ideas from [3] and [4]. The proof also gives rise to a quantitative estimate on the constant C_u . To our knowledge, Theorem 2.3 does not immediately imply exponential decay of fractional moments of the Green's function by standard methods.

2.3. Exponential decay of fractional moments of Green's function

In this subsection we assume $d = 1$ and $\text{supp } u$ compact. Denote the diameter of the support by $n - 1 \in \mathbb{N}_0$. We define the following conditions which may or may not hold:

- (A) $\text{supp } u = \{0, 1, \dots, n - 1\}$,
- (B) $\text{supp } \rho$ compact,
- (C) $\rho \in W^{1,1}(\mathbb{R})$.

For $x, y \in \mathbb{Z}^d$ and $z \in \mathbb{C} \setminus \mathbb{R}$ we denote the Green's function by $G_\omega(z; x, y) = \langle \delta_x, (H_\omega - z)^{-1} \delta_y \rangle$ and set

$$C_{u, \rho} = \left| \prod_{k \in \Theta} u(k) \right|^{-s/n} \|\rho\|_\infty \frac{2^s s^{-s}}{1-s}.$$

Theorem 2.4. *Let $s \in (0, 1)$ and assume that (A) holds. Then there exists a constant C , depending on $\|\rho\|_\infty$, s and u , such that for all $x, y \in \mathbb{Z}$ with $|x - y| \geq n$ and all $z \in \mathbb{C} \setminus \mathbb{R}$ we have*

$$\mathbb{E}\left\{|G_\omega(z; x, y)|^{s/n}\right\} \leq C \exp\left\{-m \left\lfloor \frac{|x - y|}{n} \right\rfloor\right\}, \quad (2)$$

where $m = -\ln C_{u, \rho}$ and $\lfloor \cdot \rfloor$ is defined by $\lfloor t \rfloor = \max\{k \in \mathbb{Z} : k \leq t\}$.

If $C_{u, \rho} < 1$, or equivalently $m > 0$, Ineq. (2) describes exponential decay.

Theorem 2.5. *Let $s \in (0, 1/2)$ and assume that either (B) holds with $\|\rho\|_\infty$ sufficiently small, or (C) holds with $\|\rho'\|_{L^1}$ sufficiently small. Then there exist constants $m, C \in (0, \infty)$, such that Eq. (2) holds for all $x, y \in \mathbb{Z}$ with $|x - y| \geq 4n$ and all $z \in \mathbb{C} \setminus \mathbb{R}$.*

Theorem 2.4 and 2.5 are proven in a joint paper [1] with A. Elgart. The estimates of Theorem 2.4 and 2.5 concern only off-diagonal elements. If we assume (B) and $s \in (0, 1/4n)$, then $\mathbb{E}\{|G_\omega(z; x, y)|^s\}$ is uniformly bounded for $x, y \in \mathbb{Z}$ and $z \in \mathbb{C} \setminus \mathbb{R}$, see [1, Appendix]. In spite these estimates on the Green's function, neither dynamical nor spectral localisation follow immediately using the existent methods.

Currently, the analogues of Theorem 2.4 and 2.5 in higher dimension are an open question. In a forthcoming paper with A. Elgart we intend to extend Theorem 2.5 to operators on a one-dimensional strip, i. e. $\Lambda = \Gamma \times \mathbb{Z}$ with $\Gamma \subset \mathbb{Z}^d$ finite.

3. Example

We show that discrete alloy-type models do not satisfy in general the regularity conditions required in [5, 6] for correlated Anderson models. We consider $d = 1$. Let the density function ρ be of bounded support and let $u(0) = 1$, $u(-1) = a \neq 0$ and $u(k) = 0$ for $k \in \mathbb{Z} \setminus \{-1, 0\}$. For simplicity we require that the infimum of the support of ρ is zero. Let B and A be the events $B = \{\omega \in \Omega : V_\omega(-1), V_\omega(1) \in [0, \epsilon]\}$ and $A = \{\omega \in \Omega : V_\omega(0) \in [0, \epsilon(a + \frac{1}{a})]\}$ with $\epsilon > 0$. Then, if $a > 0$ one calculates $B \subset A$ for all $\epsilon > 0$ and consequently we have

$$\mathbb{P}\{A \mid B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}} = 1 \quad \forall \epsilon > 0.$$

This shows that the regularity assumptions required in [5, 6] are violated. In the case $a < 0$ one can proceed analogously, but with a different choice of the sets A and B .

4. Proof of Theorem 2.3

Let $L \in \mathbb{N}$ such that the cube $\Lambda^+ = [-L, L]^d \cap \mathbb{Z}^d$ contains $\bigcup_{x \in \Lambda} \{k \in \mathbb{Z}^d : u(x-k) \neq 0\}$, which is the set of all lattice sites whose coupling constant influence the potential in Λ . Let $A : \ell^1(\mathbb{Z}^d) \rightarrow \ell^1(\mathbb{Z}^d)$ be the linear operator whose coefficients in the canonical orthonormal basis are $A(j, k) = u(j-k)$ for $j, k \in \mathbb{Z}^d$. Since u has compact support, the operator A is bounded. Since $\hat{u}$ does not vanish, $C_u = \|A^{-1}\|_1 < \infty$, see [4] for details. Moreover, there exists an invertible matrix $A_{\Lambda^+} : \ell^1(\Lambda^+) \rightarrow \ell^1(\Lambda^+)$ satisfying

$$A_{\Lambda^+}(j, k) = u(j-k) \quad \text{for all } j \in \Lambda \text{ and } k \in \Lambda^+ \quad (3)$$

and $\|A_{\Lambda^+}^{-1}\| \leq C_u$, see [4, Proposition 5]. We set $B_{\Lambda^+} = A_{\Lambda^+}^{-1}$. From [5] and [7], respectively, we infer that for $x, y \in \Lambda$ with $x \neq y$

$$|G_{\omega, \Lambda}(z; x, x)| = \frac{1}{|V_\omega(x) - \alpha|} \quad \text{and} \quad |G_{\omega, \Lambda}(z; x, y)| \leq \frac{2}{|V_\omega(x) - \beta|} + \frac{2}{|V_\omega(y) - \gamma|}. \quad (4)$$

Here $\alpha, \beta \in \mathbb{C}$ are functions of $V_\omega(k)$, $k \in \Lambda \setminus \{x\}$, and $\gamma \in \mathbb{C}$ is a function of $V_\omega(k)$, $k \in \Lambda \setminus \{y\}$. Set $\omega_{\Lambda^+} = (\omega_k)_{k \in \Lambda^+}$, $k(\omega_{\Lambda^+}) = \prod_{k \in \Lambda^+} \rho(\omega_k)$, $d\omega_{\Lambda^+} = \prod_{k \in \Lambda^+} d\omega_k$ and $\Omega_{\Lambda^+} = \times_{k \in \Lambda^+} \mathbb{R}$. Using the substitution $\zeta_{\Lambda^+} = (\zeta_k)_{k \in \Lambda^+} = A_{\Lambda^+} \omega_{\Lambda^+}$ and Eq. (3) we obtain

$$S = \int_{\Omega_{\Lambda^+}} \frac{1}{|V_\omega(x) - \alpha|^s} k(\omega_{\Lambda^+}) d\omega_{\Lambda^+} = \int_{\Omega_{\Lambda^+}} \frac{1}{|\zeta_x - \alpha|^s} k(B_{\Lambda^+} \zeta_{\Lambda^+}) |\det B_{\Lambda^+}| d\zeta_{\Lambda^+},$$

where $d\zeta_{\Lambda^+} = \prod_{k \in \Lambda^+} d\zeta_k$. Since by construction $\zeta_k = V_\omega(k)$ for all $k \in \Lambda$, α is now a function of ζ_k , $k \in \Lambda \setminus \{x\}$. For non-negative functions $g : \mathbb{R} \rightarrow \mathbb{R}$ with $g \in W^{1,1}(\mathbb{R})$ and $\delta \in \mathbb{C}$ one has for all $\lambda > 0$ the estimate

$$\int_{\mathbb{R}} \frac{g(\xi)}{|\xi - \delta|^s} d\xi \leq \lambda^{-s} \|g\|_{L^1} + \|g\|_{\infty} \frac{2\lambda^{1-s}}{1-s} \leq \lambda^{-s} \|g\|_{L^1} + \|g'\|_{L^1} \frac{\lambda^{1-s}}{1-s}, \quad (5)$$

where the first inequality is due to [7] and the second is the fundamental theorem of calculus. Set $\Omega_{\Lambda^+}^x = \times_{k \in \Lambda^+ \setminus \{x\}} \mathbb{R}$. Using Fubini's theorem and Ineq. (5) we obtain

$$\begin{aligned} S &\leq \int_{\Omega_{\Lambda^+}^x} \left(\lambda^{-s} \int_{\mathbb{R}} k(B_{\Lambda^+} \zeta_{\Lambda^+}) d\zeta_x + \frac{\lambda^{1-s}}{1-s} \int_{\mathbb{R}} \left| \frac{\partial}{\partial \zeta_x} k(B_{\Lambda^+} \zeta_{\Lambda^+}) \right| d\zeta_x \right) |\det B_{\Lambda^+}| \prod_{k \in \Lambda^+ \setminus \{x\}} d\zeta_k \\ &= \lambda^{-s} + \frac{\lambda^{1-s}}{1-s} \int_{\Omega_{\Lambda^+}^x} \left| \frac{\partial}{\partial \zeta_x} k(B_{\Lambda^+} \zeta_{\Lambda^+}) \right| |\det B_{\Lambda^+}| d\zeta_{\Lambda^+}. \end{aligned}$$

We calculate the partial derivative by the product rule, substitute back into original coordinates and obtain

$$\begin{aligned} S &\leq \lambda^{-s} + \frac{\lambda^{1-s}}{1-s} \int_{\Omega_{\Lambda^+}^x} \left| \sum_{j \in \Lambda^+} \rho'((B_{\Lambda^+} \zeta_{\Lambda^+})_j) B_{\Lambda^+}(j, x) \prod_{\substack{k \in \Lambda^+ \\ k \neq j}} \rho((B_{\Lambda^+} \zeta_{\Lambda^+})_k) \right| |\det B_{\Lambda^+}| d\zeta_{\Lambda^+} \\ &\leq \lambda^{-s} + \frac{\lambda^{1-s}}{1-s} \sum_{j \in \Lambda^+} |B_{\Lambda^+}(j, x)| \int_{\Omega_{\Lambda^+}^x} |\rho'(\omega_j)| \prod_{\substack{k \in \Lambda^+ \\ k \neq j}} \rho(\omega_k) d\omega_{\Lambda^+} \\ &= \lambda^{-s} + \frac{\lambda^{1-s}}{1-s} \|B_{\Lambda^+}\|_1 \|\rho'\|_{L^1} \leq \lambda^{-s} + \frac{\lambda^{1-s}}{1-s} C_u \|\rho'\|_{L^1}. \end{aligned}$$

If we choose $\lambda = s/(C_u \|\rho'\|_{L^1})$ we obtain $S \leq C_u^s \|\rho'\|_{L^1}^s s^{-s}/(1-s)$. Thus we have shown $\mathbb{E}\{|G_{\omega, \Lambda}(z; x, x)|^s\} \leq C_u^s \|\rho'\|_{L^1}^s \frac{s^{-s}}{1-s}$. Analogously we obtain for $x \neq y$ the estimate $\mathbb{E}\{|G_{\omega, \Lambda}(z; x, y)|^s\} \leq 2^{s+1} C_u^s \|\rho'\|_{L^1}^s \frac{s^{-s}}{1-s}$. The additional factor 2^{s+1} arises in the case $x \neq y$, since in Eq. (4) there are two summands of the type S , and each summand has an additional factor 2 in its numerator. $\square$

References

- [1] A. Elgart, M. Tautenhahn and I. Veselić, Exponential decay of Green's function for Anderson models on $\mathbb{Z}$ with single-site potentials of finite support, Preprint, arXiv:0903.0492 [math-ph], (2009).
- [2] I. Veselić, Wegner estimates for discrete alloy-type models, Preprint, mp_arc no. 09-100, (2009).
- [3] M. Tautenhahn, Lokalisierung für korrelierte Anderson Modelle, Diplomarbeit, Techn. Univ. Chemnitz (2007).
- [4] I. Veselić, Wegner estimates for sign-changing single site potentials, Preprint, arxiv:0806.0482 [math.SP], (2008).
- [5] M. Aizenman and S. Molchanov, *Commun. Math. Phys.* **157**, 245 (1993).
- [6] M. Aizenman, J. H. Schenker, R. M. Friedrich and D. Hundertmark, *Commun. Math. Phys.* **224**, 219 (2001).
- [7] G. M. Graf, *J. Stat. Phys.* **75**, 337 (1994).

COMPLETE DYNAMICAL LOCALIZATION IN DISORDERED QUANTUM MULTI-PARTICLE SYSTEMS

MICHAEL AIZENMAN^(a) and SIMONE WARZEL^(b)

^(a) *Departments of Mathematics and Physics, Princeton University
Princeton, NJ 08544, USA*

^(b) *Zentrum Mathematik, TU München
Boltzmannstr. 3, 85747 Garching, Germany
E-mail: aizenman@princeton.edu, warzel@ma.tum.de*

We present some recent results concerning the persistence of dynamical localization for disordered systems of n particles under weak interactions.

1. Introduction

After more than half a century, Anderson localization continues to attract the interest of a broad spectrum of researchers ranging from experimentalists, who find its effects in systems of cold atoms and in photonic crystals [7], to mathematical physicists. Considerable progress was made in the rigorous methods for the study of the localization effects of disorder in the context of the one-particle theory [6, 14, 17]. More recently attention has centered on the role of interactions, and questions related to the persistence of the localization picture in the presence of inter-particle interactions. We shall report here on some progress which was made in that area.

1.1. The one-particle theory

The discussion of weakly interacting particles often starts from the approximation in which the interactions are ignored. That is, one first considers systems of Fermions, or Bosons possibly with on-site repulsion, subject to a fixed potential which includes random terms. Such systems can be understood in terms of the one-particle theory.

For particles moving on a lattice $\mathbb{Z}^d$, the single-particle Hamiltonian may take the form

$$H^{(1)}(\omega) := -\Delta + \lambda V(x; \omega), \quad \text{in } \ell^2(\mathbb{Z}^d), \quad (1)$$

where $-\Delta$ is the lattice Laplacian and the random potential incorporates a disorder parameter $\lambda \geq 0$. For the convenience of presentation, all mathematical statements in this note are made under the assumption that the random potential $V(x; \omega)$ takes independent and identically distributed values on the lattice sites $x \in \mathbb{Z}^d$, with a distribution $\mathbb{P}(V_x \in dv) = \varrho(v) dv$ of a bounded, compactly supported density.

When talking about localization for operators such as $H^{(1)}(\omega)$, different notions have been used and established:

Spectral localization, in an energy regime $I \subset \mathbb{R}$, refers to the statement that within the specified energy regime $H^{(1)}(\omega)$ has only pure point spectrum, with exponentially localized eigenfunctions, cf. [17] and references therein.

Dynamical localization, in $I \subset \mathbb{R}$, refers to the statement that any initially localized state, which is a wave packet with energies in $I \subset \mathbb{R}$, will remain exponentially localized indefinitely under the time evolution. A convenient sufficient condition is that for some $A, \xi \in (0, \infty)$ (both depending on I, λ),

$$(DL) \quad \mathbb{E} \left[\sup_{t \in \mathbb{R}} |\langle \delta_y, e^{-itH^{(1)}} P_I(H^{(1)}) \delta_x \rangle|^2 \right] \leq A e^{-|x-y|/\xi}$$

where $\delta_x \in \ell^2(\mathbb{Z}^d)$ is the δ state at x and $P_I(H^{(1)})$ is a spectral projection.

While spectral localization provides some coarse information on the quantum time evolution via the RAGE theorem, dynamical localization in the sense of (DL) is a somewhat stronger notion. Historically, spectral localization was the first to be rigorously established by building, for $d > 1$ at extreme energies or large disorder, on the multiscale analysis of Fröhlich and Spencer [12]; cf. [14, 17]. The first proof of dynamical localization [2] relied on the fractional moment analysis of [1].

It is one of the features of the single particle model (1) that in any dimension there is an extreme disorder regime, with λ exceeding some finite λ_1 , in which one has **complete dynamical localization**, meaning that *all* states are localized and dynamical localization (DL) holds with $I = \mathbb{R}$.

For the single particle model (1) in $d = 1$ complete dynamical localization with critical disorder strength $\lambda_1 = 0$ was established already in [15]. For higher dimensions, $d \geq 1$, an explicit bound on λ_1 can be found in [2].

1.2. The influence of interaction on localization: a challenge

It is not hard to see that the above localization properties are inherited by a system of non-interacting particles with a one-particle Hamiltonian corresponding to (1). An important question however, is whether such behavior will persist under the addition of interparticle interactions. Especially interesting is the situation where there are n fermions in a region of volume $|\Lambda|$, with $|\Lambda| \rightarrow \infty$ and $n/|\Lambda| \rightarrow \rho > 0$.

A rather strong claim is being advocated by Basko, Aleiner and Altshuler [8], who argue that *under strong disorder weak interactions do not change the qualitative picture of localization as it is seen in the non-interacting model*.

The proposal is rather startling. It includes the claim that if the system of particles of an overall positive density is started at a initial state at which the distribution in space of particles and energy is far from uniform, its irregularity will persist indefinitely under the time evolution. This seems to run against the vague *equidistribution principle*, by which one expects that except under unusual

circumstances, such as in the non-interacting integrable model, the initial state will evolve in time (in a weak enough sense) towards states which maximize the entropy, in a coarse-grained sense, subject to the given energy and particle number constraints.

Persistence of localization is not undisputed among physicist. In particular, long-range repulsive interactions are conjectured to have a delocalizing effect [16].

Rigorous methods are still far from allowing one to decide whether complete localization will persist or perish in the presence of interactions. Furthermore, the analysis of even a fixed number of particles with short range interactions in the infinite-volume limit has presented difficulties, and we will report on some recent progress [5, 10, 11] made in this direction.

As an side we note that the dynamics of multiparticle systems bear some relation to non-linear evolutions. Recent results on that topic are discussed in [18].

2. Dynamical localization for multi-particle systems

2.1. The n -particle Hamiltonian

We will be concerned with a system of finitely many interacting particles in the random potential described above. The n -particle Hamiltonian is given by

$$H^{(n)}(\omega) := \sum_{j=1}^n [-\Delta_j + \lambda V(x_j; \omega)] + \mathcal{U}(\mathbf{x}; \boldsymbol{\alpha}), \quad \text{in } \ell^2(\mathbb{Z}^d)^n \quad (2)$$

acting in the Hilbert space over all configurations $\mathbf{x} = (x_1, \dots, x_n) \in (\mathbb{Z}^d)^n$. Here the last term is a p -site interaction of range $\ell_U < \infty$:

$$\mathcal{U}(\mathbf{x}; \boldsymbol{\alpha}) := \sum_{k=1}^p \alpha_k \sum_{\substack{A \subset \mathbb{Z}^d: |A|=k \\ \text{diam } A \leq \ell_U}} U_A((N_u(\mathbf{x}))_{u \in A}).$$

It is described in term of an interaction parameter $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p) \in \mathbb{R}^p$ and a function $U_A : \mathbb{N}^{|A|} \rightarrow \mathbb{R}$ which is bounded by some n -dependent constant, $\|U_A\| \leq c_n < \infty$, and translational invariant. The latter depends on the number $N_u(\mathbf{x}) := \sum_{j=1}^n \delta_{u, x_j}$ of particles of the configuration $\mathbf{x} = (x_1, \dots, x_n)$, which are at sites $u \in A$ in the pattern $A \subset \mathbb{Z}^d$.

Simply stated examples which are already of interest are short range pair interactions for which $p = 2$ and $U_{\{u, v\}}(N_u(\mathbf{x}), N_v(\mathbf{x})) = N_u(\mathbf{x})N_v(\mathbf{x})\delta_{|u-v|, 1}$.

2.2. The main result

Our main result establishes the existence, for any dimension and any number of particles, of a regime $\mathcal{L}_n^{(p)}$ in the space of basic parameters of the model, $(\lambda, \boldsymbol{\alpha}) \in \mathbb{R}_+ \times \mathbb{R}^p$, for which complete dynamical localization with a uniform localization length occurs for up to n particles. The proof gives an inductive algorithm for the

construction of such localization regimes, albeit at what may be a possibly far too restrictive manner.

The first proof of spectral localization for two particles was presented in [11], using the multiscale approach. The result presented below was derived by different means, which allow also rather simple control of the dynamical localization.

The localization regime covered by the result presented below may be best described in terms of the extreme sets it includes, namely:

Strong disorder: for each $\alpha \in \mathbb{R}^p$ there is $\lambda(\alpha)$ such that the localization regime $\mathcal{L}_n^{(p)}$ includes the cone in the parameter space $\mathbb{R}_+ \times \mathbb{R}^p$ where the interaction strengths are dominated by α , and the disorder strength exceeds $\lambda(\alpha)$.

Weak interactions: for any $\lambda > \lambda_1$, i.e. disorder strength at which the one-particle Hamiltonian exhibits complete localization, there are $\alpha_j(\lambda) > 0$, $j = \{1, \dots, p\}$, such that $\mathcal{L}_n^{(p)}$ includes all (λ, α') for which $|\alpha'_j| \leq |\alpha_j(\lambda)|$ componentwise.

Thus, the weak localization region includes some neighborhood of the *entire* localization regime $(\lambda_1, \infty) \times \{\mathbf{0}\}$ of the n particle unperturbed system. In case $d = 1$, $\lambda_1 = 0$, so this includes all positive values of λ .

Having introduced these notions we may present the result of [5]:

Theorem 2.1 ([5]). *For each $n, p \in \mathbb{N}$ there is an open set $\mathcal{L}_n^{(p)} \subset \mathbb{R}_+ \times \mathbb{R}^p$ which includes regimes of strong disorder and weak interactions, for which at some $A, \xi < \infty$ and all $(\lambda, \alpha) \in \mathcal{L}_n^{(p)}$, $k \in \{1, \dots, n\}$, and all $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}^d)^k$:*

$$\mathbb{E} \left[\sup_{t \in \mathbb{R}} \left| \langle \delta_{\mathbf{x}}, e^{-itH^{(k)}} \delta_{\mathbf{y}} \rangle \right|^2 \right] \leq A e^{-\text{dist}_{\mathcal{H}}(\mathbf{x}, \mathbf{y})/\xi}, \quad (3)$$

where the exponential decay is in terms of the Hausdorff pseudo-distance between the configurations $\mathbf{x} = (x_1, \dots, x_k)$ and $\mathbf{y} = (y_1, \dots, y_k)$

$$\text{dist}_{\mathcal{H}}(\mathbf{x}, \mathbf{y}) := \max \left\{ \max_{1 \leq i \leq k} \text{dist}(x_i, \mathbf{y}), \max_{1 \leq i \leq k} \text{dist}(y_i, \mathbf{x}) \right\} \quad (4)$$

As was mentioned, spectral localization follows from (3) by the Wiener criterion.

2.3. Comments on the result

Theorem 2.1 is formulated for distinguishable particles. For *Fermions*, or *Bosons*, the relevant Hilbert spaces are subspaces of the space which this theorem covers, and thus the results apply by restriction. Furthermore, *hard core interactions* can also be added without requiring any modification in the proof of the extended statement.

It is worth pointing out that for systems of $n > 2$ particles some subtleties show up in the decay rate seen in (3). A natural decay rate for systems with permutation symmetry is the symmetrized distance: $\text{dist}_S(\mathbf{x}, \mathbf{y}) := \min_{\pi \in S_n} \sum_{j=1}^n |x_j - y_{\pi j}|$, where S_n is the permutation group of $\{1, \dots, n\}$. In contrast, $\text{dist}_{\mathcal{H}}(\mathbf{x}, \mathbf{y})$ is not a metric, if $n > 2$. The Hausdorff pseudo-distance bounds (3) allow for the possibility that if the initial configuration had some particles within the localization distance

from each other then such ‘excess charge’ may transfer among the occupied regions.

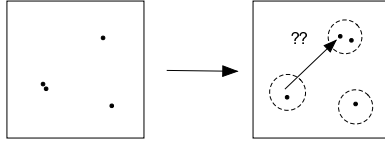


Fig. 1. A transition between two configurations which are close in the Hausdorff pseudo-distance, but not otherwise.

One could wonder whether Theorem 2.1 does not immediately follow from the existing localization results. After all, a configuration of n particles in $\mathbb{Z}^d$ may be regarded as a single particle in $\mathbb{Z}^{dn}$ evolving under a Hamiltonian which is of the form $H^{(n)}(\omega) = -\Delta^{(nd)} + U(\mathbf{x}) + \lambda V(\mathbf{x}; \omega)$ with a random potential $V(\mathbf{x}; \omega) = \sum_{j=1}^n V(x_j; \omega)$. However, the random potential $V(\mathbf{x}; \omega)$ does not take independent values at configurations with one coinciding particle and in this sense is of *infinite range*. The number of its degrees of freedom is only a fractional power ($\frac{1}{n}$) of the number of configuration. The technical problem which this causes is behind the above mentioned limitation of the bound (3).

3. Outline of the proof

The proof of Theorem 2.1 cannot be fitted in this short summary, but we may comment on the flow of the argument [5]. It consists of three steps:

- (1) Proof of the finiteness of fractional moments of the Green function.
- (2) Elucidation of the relation of a multiparticle *eigenfunction correlator* with the Green function's fractional moments.
- (3) Inductive construction of domains of uniform localization in parameter space.

The first is a *sine qua non* condition for the fractional moment technique for localization [1]. It is related to the celebrated Wegner estimate, which however does not play a direct role in the analysis. The proof is rather easy, as is also the case for the Wegner estimate for many-particle systems.

The second step follows the strategy by which dynamical localization is established for single particles [2], for which some adjustments to the multi-particle setup are required.

Once the first two preparatory steps are taken, the third step constitutes the core of the argument.

3.1. Finiteness of fractional moments of the Green function

An essential tool for the analysis is the finite-volume, $\Lambda \subset \mathbb{Z}^d$, Green function

$$G_{\Lambda}^{(n)}(\mathbf{x}, \mathbf{y}; z) := \langle \delta_{\mathbf{x}}, (H_{\Lambda}^{(n)} - z)^{-1} \delta_{\mathbf{y}} \rangle, \quad (5)$$

with $H_{\Lambda}^{(n)}$ denoting the restriction of (2) to $\ell^2(\Lambda)^n$. The first step towards *exponential bounds* is to prove finiteness of the fractional moments. In fact, even a conditional average makes them finite.

Lemma 3.1. *For any $s \in (0, 1)$ there is $C < \infty$ such that for any $\Lambda \subseteq \mathbb{Z}^d$, any two (not necessarily distinct) sites $u_1, u_2 \in \Lambda$ and any pair of configurations $\mathbf{x}, \mathbf{y}$, which have a particle at u_1, u_2 respectively,*

$$\mathbb{E} \left(\left| G_{\Lambda}^{(n)}(\mathbf{x}, \mathbf{y}; z) \right|^s \mid \{V(v)\}_{v \notin \{u_1, u_2\}} \right) \leq \frac{C}{|\lambda|^s} \quad (6)$$

for all $z \in \mathbb{C}$, and $\lambda \neq 0$.

Sketch of proof. In its dependence on $V(u_1)$ and $V(u_2)$, the Hamiltonian is of the form: $H_{\Lambda}^{(n)} = A + \lambda V(u_1) N_{u_1} + \lambda V(u_2) N_{u_2}$, where $(N_u \psi)(\mathbf{x}) := \sum_{k=1}^n \delta_{x_k, u} \psi(\mathbf{x})$ stands for the number operator. The assertion is therefore implied by the weak- L^1 estimate from [4]:

$$\int_{\mathbb{R}^2} 1 \left[\left| \left\langle \sqrt{N} \phi, (\xi N + \eta M - K)^{-1} \sqrt{M} \psi \right\rangle \right| > t \right] \varrho(\xi) \varrho(\eta) d\xi d\eta \leq \frac{C[\varrho]}{t} \|\phi\| \|\psi\|,$$

with operators $N, M \geq 0$ and K dissipative. $\square$

The finiteness (6) is the analogue of a *Wegner estimate* in the multi-scale method. As has been noticed in [9, 13], it follows similar to the one-particle case from the monotonicity used in the above proof.

3.2. Eigenfunction correlator and its relation to Green function

For the finite-volume Hamiltonian $H_{\Lambda}^{(n)}$, we define the *eigenfunction correlator*, associated with an energy regime $I \subset \mathbb{R}$, as the sum

$$Q_{\Lambda}^{(n)}(\mathbf{x}, \mathbf{y}; I) := \sum_{E \in \sigma(H_{\Lambda}^{(n)}) \cap I} \left| \langle \delta_{\mathbf{x}}, P_{\{E\}}(H_{\Lambda}^{(n)}) \delta_{\mathbf{y}} \rangle \right|, \quad (7)$$

where $P_{\{E\}}$ is a spectral projection. Ignoring some subtleties related to the passage to the infinite-volume limit $\Lambda \uparrow \mathbb{Z}^d$ (which are discussed in detail in [2, 5]), the elementary bound $\left| \langle \delta_{\mathbf{x}}, e^{-itH_{\Lambda}^{(n)}} \delta_{\mathbf{y}} \rangle \right| \leq Q_{\Lambda}^{(n)}(\mathbf{x}, \mathbf{y}; \mathbb{R})$ shows that the eigenfunction correlator is an important tool for establishing dynamical localization [2, 15].

The following key lemma states the equivalence of exponential decay in the Hausdorff pseudo-distance (though not in the symmetric distance) of the eigenfunction correlator and of the Green function's fractional moments.

Lemma 3.2. *The following statements are equivalent:*

(1) *There is $A, \xi < \infty$ such that for all $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}^d)^k$:*

$$\sup_{I \subset \mathbb{R}} \sup_{\Lambda \subset \mathbb{Z}^d} \mathbb{E} \left[Q_{\Lambda}^{(k)}(\mathbf{x}, \mathbf{y}; I) \right] \leq A e^{-\text{dist}_{\mathcal{H}}(\mathbf{x}, \mathbf{y})/\xi}. \quad (8)$$

(2) *There is $A, \xi < \infty$ and $s \in (0, 1)$ such that for all $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}^d)^k$:*

$$\sup_{\substack{I \subset \mathbb{R} \\ |I| \geq 1}} \sup_{\Lambda \subset \mathbb{Z}^d} \frac{1}{|I|} \underbrace{\int_I \mathbb{E} \left[\left| G_{\Lambda}^{(k)}(\mathbf{x}, \mathbf{y}; E) \right|^s \right] dE}_{=: \widehat{\mathbb{E}}_I[\dots]} \leq A e^{-\text{dist}_{\mathcal{H}}(\mathbf{x}, \mathbf{y})/\xi}. \quad (9)$$

In the proof of the main result we repeatedly change horses between estimates on the kernels of G and Q . The kernel $Q(\mathbf{x}, \mathbf{y})$ is what is ultimately need, and it is also a very convenient tool for the induction step. Yet, $G(\mathbf{x}, \mathbf{y})$ has a more convenient perturbation theory.

In the induction step described below one puts together two mutually non-interacting subsystems and then turns on the interaction between them. Thus, we start with the system partitioned into $J, K \subset \{1, \dots, n\}$, $J \cap K = \emptyset$, and a Hamiltonian of the form $H_{\Lambda}^{(J,K)} := H_{\Lambda}^{(J)} \oplus H_{\Lambda}^{(K)}$ acting on $\ell^2(\Lambda)^{|J|} \otimes \ell^2(\Lambda)^{|K|}$. Then:

- The Green function of the composite system is a convolution,

$$G_{\Lambda}^{(J,K)}(\mathbf{x}, \mathbf{y}; z) = \int_{\mathbb{C}} G_{\Lambda}^{(J)}(\mathbf{x}_J, \mathbf{y}_J; z - E) G_{\Lambda}^{(K)}(\mathbf{x}_K, \mathbf{y}_K; E) \frac{dE}{2\pi i} \quad (10)$$

involving a contour integral encircling the spectrum of $H_{\Lambda}^{(K)}$. Moment estimates for the combined system are complicated by the unboundedness of G and possible correlations between the eigenvalues of the subsystems.

- On the other hand, the eigenfunction correlator is bounded and we may use:

$$Q_{\Lambda}^{(J,K)}(\mathbf{x}, \mathbf{y}; I) \leq Q_{\Lambda}^{(J)}(\mathbf{x}_J, \mathbf{y}_J; \mathbb{R}) Q_{\Lambda}^{(K)}(\mathbf{x}_K, \mathbf{y}_K; \mathbb{R}), \quad (11)$$

by which the exponential decay of the subsystems passes to the joint system.

3.3. Inductive construction of domains of localization in parameter space

In view of the equivalence of the exponential decay in the Hausdorff distance of the eigenfunction correlator and of the fractional moments of the Green function, it is natural to define localization regimes as follows.

Definition 3.1. An open subset of the parameter space, $\mathcal{L} \subset \mathbb{R}_+ \times \mathbb{R}^p$ is said to be a *domain of uniform n -particle localization* if for some $s \in (0, 1)$ there exists $\xi < \infty$ and $A < \infty$ such that

$$(UL) \quad \sup_{\substack{I \subset \mathbb{R} \\ |I| \geq 1}} \sup_{\Lambda \subset \mathbb{Z}^d} \frac{1}{|I|} \int_I \mathbb{E} \left[\left| G_{\Lambda}^{(k)}(\mathbf{x}, \mathbf{y}; E) \right|^s \right] dE \leq A e^{-\text{dist}_{\mathcal{H}}(\mathbf{x}, \mathbf{y})/\xi}$$

holds for all $(\lambda, \alpha) \in \mathcal{L}$, all $k \in \{1, \dots, n\}$, and all $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}^d)^k$.

Our aim is to inductively construct $\mathcal{L}_n^{(p)}$ starting from the domain $\mathcal{L}_1^{(p)} = (\lambda_1, \infty) \subset \mathbb{R}_+$ of uniform one-particle localization, whose existence is guaranteed in [2] (and [15] for $d = 1$). For the induction step we assume that $\mathcal{L}_{n-1}^{(p)}$ is a domain of uniform $(n-1)$ -particle localization and proceed by distinguishing two cases:

- (i) **Localization for non-clustered configurations.** Here we deal with two configuration $\mathbf{x}, \mathbf{y}$ of which at least one is of diameter ($\text{diam}(\mathbf{x}) := \max_{j,k} |x_j - x_k|$) comparable with their Hausdorff distance. A key lemma provides the bound:

$$\sup_{\substack{I \subset \mathbb{R} \\ |I| \geq 1}} \sup_{\Lambda \subseteq \mathbb{Z}^d} \widehat{\mathbb{E}}_I \left[|G_\Lambda^{(n)}(\mathbf{x}, \mathbf{y})|^s \right] \leq A \exp \left(-\frac{1}{\xi} \min \left\{ \text{dist}_{\mathcal{H}}(\mathbf{x}, \mathbf{y}), \frac{\max\{\text{diam}(\mathbf{x}), \text{diam}(\mathbf{y})\}}{n-1} \right\} \right) \quad (12)$$

for all $(\lambda, \alpha) \in \mathcal{L}_{n-1}^{(p)}$, (which, strictly speaking, is only valid under the additional assumption of $\mathcal{L}_{n-1}^{(p)}$ being sub-conical, cf. [5]).

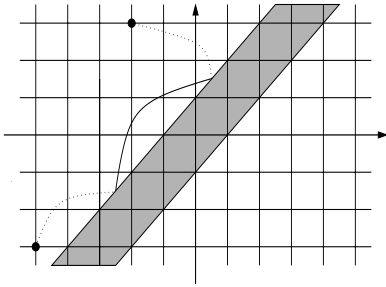


Fig. 2. The solid line is the full and the dotted lines represent the Green function of the partially non-interacting system which, by assumption and due to arguments based on (11) decay exponentially. Due to finite range, the interaction is localized to the shaded strip of width $2\ell_U$ about the diagonal.

The proof of (12) proceeds by breaking the configuration in two parts and removing the *intercluster* interaction. Standard perturbation theory then yields an expansion in U which in case $n = 2$, $d = 1$ is pictorially explained in *configuration space* in Fig 2.

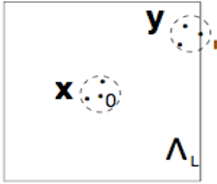
The bound (12) is rather crude in its dependence on the particle number. It is the main reason why our bounds on the localization length degrade rapidly with n .

- (ii) **Localization for clustered configurations.** This refers to configurations with diameter less than half their separation. The issue which is to be addressed is the possible formation of a quasi-particle which is not constrained by previous localization bounds.

A convenient quantity to monitor is

$$B_s^{(n)}(L) := \sup_{\substack{I \subset \mathbb{R} \\ |I| \geq 1}} \sup_{\Omega \subseteq \Lambda_L} |\partial \Lambda_L| \sum_{y \in \partial \Lambda_L} \sum_{\substack{\mathbf{x} \in \mathcal{C}_L^{(n)}(\Omega; 0) \\ \mathbf{y} \in \mathcal{C}_L^{(n)}(\Omega; y)}} \widehat{\mathbb{E}}_I \left[|G_\Omega^{(n)}(\mathbf{x}, \mathbf{y})|^s \right], \quad (13)$$

where $\mathcal{C}_L^{(n)}(\Lambda; x)$ denotes the collection of configurations of diameter less than $L/2$ and at least one particle at x .



- one has at least one particle at 0, and
- the other one has at least one particle at the boundary $\partial\Lambda_L$ of the box; cf. illustration in *real space*.

Localization bounds for clustered configurations proceed through *rescaling inequalities* for $B_s^{(n)}(L)$. The key statement is:

Lemma 3.3. *There exists $s \in (0, 1)$, $a, A, p < \infty$, and $\nu > 0$ such that*

$$B_s^{(n)}(2L) \leq \frac{a}{|\lambda|^s} B_s^{(n)}(L)^2 + A L^{2p} e^{-2\nu L} \quad (14)$$

for all $(\lambda, \alpha) \in \mathcal{L}_{n-1}^{(p)}$, (which is again assumed to be sub-conical, cf. [5]).

Part of the proof of this lemma resembles the proof of finite-volume criterion for one particle localization in [3]. In view of the infinite-range correlations of the random potential in configuration space discussed at the beginning of this section, a crucial observation is the following. When placing two boxes of length $L/2$ about configurations $\mathbf{x}$ and $\mathbf{y}$ in the above picture, the random variables associated with those boxes are independent due to the spatial separation. Moreover, the error when restricting the sum in the definition of $B_s^{(n)}(2L)$ to configurations $\mathbf{x}, \mathbf{y}$ which have a diameter less than $L/4$ can be controlled by (12) yielding the second term on the right side in (14).

It is not hard to see that rescaling inequalities such as (14) imply exponential decay provided the quantity $\frac{a}{|\lambda|^s} B_s^{(n)}(L)$ is small on some scale L . This is the requirement which determines $\mathcal{L}_n^{(p)}$. Namely,

- in case of *strong disorder* and arbitrary value of the interaction α we choose L fixed and λ small enough.
- in case of *weak interaction* and arbitrary value of the disorder $\lambda > \lambda_1$ we appeal to [1, 3] which imply that $B_s^{(n)}(L) \rightarrow 0$ as $L \rightarrow \infty$ for $\alpha = \mathbf{0}$. Since the finite-volume quantity $B_s^{(n)}(L)$ is continuous in some neighborhood of $\alpha = \mathbf{0}$ we choose L and α accordingly.

4. Some remaining challenges

While Theorem 2.1 is formulated for interactions of finite range, its proof allows for extension to interactions of with exponential falloff. However, it does not address questions about the effects of Coulomb interactions.

An important outstanding challenge it to resolve the question which is commented upon in Section 2.3: does localization persists (with uniform bounds) even for large systems at a positive density of particles?

Acknowledgments

This work was partially supported by the National Science Foundation under grants DMS-0602360 (MA), DMS-0701181 (SW) and a Sloan Fellowship (SW).

References

- [1] M. Aizenman and S. Molchanov, Localization at large disorder and at extreme energies: an elementary derivation, *Comm. Math. Phys.*, **157** 245–278 (1993).
- [2] M. Aizenman, Localization at weak disorder: some elementary bounds, *Rev. Math. Phys.* **6**, 1163–1182 (1994).
- [3] M. Aizenman, J. H. Schenker, R. M. Friedrich, and D. Hundertmark, Finite-volume criteria for Anderson localization. *Comm. Math. Phys.* **224** 219–253 (2001).
- [4] M. Aizenman, A. Elgart, S. Naboko, J. H. Schenker, and G. Stolz, Moment analysis for localization in random Schrödinger operators, *Invent. Math.* **163**, 343–413 (2006).
- [5] M. Aizenman and S. Warzel, Localization Bounds for Multiparticle Systems, *Commun. Math. Phys.* **293**, 903–934 (2009).
- [6] M. Aizenman and S. Warzel, Random Schrödinger Operators, Lecture Notes, in preparation.
- [7] A. Aspect and M. Inguscio, Anderson localization of ultracold atoms, *Phys. Today* August 2009.
- [8] D. M. Basko, I. L. Aleiner, and B. L. Altshuler, Metal-insulator transition in a weakly interacting many-electron system with localized single-particle states, *Annals of Physics* **321** 1126–1205 (2006).
- [9] V. Chulaevsky, and Y. M. Suhov, Wegner bounds for a two-particle tight binding model, *Comm. Math. Phys.* **283**, 479–489 (2008).
- [10] V. Chulaevsky, and Y. M. Suhov, Eigenfunctions in a two-particle Anderson tight binding model, *Comm. Math. Phys.* **289**, 701–723 (2009).
- [11] V. Chulaevsky and Y. Suhov, Multi-particle Anderson localisation: Induction on the number of particles, *Math. Phys. Anal. Geom.* **12**, 117–139 (2009).
- [12] J. Fröhlich and T. Spencer, *Absence of diffusion in the Anderson tight binding model for large disorder or low energy*, *Comm. Math. Phys.* **88**, 151–184 (1983).
- [13] W. Kirsch, A Wegner estimate for multi-particle random Hamiltonians, *J. Math. Phys. Anal. Geom.* **4**, 121–127 (2008).
- [14] W. Kirsch, An invitation to random Schrödinger operators (with appendix by F. Klopp). pp. 1–119 in: *Random Schrödinger Operators*. M. Disertori, W. Kirsch, A. Klein, F. Klopp, V. Rivasseau, Panoramas et Synthèses 25 (2008).
- [15] H. Kunz and B. Souillard, *Sur le spectre des opérateurs aux différences finies aléatoires*, *Commun. Math. Phys.* **78**, 201–246 (1980).
- [16] D. L. Shepelyansky, Three-dimensional Anderson transition for two electrons in two dimensions, *Phys. Rev. B*, **61** 4588 – 4591 (2000).
- [17] P. Stollmann, *Caught by disorder: bound states in random media*, (Birkhäuser, 2001).
- [18] W.-M. Wang and Z. Zhang, Long time Anderson localization for the nonlinear random Schrödinger equation, *J. Stat. Phys.*, **134**, 953–968 (2009).

OTHER TALKS

INVITED TALKS

FROM AREA LAWS TO THE SIMULATION OF QUANTUM SYSTEMS

MARTIN PLENIO

Imperial College London

E-mail: m.plenio@imperial.ac.uk

Physical interactions in quantum many-body systems are typically local: Individual constituents interact mainly with their few nearest neighbors. This locality of interactions is inherited by a decay of correlation functions, but also reflected by scaling laws of the entanglement entropy of ground states. This entropy of the reduced state of a subregion often merely grows like the boundary area of the subregion, and not like its volume, in sharp contrast to an extensive behavior. Such “area laws” for the entanglement entropy and related quantities have received considerable attention in recent years initiated by [1]. They emerge in several seemingly unrelated fields, in the context of black hole physics, quantum information science, and quantum many-body physics. In this talk I will present an overview over recent results on this topic. Center stage is taken by rigorous results on lattice models in one and higher spatial dimensions [2,3,4] and the connection between the entanglement content of states and the possibility of their efficient numerical simulation as they were recently reviewed in [5].

- [1] K. Audenaert, J. Eisert, M.B. Plenio and R.F. Werner, Phys.Rev. A 66, 042327 (2002)
- [2] M.B. Plenio, J. Eisert, J. Dreissig, and M. Cramer, Phys. Rev. Lett. 94, 060503 (2005)
- [3] M. Cramer, J. Eisert, M.B. Plenio, and J. Dreissig, Phys. Rev. A 73, 012309 (2006)
- [4] M. Cramer, J. Eisert, M.B. Plenio, Phys. Rev. Lett. 98, 220603 (2007)
- [5] J. Eisert, M. Cramer, and M.B. Plenio, to appear in Rev. Mod. Phys. 2009, see also e-print arXiv:0808.3773 [quant-ph]

CONTRIBUTED TALKS

AHARONOV-BOHM ANSATZ AND TONOMURA ET AL. EXPERIMENTS: RIGOROUS PROOFS

MIGUEL BALLESTEROS, RICARDO WEDER

INRIA, Paris, and UNAM, Mexico City

E-mail: Miguel.Ballestros_Montero@inria.fr

The Aharonov-Bohm effect is a fundamental issue in physics, it describes the physical important electromagnetic quantities in quantum mechanics and its experimental verification constitutes a test of the theory of quantum mechanics itself. In the fundamental paper of Aharonov and Bohm (1959), they provide an approximate solution to the Schrödinger equation when the electron stays in simple connected regions where the magnetic field is zero, by a change of gauge formula from the zero vector potential. The formula that they give shows explicitly the dependence on the magnetic potential over regions where the magnetic field is zero. The formula of Aharonov and Bohm was experimentally confirmed by the fundamental experiments of Tonomura et al., giving a conclusive evidence of the existence of the Aharonov-Bohm effect. In this talk we provide a rigorous proof that the Aharonov-Bohm formula is a good approximation in the case of the Tonomura et al. experiments. We give a simple (uniform in time) formula for the error bound that estimates the difference between the Aharonov-Bohm approximated solution and the exact solution to the Schrödinger equation and we show that under certain properties of the electron, the Aharonov-Bohm solution differs from the real one by 10^{-314} .

ONE-SHOT QUANTUM CAPACITIES OF QUANTUM CHANNELS

FRANCESCO BUSCEMI, NILANJANA DATTA

Statistical Laboratory, University of Cambridge

E-mail: buscemi@statslab.cam.ac.uk

We consider the protocol in which Alice sends one part of a maximally entangled state through a quantum channel to Bob, who then performs a quantum operation on the received state, with the final objective of obtaining a nearly maximally entangled state, shared with Alice. We find that the one-shot capacity of this protocol is characterized by appropriate smoothing of the 0-conditional Rényi entropy. This in turn provides a characterization of the one-shot quantum capacity of the channel. In the limit of asymptotically many uses of a memoryless channel, we recover the familiar expression of the quantum capacity given by the regularized coherent information. An important off-shoot of our result is that it also yields an expression for the asymptotic quantum capacity of an arbitrary sequence of channels, possibly with memory.

SIGNATURES OF NONCLASSICALITY IN MIXED-STATE QUANTUM COMPUTATION

ANIMESH DATTA*, SEVAG GHARIBIAN, ANIL SHAJI, CARLTON M. CAVES

**Imperial College, London*

E-mail: a.datta@imperial.ac.uk

We investigate signatures of non-classicality in quantum states, in particular, those involved in the DQC1 model of mixed-state quantum computation [Phys. Rev. Lett. 81, 5672 (1998)]. To do so, we consider two known non-classicality criteria. The first quantifies disturbance of a quantum state under locally noneffective unitary operations (LNU), which are local unitaries acting invariantly on a subsystem. The second quantifies measurement induced disturbance (MID) in the eigenbasis of the reduced density matrices.

We study the role of both figures of non-classicality in the exponential speedup of the DQC1 model and compare them vis-a-vis the interpretation provided in terms of quantum discord. In particular, we prove that a non-zero quantum discord implies a non-zero shift under LNUs. We also use the MID measure to study the locking of classical correlations [Phys. Rev. Lett. 92, 067902 (2004)] using two mutually unbiased bases (MUB). We find the MID measure to exactly correspond to the number of locked bits of correlation. For three or more MUBs, it predicts the possibility of superior locking effects.

FROM FDE THROUGH ODE TO PDE: UNIFORM EXISTENCE OF THE IDS

MICHAEL J. GRUBER*, DANIEL LENZ, IVAN VESELIĆ

* *TU Clausthal*

E-mail: gruber@math.tu-clausthal.de

The integrated density of states (IDS) is an important characteristic of random operators. It is often defined by a limiting process, although typically the limit exists only almost surely and only for certain values of the spectral parameter (points of continuity of the IDS). We show that for a specific class of random models the IDS exists as a uniform limit, where the notion of uniformity depends on the type of random operator: finite difference operator (combinatorial graph), ordinary differential operator (metric graph), and partial differential operator (on $\mathbb{R}^d$), leading to globally uniform, weighted globally or locally uniform, and locally averaged uniform convergence. Applications include various flavours of percolation models.

DIFFUSION OF WAVEPACKETS IN A MARKOV RANDOM POTENTIAL

YANG KANG, JEFFREY SCHENKER

Michigan State University

E-mail: ykang@math.msu.edu

We consider the evolution of a tight binding wave packet propagating by the Schrödinger equation with a time dependent random potential. If the potential evolves according to a stationary Markov process, we show that the square amplitude of the wave packet converges, after diffusive rescaling, to a solution of a heat equation.

SPECTRAL ANALYSIS OF 1-D SCHRÖDINGER OPERATORS WITH LOCAL POINT INTERACTIONS

ALEKSEY S. KOSTENKO*, MARK M. MALAMUD

* *Dublin Institute of Technology*

E-mail: duzer80@mail.ru

Consider formal differential expressions

$$\ell_{X,\alpha} := -\frac{d^2}{dx^2} + \sum_{n \in \mathbb{Z}} \alpha_n \delta(x - x_n), \quad \ell_{X,\beta} := -\frac{d^2}{dx^2} + \sum_{n \in \mathbb{Z}} \beta_n \delta'(x - x_n), \quad \alpha_n, \beta_n \in \mathbb{R},$$

where $\delta(\cdot)$ is a Dirac delta-function and $\{x_n\}_{n \in \mathbb{Z}}$ is a strictly increasing sequence satisfying $x_n \rightarrow \pm\infty$ as $n \rightarrow \pm\infty$. Let $H_{X,\alpha}$ and $H_{X,\beta}$ be the minimal operators associated in $L^2(\mathbb{R})$ with the expressions $\ell_{X,\alpha}$ and $\ell_{X,\beta}$, respectively. The spectral properties of these operators are well studied in the case when the interactions sites are uniformly distributed, $d_* := \inf_{n \neq k} |x_n - x_k| > 0$ (numerous results as well as a comprehensive list of references might be found in the monograph of Albeverio, Gesztesy, Hoegh-Krohn, Holden, *Solvable Models in Quantum Mechanics*, AMS Chelsea Publ., 2005; see also a survey of recent result given by Exner in Appendix K).

If the assumption $d_* > 0$ is dropped, then the spectral analysis of operators $H_{X,\alpha}$ and $H_{X,\beta}$ becomes more complicated. For instance, for the operator $H_{X,\alpha}$ it is known only that it might be symmetric with nontrivial deficiency indices (the example of Shubin, Christ and Stolz, *J. Math. Anal. Appl.* **184** (1994)).

The main aim of our talk is the spectral analysis of operators $H_{X,\alpha}$ and $H_{X,\beta}$ in the case $d_* = 0$. We show that spectral properties of the operators $H_{X,\alpha}$ and $H_{X,\beta}$ are closely related with the spectral properties of certain classes of unbounded Jacobi matrices. We exploit this connection to investigate self-adjointness, lower semiboundedness, and discreteness of operators with local point interactions.

ASYMPTOTIC HYPOTHESIS TESTING FOR CORRELATED QUANTUM STATES

N. DATTA, F. HIAI, M. HAYASHI, M. FANNES, M. MOSONYI*, T. OGAWA

*Centre for Quantum Technologies, National University of Singapore
E-mail: milan.mosonyi@gmail.com

Asymptotic hypothesis testing in its simplest form is about discriminating two states of a lattice system, based on measurements on finite blocks that asymptotically cover the whole lattice. In general, it is not possible to discriminate the local states with certainty, and one's aim is to minimize the probability of error, subject to certain constraints. Hypothesis testing results show that, in various settings, the error probabilities vanish with an exponential speed, and the decay rates coincide with certain relative-entropy like quantities. Apart from giving computable closed expressions for the error exponents, the importance of these results lies in providing an operational interpretation for the given relative entropy-like quantities. Settings of main interest are that of Stein's lemma as well as the Chernoff and the Hoeffdings bounds. Here I give an overview on the general method to handle such problems on quantum lattice systems and show applications to state discrimination problems on quasifree fermionic and bosonic lattices, Gibbs states of spin chains and group symmetry-restricted discrimination problems. I also show an application of state discrimination results to derive a lower bound on the classical capacity for finitely many uses of a quantum channel. The talk is based on the works

- [1] F. Hiai, M. Mosonyi, T. Ogawa: Error exponents in hypothesis testing for correlated states on a spin chain; *J. Math. Phys.* **49**, 032112, (2008),
- [2] M. Mosonyi, F. Hiai, T. Ogawa, M. Fannes: Asymptotic distinguishability measures for shift-invariant quasi-free states of fermionic lattice systems; *J. Math. Phys.* **49**, 072104, (2008),

- [3] M. Mosonyi: Hypothesis testing for Gaussian states on bosonic lattices; J. Math. Phys. 50, 032105 (2009),
- [4] M. Mosonyi, N. Datta: Generalized relative entropies and the capacity of classical-quantum channels; arXiv:0810.3478,
- [5] M. Hayashi, F. Hiai, M. Mosonyi: Hypothesis testing under group symmetry; in preparation

VERTEX COUPLINGS IN QUANTUM GRAPHS

ONDŘEJ TUREK*, PAVEL EXNER, TAKSU CHEON

* *Department of Mathematics, FNSPE, Czech Technical University*

E-mail: turekond@fjfi.cvut.cz

(i) It is a well known fact that a general vertex coupling in a quantum graph vertex with n outgoing edges can be parametrized by n^2 real parameters. In 1999, Kostykin and Schrader introduced a concrete parametrization, simple and clear, but with the defect of ambiguity. A year later, independently Harmer and Kostykin & Schrader have shown that this parametrization can be made unique using a unitary matrix. However, in this unique form it is extremely difficult to see which role each of the n^2 real parameters plays, and moreover, even the couplings that are simple from the physical point of view, for example the δ coupling, have a very complicated expression here. In this work we suggest an alternative way how to parametrize vertex couplings that combines the advantages of the both ways mentioned above. Namely, our parametrization is unique and at the same time it is relatively simple to understand the role of the involved parameters.

(ii) Although a general vertex coupling is well explored from the mathematical point of view, the explanation of the physical meaning of the full family of vertex couplings has not been given yet. Here we propose the following solution. We begin with a star graph with n arms and a *general* coupling in the vertex - the *approximated* system. Then we consider a system of n half lines whose endpoints are connected by lines of a length d such that there are δ couplings and δ interactions placed in the endpoints of the half lines and in the centers of the connecting lines; at the same time, the connecting lines are supposed to support vector potentials - the *approximating* system. Then we show that the approximated system can be considered as a limit of the family of approximating systems as $d \rightarrow 0$, and we support our claim by a proof that this convergence is in the norm-resolvent sense.

POSTERS

L'UBOMÍRA BALKOVÁ (Department of Mathematics, Czech Technical University, Prague):
SCHRÖDINGER OPERATORS ASSOCIATED WITH PARRY NUMBERS

IRINA V. BLINOVA (St.-Petersburg State University of Information Technologies, Mechanics and Optics), with IGOR S. LOBANOV, VLADIMIR YU. LOTOREYCHIK, and IGOR YU. POPOV: **QUANTUM GRAPH OF SIERPINSKI GASKET TYPE: MODEL OF SCATTERING**

THOMAS BROUGHAM (Doppler Institute, Czech Technical University), with IGOR JEX and G.M. NIKOPOULOS: **PERFECT STATE TRANSFER IN NETWORKS OF ARBITRARY TOPOLOGY**

CLAUDIO CACCIAPUOTI (Hausdorff Center for Mathematics Bonn), with RAFAELLE CARLONE, GIANFAUSTO DELL' ANTONIO, RODOLFO FIGARI, DOMENICO FINCO, and ALESSANDRO TETA: **THE MOTT'S CONJECTURE**

SEBASTIAN ENDRES (Institut für Theoretische Physik, Universität Ulm), with FRANK STEINER: **BERRY-KEATING OPERATOR ON COMPACT GRAPHS**

AURÉL GÁBRIS (Doppler Institute, Prague), with VÁCLAV POTOČEK, TAMÁS KISS, and IGOR JEX: **OPTIMIZING THE QUANTUM RANDOM WALK SEARCH ALGORITHM ON THE HYPERCUBE**

SYLVAIN GOLENIA (Universität Erlangen) with NABILE BOUSSAID: **LIMITING ABSORPTION PRINCIPLE FOR SOME LONG RANGE PERTURBATIONS OF DIRAC SYSTEMS AT THRESHOLD ENERGIES**

UWE GÜNTHER (Research Center Dresden-Rossendorf), with BORIS F. SAMSONOV: **ULTRA-FAST EVOLUTION VIA PT-SYMMETRY AND ENTANGLEMENT**

NINO KHATIASHVILI (I. Vekua Institute of Applied Mathematics, Tbilisi): **QUANTUM BILLIARD IN THE HEXAGONAL LATTICED DOMAIN**

HYNEK LAVIČKA (FNSPE, Czech Technical University), with IGOR JEX and MARTIN ŠTEFAŇÁK: **QUANTUM RANDOM WALK WITH JUMPS**

JIŘÍ LIPOVSKÝ (Doppler Institute and Charles University, Prague), with PAVEL EXNER: **RESONANCES FROM PERTURBATIONS OF QUANTUM GRAPHS WITH RATIONALLY RELATED EDGES**

RUBEN MARES (Escuela Superior de Física y Matemáticas – IPN), with ARES DE PARGA: **QUANTUM PROPAGATOR VIA THE SCHWINGER ACTION PRINCIPLE**

TAKUYA MINE (Kyoto Institute of Technology): **AHARONOV-BOHM SOLENOIDS ON THE HYPERBOLIC PLANE**

JAN NAUDTS (Universiteit Antwerpen), with TOBIAS VERHULST, and BEN ANTHONIS: **COUNTING OPERATOR ANALYSIS OF THE DISCRETE SPECTRUM OF SOME MODEL HAMILTONIANS**

HAGEN NEIDHARDT (WIAS Berlin), with PAVEL EXNER: **TROTTER-KATO PRODUCT FORMULA FOR IMAGINARY TIMES**

ROGER NICHOLS (University of Alabama at Birmingham), with GÜNTHER STOLZ: **THE SPECTRAL MINIMUM OF THE DISCRETE DISPLACEMENT MODEL**

JOSE JUAN PEÑA (Univ. Autònoma Metropolitana), with J. MORALES, CH. PACHECO-GARCÍA, and J. GARCÍA-RAVELO: **A HARMONIC-LIKE POTENTIAL AND THE SCHRÖDINGER EQUATION WITH POSITION-DEPENDENT MASS**

IGOR YU. POPOV (St.-Petersburg State University of Information Technologies, Mechanics and Optics): **SCATTERING IN NANOSTRUCTURES AND QUANTUM GATES**

LESZEK SIRKO (Institute of Physics, Polish Academy of Sciences), with MICHAŁ LAWNICZAK, OLEH HUL, and SZYMON BAUCH: **INVESTIGATION OF GRAPHS WITH AND WITHOUT TIME REVERSAL SYMMETRY**

MARTIN ŠTEFAŇÁK (Czech Technical University, Prague), with IGOR JEX and TAMÁS KISS: **RECURRENCES IN QUANTUM WALKS**

MILOŠ TATER (Nuclear Physics Institute AS CR, Řež), with BORIS SHAPIRO and KOUICHI TAKEMURA: **ON SPECTRAL POLYNOMIALS OF THE HEUN EQUATION**

RAFAEL TIEDRA DE ALDECOA (Pontifical Catholic University of Chile): **TIME DELAY FOR DISPERSIVE QUANTUM HAMILTONIANS**

JIRÍ TOLAR (Czech Technical University, Prague), with GOCE CHADZITASKOS: **FEYNMAN'S PATH INTEGRAL AND MUTUALLY UNBIASED BASES**

MATĚJ TUŠEK (Czech Technical University, Prague), with PIERRE DUCLOS and PAVEL ŠTOVÍČEK: **ON A HYDROGEN ATOM IN A THIN SLAB**

NAOMASA UEKI (Graduate School of Human and Environmental Studies, Kyoto University), with RYOKI FUKUSHIMA: **ASYMPTOTICS OF IDS FOR A RANDOMLY PERTURBED LATTICE**

TOMIO UMEDA (University of Hyogo), with YOSHIMI SAITO: **EIGENFUNCTIONS AT THE THRESHOLD ENERGIES OF MAGNETIC DIRAC OPERATORS WITH POSITIVE MASS**

OSANOBU YAMADA (Ritsumeikan University), with HIROSHI ITO: **DIRAC OPERATORS WITH DILATION ANALYTIC POTENTIALS DIVERGING AT INFINITY**

MILOSLAV ZNOJIL (Nuclear Physics Institute ASCR, Řež): **THREE-HILBERT-SPACE FORMULATION OF QUANTUM MECHANICS**

String Theory and Quantum Gravity

Session organizers:

Gary Gibbons
Hermann Nicolai

EXTREMAL BLACK HOLES AND NILPOTENT ORBITS

GUILLAME BOSSARD

*AEI, Max-Planck-Institut für Gravitationsphysik
Am Mühlenberg 1, D-14476 Potsdam, Germany
E-mail: bossard@aei.mpg.de*

The stationary solutions of a large variety of (super)gravity theories can be described within a non-linear sigma model $\mathfrak{G}/\mathfrak{H}^*$ coupled to Euclidean gravity in three-dimensions, for which $\mathfrak{G}$ is a simple group and $\mathfrak{H}^*$ a non-compact real form of its maximal compact subgroup. The absence of naked singularities in four dimensions requires the $\mathfrak{G}$ Noether charge in 3D to satisfy a characteristic equation that determines it in function of the mass, the NUT charge and the electro-magnetic charges of the solution. It follows that the Noether charge associated to extremal black holes must lie in a certain Lagrangian submanifold of a nilpotent orbit of $\mathfrak{G}$. Constructing a suitable parameterisation of this Lagrangian, we are able to determine the so-called ‘fake superpotential’ that governs the radial dependency of the scalar fields.

Keywords: Supergravity, black holes, Lie algebras.

In trying to extend our quantum gravity understanding of black holes, extremality is often a key simplifying assumption. Firstly, it eliminates Hawking radiation and ensures that the solution is semi-classically stable. Secondly, it guarantees that the near-horizon solution is entirely determined by the conserved charges measurable at spatial infinity, and therefore insensitive (away from lines of marginal stability) to variations to the asymptotic value of the scalar fields at infinity. This attractor behaviour, first discovered for supersymmetric (BPS) black holes [1, 2], holds for all extremal solutions [3–5], and is arguably responsible for the validity of certain weakly coupled description of non-BPS black hole micro-states [6] in string theory.

Considering supergravity theories, black holes are solutions of the Einstein equations coupled to abelian vector fields and scalar fields parameterising a Riemannian space $\mathcal{M}_4$. Assuming spherical symmetry and extremality, the supergravity equations of motion become equivalent to light-like geodesic motion on a pseudo-Riemannian space $\mathcal{M}_3^*$, with the affine parameter τ identified as the inverse radial distance $\tau = 1/r$. Restricting ourselves to static solutions, or equivalently considering asymptotically Minkowski solutions, the equations of motion reduce furthermore to the motion of a fiducial particle on $\mathbb{R}_+^* \times \mathcal{M}_4$ subject to a negative definite V depending on the electromagnetic charges. There are two main approaches to determine such solutions that we combined in [7].

The first one consists in reducing the equations of motion to first order gradient flow equations through the determination of a ‘fake superpotential’ W satisfying

$$V = -e^{2U} (W^2 + 2g^{ij}\partial_i W \partial_j W) \quad (1)$$

where g_{ij} is the Riemannian metric on $\mathcal{M}_4$ and U is the scale function defining the extremal static spherically symmetric metric Ansatz

$$ds^2 = -e^{2U} dt^2 + e^{-2U} (dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2)) \quad (2)$$

Whenever the solution is BPS, *i.e.* that it preserves some supersymmetry generators, the first order gradient flow equations correspond to the vanishing of the supersymmetry transformation of the fermion fields. This is most familiar in the framework of $D = 4, \mathcal{N} = 2$ supergravity, for which the BPS superpotential is determined in function of the central charge Z as

$$W_{\text{BPS}} = |Z_{p,q}(\phi^i)| \quad (3)$$

This permitted to determine the most general BPS solutions explicitly [8, 9]. However, the solution to (1) is not unique, and the determinations of inequivalent ‘fake superpotentials’ W were obtained in [10–17].

On the other-hand, when $\mathcal{M}_4$ is a symmetric space $\mathfrak{G}_4/\mathfrak{H}_4$, such that its isometry group $\mathfrak{G}_4$ defines a symmetry of the theory by acting faithfully on the electromagnetic field strength, the pseudo-Riemannian manifold $\mathcal{M}_3^*$ on which is defined the light-like geodesic motion is itself a symmetric space $\mathfrak{G}/\mathfrak{H}^*$ [18], with simple isometry group $\mathfrak{G}$ and a non-compact maximal subgroup $\mathfrak{H}^*$ which is a non-compact real form of the maximal compact subgroup of $\mathfrak{G}$. The geodesic motion on $\mathcal{M}_3^*$ is then integrable, and in fact all geodesics on $\mathcal{M}_3^*$ can be obtained by exponentiating a generator $-P_0\tau \in \mathfrak{g} \ominus \mathfrak{h}^*$, where P_0 determines the momentum along the trajectory. P_0 is conjugate to the Noether charge Q via the coset representative $\mathcal{V}$ in $\mathfrak{G}/\mathfrak{H}^*$,

$$P \equiv -\left(\mathcal{V}^{-1} \dot{\mathcal{V}}\right)\big|_{\mathfrak{g} \ominus \mathfrak{h}^*} = \mathcal{V}^{-1} Q \mathcal{V} \quad (4)$$

where the dot denotes the derivative with respect to τ . Extremal solutions correspond to special geodesics which reach the boundary $U = -\infty$ in infinite proper time [18]. It is necessary but not sufficient that the geodesic be light-like. For BPS black holes, it was observed in [19] that the Noether charge must satisfy $[\text{ad}(Q)]^5 = 0$, *i.e.* belong to a nilpotent orbit of degree 5. More recently, the supersymmetry and extremality conditions on the Noether charge for symmetric supergravity models were re-analyzed in [20]. It was shown in all cases where $\mathfrak{G}$ is simple that extremality requires to $[Q|_{\mathbf{R}}]^3 = 0$, where $\mathbf{R}$ denotes the “fundamental representation” of $\mathfrak{G}$: for example the spinor representation if $\mathfrak{G}$ is an orthogonal group $SO(2+m, 2+n)$ or $SO^*(2m+4)$. The only exception is for $\mathfrak{G} = E_{8(8)}$ or $E_{8(-24)}$, where the condition becomes $[Q|_{\mathbf{3875}}]^5 = 0$, with **3875** being the 3875-dimensional irreducible representation appearing in the symmetric tensor product of two adjoints.

More precisely, any generic extremal spherically symmetric black hole (*i.e.* with a non-zero horizon area) is characterized by a nilpotent Noether charge Q which lies inside the grade-two component $\mathfrak{l}_4^{(2)}$ of $\mathfrak{g}$ with respect to the 5-grading (more appropriately, even 9-grading) which arises in the reduction from 4 to 3 dimensions:

$$\mathfrak{g} \cong \mathbf{1}^{(-4)} \oplus \mathfrak{l}_4^{(-2)} \oplus (\mathfrak{gl}_1 \oplus \mathfrak{g}_4)^{(0)} \oplus \mathfrak{l}_4^{(2)} \oplus \mathbf{1}^{(4)} \quad (5)$$

The nilpotent orbit $\mathcal{O}_{\mathfrak{G}}$ of $Q \in \mathfrak{g}$ under $\mathfrak{G}$ is characterized by the isotropy subgroup ^{a} of Q in $\mathfrak{G}$. On the other hand, the momentum P_0 is valued in the coset $\mathfrak{g} \ominus \mathfrak{h}^*$, and therefore defines a $\mathfrak{H}^*$ -orbit $\mathcal{O}_{\mathfrak{H}^*}$ inside $\mathfrak{g} \ominus \mathfrak{h}^*$.

Since the coset component of the Maurer–Cartan form is conjugate to the Noether charge via $P = \mathcal{V}^{-1}Q\mathcal{V}$, it defines a representative $\mathbf{e} \equiv P$ of the corresponding nilpotent orbit inside the coset component $\mathfrak{g} \ominus \mathfrak{h}^*$, and therefore defines a $\mathfrak{H}^*$ -orbit inside this coset. A general fact about nilpotent elements is that one can always find another nilpotent element $\mathbf{f}$ and a semi-simple generator $\mathbf{h}$ such that the triplet $(\mathbf{e}, \mathbf{f}, \mathbf{h})$ defines an $\mathfrak{sl}_2$ subalgebra of $\mathfrak{g}$, *i.e.*

$$[\mathbf{e}, \mathbf{f}] = \mathbf{h} \quad [\mathbf{h}, \mathbf{e}] = 2\mathbf{e} \quad [\mathbf{h}, \mathbf{f}] = -2\mathbf{f} \quad (6)$$

The eigenspaces of $\mathbf{h}$ furnish a graded decomposition of $\mathfrak{g}$ which uniquely characterizes the complex nilpotent $\mathfrak{G}_{\mathbb{C}}$ orbit [24]. Extremal solutions are such that the $\mathfrak{H}^*$ -orbit of P is characterized by a graded decomposition of $\mathfrak{h}^*$ of the same form as (5) [25],

$$\mathfrak{h}^* \cong \mathfrak{h}^{(-4)} \oplus \mathfrak{h}^{(-2)} \oplus \mathfrak{gl}_1 \oplus \mathfrak{h}^{(0)} \oplus \mathfrak{h}^{(2)} \oplus \mathfrak{h}^{(4)} \quad (7)$$

In the case of maximally supersymmetric supergravity [26], for static solutions (*i.e.* with zero NUT charge) the semi-simple element $\mathbf{h}$ associated to the nilpotent element P can be computed in terms of the central charges Z_{ij} alone, and more generally, in terms of the central and matter charges which we write collectively Z_I .^b Decomposing $P \in \mathfrak{g} \ominus \mathfrak{h}^*$ with respect to the Ehlers $U(1)$ and the four-dimensional R-symmetry group $\mathfrak{H}_4$ [7],

$$P = -\dot{U}\mathbf{H} + e^U Z_I \mathbf{L}^I - e_i{}^j \dot{\phi}^i \mathbf{G}_j \quad \in \quad \mathbb{C} \oplus \mathfrak{l}_4 \oplus (\mathfrak{g}_4 \ominus \mathfrak{h}_4) \quad (8)$$

where $e_i{}^j$ is a vielbein for the metric g_{ij} , one may recast the middle equation in (6) into a system of first order differential equations of the form

$$\dot{U} = -e^U W \quad g_{ij} \dot{\phi}^j = -e^U W_i \quad (9)$$

where W and W_i depend on the moduli ϕ^i and electromagnetic charges Q_I through the charges Z_I only; moreover, we proved that [7]

$$W_i = \partial_{\phi^i} W \quad (10)$$

^aFor extremal black holes, this isotropy subgroup coincides with the isotropy subgroup of the electromagnetic charges in the four-dimensional duality group $\mathfrak{G}_4$ computed in [21–23].

^bHere Z_I are the scalar field dependent linear combinations of the electromagnetic charges, transforming in a complex representation of $\mathfrak{H}_4$ and such that $V_{\text{BH}} = Z_I Z^I$.

Thus, extremal solutions attached to the given nilpotent orbit satisfy a gradient flow under the fake superpotential W . In particular, it follows from the nilpotency of P that

$$\text{Tr } P^2 = 0 = e^{2U} \left(W^2 - Z_I Z^I + 2g^{ij} W_i W_j \right) = 0, \quad (11)$$

and therefore that (1) is obeyed.

Applying this strategy to $\mathcal{N} = 8$ supergravity with $\mathfrak{G} = E_{8(8)}$, we are able to determine the fake superpotentials for both BPS and non-BPS extremal black holes, and express them in terms of the $SU(8)$ invariant combinations of the central charges. In this case P_0 transforms as a Majorana–Weyl spinor under $Spin^*(16)$. It can be conveniently parameterised using a fermionic oscillator basis [20],

$$\begin{aligned} |P_0\rangle &= \left(w + Z_{ij} a^i a^j + \Sigma_{ijkl} a^i a^j a^k a^l + \frac{1}{6!} \varepsilon_{ijklmnpq} (Z^{pq} a^i \cdots a^n + \frac{1}{56} \bar{w} a^i \cdots a^q) \right) |0\rangle \\ &= (1 + \star) \left(w + Z_{ij} a^i a^j + \frac{1}{2} \Sigma_{ijkl} a^i a^j a^k a^l \right) |0\rangle \end{aligned} \quad (12)$$

where $\star$ is the anti-involution defining the chiral Majorana–Weyl representation of $Spin^*(16)$, $w = M + iN$ where M is the mass and N the NUT charge, Z_{ij} are the supersymmetric central charges and Σ_{ijkl} are the “scalar charges”.

There are two $E_{8(8)}$ orbits associated to the nilpotency condition $[Q]_{\mathbf{3875}}^{(5)} = 0$, whose union is dense in the space of solutions of this equation [20, 27]. They both lie in a single $E_8(\mathbb{C})$ orbit, associated to the same five graded decomposition,

$$\mathfrak{e}_{8(8)} \cong \mathbf{1}^{(-4)} \oplus \mathbf{56}^{(-2)} \oplus (\mathfrak{gl}_1 \oplus \mathfrak{e}_{7(7)})^{(0)} \oplus \mathbf{56}^{(2)} \oplus \mathbf{1}^{(4)} \quad (13)$$

A representative $\mathbf{E}$ of such a nilpotent orbit is a generic element of the grade two component $\mathbf{56}^{(2)}$. There are two classes of such elements which are distinguished by their isotropy subgroup inside $E_{7(7)}$, respectively $E_{6(2)}$ and $E_{6(6)}$ [21, 22]. The $\mathfrak{gl}_1$ generator $\mathbf{h}$ of $\mathfrak{so}^*(16)$ which defines a corresponding graded decomposition of $\mathfrak{so}^*(16)$ and its Majorana–Weyl representation $\mathbf{128}_+$ are such that a representative of the orbit lies in the component of grade two of $\mathbf{128}_+$

$$\mathbf{h} |P_0\rangle = 2 |P_0\rangle \quad (14)$$

The two orbits of $Spin^*(16)$ associated to generic extremal black holes (*i.e.* black holes with a non-vanishing horizon area) correspond to the 1/8 BPS and the non-BPS extremal black holes, respectively.

In order for the solution to be supersymmetric, the corresponding Noether charge state must satisfy the ‘Dirac equation’ [20]

$$(\epsilon_\alpha^i a_i + \varepsilon_{\alpha\beta} \epsilon_i^\beta a^i) |P_0\rangle = 0 \quad (15)$$

where a^i and a_j (for $i, j, \dots = 1, \dots, 8$) are the fermionic oscillators from which the spinor representations of $Spin^*(16)$ are built. In the case of a 1/8 BPS solution, this equation permits to determine the generator $\mathbf{h}$ from the value of the central charge as

$$\mathbf{h}_{\frac{1}{8}} \equiv e^{i\varphi} R_i{}^k R_j{}^l \omega_{kl}^o a^i a^j - e^{-i\varphi} R^i{}_k R^j{}_l \omega^{o\,kl} a_i a_j \quad (16)$$

for a central charge decomposing on a basis of real orthonormal antisymmetric tensor ω_{ij}^n as

$$Z_{ij} = \frac{1}{2} e^{i\varphi} R_i^k R_j^l \left(\rho_0 \omega_{ij}^0 + \rho_1 \omega_{ij}^1 + \rho_2 \omega_{ij}^2 + \rho_3 \omega_{ij}^3 \right) \quad (17)$$

such that $\rho_0 \geq \rho_1 \geq \rho_2 \geq \rho_3$. This determines the BPS ‘fake superpotential’ to be

$$W_{\text{BPS}} = \rho_0(Z_{ij}) \quad (18)$$

in agreement with [11].

In the non-BPS case, one use the fact that the generator $\mathfrak{h}$ also define the nilpotent elements associated to 1/2 BPS solutions by its grade four component, to use the ‘Dirac equation’ (15) to determine $\mathbf{h}$ in this case as well. As a result,

$$\mathbf{h} = \frac{1}{2 \cos(2\alpha)} \left(e^{-i\alpha} \tilde{R}_i^k \tilde{R}_j^l \Omega_{kl} a^i a^j - e^{i\alpha} \tilde{R}_k^i \tilde{R}_l^j \Omega^{kl} a_i a_j \right) + i \tan(2\alpha) (a^i a_i - 4) \quad (19)$$

for a central charge decomposing as

$$Z_{ij} = e^{-i\alpha} \tilde{R}_i^k \tilde{R}_j^l \left(\frac{1}{2} \left(e^{2i\alpha} + i \sin(2\alpha) \right) \varrho \Omega_{kl} + \Xi_{kl} \right) \quad (20)$$

such that Ω_{ij} is a symplectic form of overall phase factor $e^{i\frac{\pi}{4}}$ and Ξ_{ij} satisfies

$$\Xi_{ij} = \Omega_{ik} \Omega_{jl} \Xi^{kl} \quad \Omega^{ij} \Xi_{ij} = 0 \quad (21)$$

The ‘fake superpotential’ is thus defined in the non-BPS case by

$$W = 2\varrho(Z_{ij}) \quad (22)$$

ϱ being also a particular root of an irreducible sextic polynomial [7].

By truncation, one then obtains the ‘fake superpotential’ for all magic $\mathcal{N} = 2$ supergravity models, and in fact for all supergravity theories with $\mathcal{N} \geq 2$ with a symmetric moduli space.

References

- [1] S. Ferrara, R. Kallosh and A. Strominger, “ $\mathcal{N} = 2$ extremal black holes,” Phys. Rev. D **52** (1995) 5412 [[hep-th/9508072](#)].
- [2] S. Ferrara and R. Kallosh, “Universality of supersymmetric attractors,” Phys. Rev. D **54** (1996) 1525 [[hep-th/9603090](#)].
- [3] S. Ferrara, G. W. Gibbons and R. Kallosh, “Black holes and critical points in moduli space,” Nucl. Phys. B **500** (1997) 75 [[hep-th/9702103](#)].
- [4] A. Sen, “Black hole entropy function and the attractor mechanism in higher derivative gravity,” JHEP **0509** (2005) 038 [[hep-th/0506177](#)].
- [5] K. Goldstein, N. Iizuka, R. P. Jena and S. P. Trivedi, “Non-supersymmetric attractors,” Phys. Rev. D **72** (2005) 124021 [[hep-th/0507096](#)].
- [6] A. Dabholkar, A. Sen and S. P. Trivedi, “Black hole microstates and attractor without supersymmetry,” JHEP **0701** (2007) 096 [[hep-th/0611143](#)].
- [7] G. Bossard, Y. Michel and B. Pioline, “Extremal black holes, nilpotent orbits and the true fake superpotential,” [0908.1742](#) [[hep-th](#)].

- [8] F. Denef, “Supergravity flows and D-brane stability,” JHEP **0008** (2000) 050 [arXiv:hep-th/0005049].
- [9] B. Bates and F. Denef, “Exact solutions for supersymmetric stationary black hole composites,” arXiv:hep-th/0304094.
- [10] A. Ceresole and G. Dall’Agata, “Flow equations for non-BPS extremal black holes,” JHEP **0703**, 110 (2007) [hep-th/0702088].
- [11] L. Andrianopoli, R. D’Auria, E. Orazi and M. Trigiante, “First order description of black holes in moduli space,” JHEP **0711** (2007) 032 0706.0712 [hep-th].
- [12] G. Lopes Cardoso, A. Ceresole, G. Dall’Agata, J. M. Oberreuter and J. Perz, “First-order flow equations for extremal black holes in very special geometry,” JHEP **0710** (2007) 063 0706.3373 [hep-th].
- [13] J. Perz, P. Smyth, T. Van Riet and B. Vercnocke, “First-order flow equations for extremal and non-extremal black holes,” JHEP **0903** (2009) 150 0810.1528 [hep-th].
- [14] K. Hotta and T. Kubota, “Exact solutions and the attractor mechanism in non-BPS black holes,” Prog. Theor. Phys. **118** (2007) 969 0707.4554 [hep-th].
- [15] E. G. Gimon, F. Larsen and J. Simon, “Black holes in supergravity: the non-BPS branch,” JHEP **0801** (2008) 040 0710.4967 [hep-th].
- [16] S. Bellucci, S. Ferrara, A. Marrani and A. Yeranyan, “*stu* black holes unveiled,” 0807.3503 [hep-th].
- [17] E. G. Gimon, F. Larsen and J. Simon, “Constituent Model of Extremal non-BPS Black Holes,” JHEP **0907** (2009) 052 0903.0719 [hep-th].
- [18] P. Breitenlohner, D. Maison and G. W. Gibbons, “Four-dimensional black holes from Kaluza–Klein theories,” Commun. Math. Phys. **120** (1988) 295.
- [19] M. Günaydin, A. Neitzke, B. Pioline and A. Waldron, “BPS black holes, quantum attractor flows and automorphic forms,” Phys. Rev. D **73** (2006) 084019 [hep-th/0512296].
- [20] G. Bossard, H. Nicolai and K. S. Stelle, “Universal BPS structure of stationary supergravity solutions,” 0902.4438 [hep-th].
- [21] D. Ž. Đoković, “Classification of nilpotent elements in simple exceptional real Lie algebras of inner type and description of their centralizers,” J. of Algebra **112** (1988) 503.
- [22] S. Ferrara and M. Günaydin, “Orbits of exceptional groups, duality and BPS states in string theory,” Int. J. Mod. Phys. A **13** (1998) 2075 [hep-th/9708025].
- [23] S. Bellucci, S. Ferrara, M. Günaydin and A. Marrani, “Charge orbits of symmetric special geometries and attractors,” Int. J. Mod. Phys. A **21** (2006) 5043 [hep-th/0606209].
- [24] D. Collingwood and W. McGovern, “Nilpotent orbits in semisimple Lie algebras” Van Nostrand Reinhold Mathematics Series, New York, 1993.
- [25] G. Bossard and H. Nicolai, “Multi-black holes from nilpotent Lie algebra orbits,” 0906.1987 [hep-th].
- [26] E. Cremmer and B. Julia, “The $SO(8)$ Supergravity,” Nucl. Phys. B **159** (1979) 141.
- [27] D. Ž. Đoković, “The closure diagram for nilpotent orbits of the split real form of E_8 ,” CEJM **4** (2003) 573.

BLACK HOLE SPACETIMES WITH KILLING-YANO SYMMETRIES

DAVID KUBIZŇÁK

*DAMTP, University of Cambridge,
Wilberforce Road, Cambridge CB3 0WA, UK
E-mail: D.Kubiznak@damtp.cam.ac.uk
www.damtp.cam.ac.uk/people/d.kubiznak*

We present a brief overview of black hole spacetimes admitting Killing–Yano tensors. In vacuum these include Kerr–NUT–(A)dS metrics and certain black brane solutions. In the presence of matter fields, (conformal) Killing–Yano symmetries are known to exist for the Plebanski–Demianski solution and (trivially) for any spacetime with spherical symmetry. Special attention is devoted to generalized Killing–Yano tensors of black holes in minimal gauged supergravity. Several aspects directly related to the existence of Killing–Yano tensors—such as the Kerr–Schild form, algebraic type of spacetimes, and separability of field equations—are also briefly discussed.

Keywords: Killing–Yano symmetries, black holes, separability of field equations

1. Introduction

It is now more than forty years since hidden symmetries have been recognized to play an important role in black hole physics. The foundations of the subject dates back to 1968 and the seminal paper of Brandon Carter [1] who discovered a mysterious and unexpected constant of motion for geodesic trajectories in the rotating black hole spacetime of the Kerr geometry [2]. In the following years it was recognized that the Carter’s constant can be traced to the existence of hidden symmetries described by the *Killing* [3, 4] and *Killing–Yano* (KY) tensors [5–7]. Subsequent studies have also shown that these symmetries are deeply connected with complete integrability of geodesics motion, separability of fundamental field equations and existence of corresponding symmetry operators, various conserved quantities, worldline supersymmetry of spinning particles, and many other exceptional properties of the Kerr, and, more generally, Kerr–NUT–(A)dS [8, 9] spacetimes (for recent review see [10]).

With motivations from string theory and braneworld scenarios, hidden symmetries and spacetimes with hidden symmetries were recently studied in higher dimensions. It turns out, that vacuum (including a cosmological constant) spacetimes admitting KY tensors include physically interesting geometries such as the most general known higher-dimensional Kerr–NUT–(A)dS metrics [11]. As a consequence, the properties of these (spherical) black holes are very similar to the properties of

their four-dimensional counterparts [10, 12, 13]. Another class of spacetimes which possesses hidden symmetries is a class of black branes where the KY symmetry is inherited from the black hole base space. On the other hand, the existence of a non-degenerate CKY 2-form restricts the spacetime to a special algebraic type D [14]. This implies that (non-degenerate) KY tensors do not exist for 5D non-spherical black holes, such as black rings or black saturns. An interesting question is whether hidden symmetries can be found for black holes in the presence of matter fields, such as Kaluza–Klein black holes or black holes of various supergravities. Although (conformal) Killing tensors are known for a wide class of supergravity black holes, see, e.g., [15] and references therein, this is not the case for KY tensors. An example of spacetime admitting a KY symmetry is the Kerr–Newman black hole, which may be viewed as a solution of $D = 4$ supergravity. A proper generalization of this symmetry was recently found for black holes of $D = 5$ minimal gauged supergravity. For black holes of other supergravities KY tensors are (if they exist) yet to be discovered.

In this paper, we try to briefly overview the picture of recent progress in the field of hidden symmetries. As (at least for the vacuum case) there exist several extended reviews on the subject, see, e.g., [10, 12, 13], we mainly concentrate on two things: i) presenting black hole spacetimes for which hidden symmetries are known to exist ii) summarizing certain aspects related to generalized KY tensors in minimal gauged supergravity. Throughout the paper we follow notations of [10].

2. Vacuum spacetimes with hidden symmetries

2.1. Black holes

In vacuum D -dimensional (rotating) black hole spacetimes the central hidden symmetry is associated with a *closed conformal Killing–Yano* (CCKY) 2-form [16, 17], which is a special case of a *conformal Killing–Yano* (CKY) tensor introduced by Kashiwada and Tachibana [18, 19]. The CKY tensor $\mathbf{k}$ of rank p is a p -form which for any vector field $\mathbf{X}$ obeys (the Hodge star invariant) equation

$$\nabla_X \mathbf{k} = \frac{1}{p+1} \mathbf{X} \lrcorner d\mathbf{k} - \frac{1}{D-p+1} \mathbf{X}^\flat \wedge \delta \mathbf{k}. \quad (1)$$

If the first term on the r.h.s. vanishes one has a CCKY tensor. The vanishing of the second term means that one is dealing with a KY tensor. These ‘dual’ special cases (the Hodge star transforms a CCKY tensor into a KY tensor and vice versa) are of particular importance. They give rise to Killing tensors and associated conserved quantities. Moreover, the CCKY tensors form an algebra with respect to the wedge product; a wedge product of two CCKY tensors is again a CCKY tensor [20].

The most general metric element admitting a CCKY 2-form $\mathbf{h}$,

$$\nabla_c h_{ab} = 2g_{c[a} \xi_{b]}, \quad \xi_a = \frac{1}{D-1} \nabla_c h^c_a, \quad (2)$$

was constructed by Houri *et al.* [21, 22]. Depending on the character of eigenvalues of $\mathbf{h}^2$, such a metric consists of three parts: i) a ‘canonical Kerr–NUT-(A)dS

spacetime' [23, 24] which corresponds to functionally independent eigenvalues ii) the (unspecified) Kähler manifolds—corresponding to nonzero constant eigenvalues and iii) an 'arbitrary metric' on a space of zero eigenvalue. Specifically, when all the eigenvalues are functionally independent $\mathbf{h}$ was called a *principal conformal Killing–Yano* (PCKY) tensor [16, 17, 20] and the metric possesses many interesting properties. Namely, since the PCKY tensor is completely non-degenerate, one can extract from it a sufficient number of explicit and hidden symmetries [20] which ensure complete integrability of geodesic motion [25–27], separability of the Hamilton–Jacobi [28], Klein–Gordon [28, 29], Dirac [30–32], and stationary string [33] equations, as well as separability of certain gravitational [34, 35] perturbations. The metric is of the algebraic type D [36]. The problem of parallel transport reduces to a set of first order differential equations [37, 38]. When the vacuum Einstein equations are imposed the metric describes the most general known Kerr–NUT–(A)dS black hole spacetime [11]. Performing a certain limit in which the PCKY tensor becomes completely degenerate one can obtain [39] the most general known Einstein–Kähler and Einstein–Sasaki manifolds [11, 40]. The (Wick-rotated) metric $\mathbf{g}$, together with its PCKY tensor $\mathbf{h} = d\mathbf{b}$, can be written in the following multi Kerr–Schild form [41] ($n = [D/2], \varepsilon = D - 2n$):^a

$$\mathbf{g} = \mathbf{g}_{(\text{A})\text{dS}} + \sum_{\mu=1}^n \frac{2b_{\mu}x_{\mu}^{1-\varepsilon}}{U_{\mu}} \mathbf{l}_{\mu}^2, \quad \mathbf{h} = \sum_{\mu=1}^n x_{\mu} d x_{\mu} \wedge \mathbf{l}_{\mu}, \quad (3)$$

where parameters b_{μ} stand for mass and NUT charges, and the (A)dS metric $\mathbf{g}_{(\text{A})\text{dS}}$ is written as

$$\begin{aligned} \mathbf{g}_{(\text{A})\text{dS}} = & - \sum_{\mu=1}^n \left(\frac{X_{\mu}}{U_{\mu}} \mathbf{l}_{\mu}^2 - \mathbf{l}_{\mu} d x_{\mu} - d x_{\mu} \mathbf{l}_{\mu} \right) + \frac{\varepsilon c}{A^{(n)}} \mathbf{l}_0^2, \quad X_{\mu} = \sum_{k=\varepsilon}^n c_k x_{\mu}^{2k} + \frac{\varepsilon c}{x_{\mu}^2}, \\ \mathbf{l}_{\mu} = & \sum_{k=0}^{n-1} A_{\mu}^{(k)} d\psi_k, \quad \mathbf{l}_0 = \sum_{k=0}^n A^{(k)} d\psi_k, \quad \mathbf{b} = \frac{1}{2} \sum_{k=0}^{n-1} A^{(k+1)} d\psi_k, \\ U_{\mu} = & \prod_{\nu \neq \mu} (x_{\nu}^2 - x_{\mu}^2), \quad A_{\mu}^{(k)} = \sum_{\substack{\nu_1 < \dots < \nu_k \\ \nu_i \neq \mu}} x_{\nu_1}^2 \dots x_{\nu_k}^2, \quad A^{(k)} = \sum_{\nu_1 < \dots < \nu_k} x_{\nu_1}^2 \dots x_{\nu_k}^2. \end{aligned} \quad (4)$$

Parameter c_n is proportional to the cosmological constant, other parameters c_k , c are related to rotations; metric functions $X_{\mu}(x_{\mu})$ are linear in all of them. Forms $\mathbf{l}_{\mu}$ and $\mathbf{l}_0$ are nonzero(zero)-eigenvalue eigenforms of $\mathbf{h}$ and correspond to principal null directions (WANDs) of the metric [12, 14, 24, 36]. The expression (3) nicely demonstrates the connection between the existence of the PCKY tensor and the form of the corresponding Kerr–Schild structure. (For a more general discussion on higher-dimensional Kerr–Schild spacetimes see [42]).

^aThe expressions for $\mathbf{g}$ and $\mathbf{h}$ in so called Darboux basis can be found, e.g., in reviews [10, 12, 13].

2.2. Black branes

Another class of higher-dimensional spacetimes which possess KY symmetries and do not necessary fall into the class described above are black branes with the base that allows a KY tensor. Let g_B be a base space admitting a KY tensor f (CKY is not enough). Then, a p -brane

$$g = g_B + \sum_{a=1}^p dx^a dx^a, \quad (5)$$

possesses the same KY tensor f (now understood as a higher-dimensional one). An obvious example of this is a Kerr-string which possesses a KY 2-form inherited from the Kerr spacetime. Slightly more generally, let g_B be a base space admitting a KY q -form f_B , then $g(x, y) = h^2(x)g_B(y) + \gamma_{ab}(x)dx^a dx^b$ inherits a KY tensor $f = h^{q+1}(x)f_B$. In particular, a product spacetime inherits all the KY tensors of individual constituents. See also metric (6) below.

3. Non-vacuum spacetimes with hidden symmetries

As mentioned in the introduction, there are only few known examples of non-vacuum black hole spacetimes admitting CKY symmetries. In 4D the most important of them is a (7-parametric) family of type D solutions of Einstein–Maxwell equations obtained by Plebanski and Demianski [43]. The metric element describes the Kerr–Newman black hole, the various generalizations of the C-metric, or even the, recently discovered, black hole spacetimes with a conformally coupled scalar field [44, 45]. The metric admits a CKY 2-form, which becomes a KY 2-form when the ‘acceleration’ parameter is removed.^b Explicit expressions for the (conformal) KY tensors of the general metric and its various subclasses can be found in [46]. Another example of spacetime admitting CKY 2-form is a dilatonic C-metric obtained recently by KK-reduction of a single spinning black ring [47].

Different class admitting KY symmetries are spherically symmetric spacetimes. KY tensors there are ‘inherited from the sphere’. Namely, a ‘generalized’ spherically symmetric spacetime

$$g = \gamma_{ab}(x)dx^a dx^b + h^2(x)d\omega_d^2, \quad (6)$$

where $d\omega_d^2$ denotes the metric induced on a unit d -sphere (parameterized by angles $\theta_i \in [0, \pi]$, $i = 1, \dots, d-1$ and $\varphi \in [0, 2\pi]$)

$$d\omega_d^2 = \sum_{i=1}^{d-1} e_i^2 + e_\varphi^2, \quad e_i = \prod_{j<i} \sin \theta_j d\theta_i, \quad e_\varphi = \prod_{j=1}^{d-1} \sin \theta_j d\varphi, \quad (7)$$

^bIn fact, after the acceleration is removed, the metric element attains a form of the four-dimensional canonical Kerr–NUT-(A)dS spacetime described in Sec. 2. Such an element was already obtained by Carter in 1968 [9] as the most general metric admitting separability of the Hamilton–Jacobi and Klein–Gordon equations.

possesses the following ($i = 1, \dots, d-1$) KY tensors of decreasing rank:^c

$$\mathbf{f}_i = h^{d-i+2}(x) \mathbf{f}_i^{(\omega_d)}, \quad \mathbf{f}_i^{(\omega_d)} = \left(\prod_{j < i} \sin \theta_j \right) \mathbf{e}_i \wedge \mathbf{e}_{i+1} \wedge \dots \wedge \mathbf{e}_{d-1} \wedge \mathbf{e}_\varphi. \quad (8)$$

In particular, this applies to spherically symmetric black holes of a-model [49], or to charged black strings (see, e.g., [50] and references therein).

4. Generalized KY symmetries: Black hole of minimal gauged supergravity

When dealing with CKY symmetries, defined by Eq. (1), there is an important limitation which restricts the spacetime to a spacetime of special algebraic type. Namely, a spacetime admitting a (non-degenerate) CKY 2-form must be of the special algebraic type D [14, 51–53]. Such a restriction a priori excludes most of the black holes with matter fields, such as those of various supergravities. On the other hand, some of these black holes are known to admit Killing tensors and allow separability of the Hamilton–Jacobi and Klein–Gordon equations, see, e.g., [15] and references therein. An interesting question is whether it is possible to generalize the concept of KY symmetries for these solutions, while preserving most of its fundamental physical properties and implications. The following example demonstrates that a natural generalization exists for a black hole of minimal gauged supergravity.

It was shown by Wu [54] that the most general known black hole solution of $D = 5$ minimal gauged supergravity [55] (which is a solution of the Einstein–Maxwell theory with a specific Chern–Simons coupling) admits a *generalized PCKY* 2-form $\mathbf{h}$. Such a 2-form obeys a modified Eq. (2), where an additional term, proportional to the electromagnetic flux $\star \mathbf{F}$, is present [54, 56],

$$\nabla_c h_{ab} = 2g_{c[a} \xi_{b]} - \frac{1}{\sqrt{3}} (\star \mathbf{F})_{cd} h^d_{[a} h^d_{b]}. \quad (9)$$

This generalization can be naturally described [56] by the original Eq. (2) if all the derivatives therein are understood as in the presence of a torsion field $\mathbf{T} = \star \mathbf{F} / \sqrt{3}$. Such a torsion is ‘harmonic’, $\delta^T \mathbf{T} = 0 = d^T \mathbf{T}$, due to Maxwell’s equations, and the identification has many appealing features. CKY tensors in the presence of torsion preserve most of the properties of ‘standard’ CKY tensors. In particular, the generalized PCKY tensor generates a Killing tensor and implies the existence of symmetry operators for the Hamilton–Jacobi, Klein–Gordon, and torsion-modified Dirac equations. In the background [55] these equations and their charged versions were separated in [54, 57, 58], stationary string equations for equal angular momenta

^cIt should be stressed that tensors $\mathbf{f}_i^{(\omega_d)}$ are not the only KY forms on a unit sphere. In fact, S^d admits a maximum possible number, $n = (d+1)!/[(d-p)!(p+1)!]$, of independent KY tensors of each rank p , which coincide with the coclosed minimal eigenvalue eigenforms of the Laplace operator [48].

were separated in [59]. The (Wick-rotated) metric $\mathbf{g}$ and the electromagnetic flux $\mathbf{F} = d\mathbf{A}$ are given by [55, 56, 60]

$$\mathbf{g} = \frac{x-y}{4X} d\mathbf{x}^2 - \frac{x-y}{4Y} d\mathbf{y}^2 - \frac{X(dt+yd\phi)^2}{x(x-y)} + \frac{Y(dt+xd\phi)^2}{y(x-y)} - \frac{\mu^2}{xy} l_0^2, \\ \mathbf{A} = \frac{\sqrt{3}\mu}{x-y} [q(dt+yd\phi) - p(dt+xd\phi)], \quad (10)$$

where

$$l_0 = dt + (x+y)d\phi + xy d\psi - \frac{yy(dt+yd\phi)}{x-y} + \frac{px(dt+xd\phi)}{x-y}, \\ X = \mu^2(1+q)^2 + ax + cx^2 + \frac{1}{12}\Lambda x^3, \quad Y = \mu^2(1+p)^2 + by + cy^2 + \frac{1}{12}\Lambda y^3. \quad (11)$$

The generalized PCKY tensor $\mathbf{h}$ in this background is closed and can be generated from a potential, $\mathbf{h} = d\mathbf{b}$, with $\mathbf{b} = -\frac{1}{2}[(x+y)dt + xy d\phi]$. The metric is of type I_i and the eigenvectors of $\mathbf{h}$ are principal null directions. Similar to the vacuum case, all the Killing fields can be ‘derived’ from the knowledge of $\mathbf{h}$

$$\xi_0^a = \xi^a = (\partial_t)^a, \quad \xi_1^a = K^a_b \xi^b + \frac{1}{4\sqrt{3}}(*F)^{abc} K_{bd} h^d_c = (\partial_\phi)^a, \quad \xi_2^a = \frac{\mu}{2}(*h^{\wedge 2})^a = (\partial_\psi)^a. \quad (12)$$

Here, $K_{ab} = (*h)_{acd}(*h)_b{}^{cd} = h_{ac}h_b{}^c - \frac{1}{2}g_{ab}h^2$, is a Killing tensor associated with $\mathbf{h}$. The construction (12) is a ‘natural’ generalization of the construction known from vacuum [20] and underlines the following uniqueness result [61]: the spacetime (10)–(11) is the most general spacetime with ‘harmonic’ torsion field which admits a closed generalized PCKY tensor.

The motion of charged test particles is governed by the charged Hamilton–Jacobi equation

$$\frac{\partial S}{\partial \lambda} + g^{ab}(\partial_a S + eA_a)(\partial_b S + eA_b) = 0. \quad (13)$$

The separability of this equation in the background (10) follows from the separability of the complex Klein–Gordon field [58] and the geometric optics approximation. This results in the following formula for the particle’s momentum $\mathbf{p}$ (obeying $\dot{p}_a = eF_{ab}p^b$):

$$\mathbf{p} = (E - e\sqrt{3}\mu \frac{q-p}{x-y})dt + (\Phi - e\sqrt{3}\mu \frac{qy-px}{x-y})d\phi + \Psi d\psi \\ + \sigma_x \sqrt{\frac{W_x^2}{4xX^2} + \frac{V_x}{4X}} d\mathbf{x} + \sigma_y \sqrt{\frac{W_y^2}{4yY^2} + \frac{V_y}{4Y}} d\mathbf{y}, \quad (14)$$

where $\sigma_x, \sigma_y = \pm$ are independent signs,

$$V_x = \kappa_0 x - \kappa_1 - \frac{\Psi^2}{\mu^2 x}, \quad V_y = \kappa_0 y - \kappa_1 - \frac{\Psi^2}{\mu^2 y}, \quad \kappa_0 = p_a p^a, \\ W_x = Ex^2 - \Phi x + \Psi + q(\Psi - \sqrt{3}\mu ex), \quad W_y = Ey^2 - \Phi y + \Psi + p(\Psi - \sqrt{3}\mu ey), \quad (15)$$

the parameters E, Φ, Ψ are separation constants corresponding to the Killing fields (12), and κ_1 is a separation constant associated with the Killing tensor $\mathbf{K}$. Let us finally mention that metric (10) can be cast in a certain ‘generalized’ Kerr–Schild form [62]. This form is, however, fundamentally different to (3); in addition to standard Kerr–Schild terms (proportional to the squares of principal null directions) there appear additional cross terms of principal null directions with extra space-like vector.

References

- [1] B. Carter, *Phys. Rev.* **174**, 1559 (1968).
- [2] R. P. Kerr, *Phys. Rev. Lett.* **11**, 237 (1963).
- [3] P. Stackel, *C. R. Acad. Sci. Paris Ser. IV* **121**, p. 489 (1895).
- [4] M. Walker and R. Penrose, *Commun. Math. Phys.* **18**, 265 (1970).
- [5] K. Yano, *Ann. Math.* **55**, 328 (1952).
- [6] R. Penrose, *Annals N. Y. Acad. Sci.* **224**, 125 (1973).
- [7] R. Floyd, The dynamics of kerr fields, PhD thesis, London University, (London, United Kingdom, 1973).
- [8] B. Carter, *Phys. Lett.* **26A**, 399 (1968).
- [9] B. Carter, *Commun. Math. Phys.* **10**, 280 (1968).
- [10] V. P. Frolov and D. Kubizňák, *Class. Quantum Grav.* **25**, p. 154005 (2008), arXiv:0802.0322.
- [11] W. Chen, H. Lü and C. N. Pope, *Class. Quantum Grav.* **23**, 5323 (2006), hep-th/0604125.
- [12] D. Kubizňák, Hidden symmetries of higher-dimensional rotating black holes, PhD thesis, University of Alberta, (Edmonton, Alberta, Canada, 2008), arXiv:0809.2452.
- [13] V. P. Frolov (2009), arXiv:0901.1472.
- [14] L. Mason and A. Taghavi-Chabert (2008), arXiv:0805.3756.
- [15] D. D. K. Chow (2008), arXiv:0811.1264.
- [16] V. P. Frolov and D. Kubizňák, *Phys. Rev. Lett.* **98**, p. 011101 (2007), gr-qc/0605058.
- [17] D. Kubizňák and V. P. Frolov, *Class. Quantum Grav.* **24**, F1 (2007), gr-qc/0610144.
- [18] T. Kashiwada, *Nat. Sci. Rep. Ochanomizu Univ.* **19**, 67 (1968).
- [19] S. Tachibana, *Tôhoku Math. J.* **21**, p. 56 (1969).
- [20] P. Krtouš, D. Kubizňák, D. N. Page and V. P. Frolov, *J. High Energy Phys.* **02**, p. 004 (2007), hep-th/0612029.
- [21] T. Houri, T. Oota and Y. Yasui, *Phys. Lett.* **B666**, 391 (2008), arXiv:0805.0838.
- [22] T. Houri, T. Oota and Y. Yasui, *Class. Quantum Grav.* **26**, p. 045015 (2009), arXiv:0805.3877.
- [23] T. Houri, T. Oota and Y. Yasui, *Phys. Lett.* **B656**, 214 (2007), arXiv:0708.1368.
- [24] P. Krtouš, V. P. Frolov and D. Kubizňák, *Phys. Rev.* **D78**, p. 064022 (2008), arXiv:0804.4705.
- [25] D. N. Page, D. Kubizňák, M. Vasudevan and P. Krtouš, *Phys. Rev. Lett.* **98**, p. 061102 (2007), hep-th/0611083.
- [26] P. Krtouš, D. Kubizňák, D. N. Page and M. Vasudevan, *Phys. Rev.* **D76**, p. 084034 (2007), arXiv:0707.0001.
- [27] T. Houri, T. Oota and Y. Yasui, *J. Phys.* **A41**, p. 025204 (2008), arXiv:0707.4039.
- [28] V. P. Frolov, P. Krtouš and D. Kubizňák, *J. High Energy Phys.* **02**, p. 005 (2007), hep-th/0611245.
- [29] A. Sergeyev and P. Krtouš, *Phys. Rev.* **D77**, p. 044033 (2008), arXiv:0711.4623.
- [30] T. Oota and Y. Yasui, *Phys. Lett.* **B659**, 688 (2008), arXiv:0711.0078.

- [31] S.-Q. Wu, *Phys. Rev. D* **78**, p. 064052 (2008), arXiv:0807.2114.
- [32] S.-Q. Wu, *Class. Quantum Grav.* **26**, p. 055001 (2009), arXiv:0808.3435.
- [33] D. Kubizňák and V. P. Frolov, *J. High Energy Phys.* **02**, p. 007 (2008), arXiv:0711.2300.
- [34] H. K. Kunduri, J. Lucietti and H. S. Reall, *Phys. Rev.* **D74**, p. 084021 (2006), hep-th/0606076.
- [35] T. Oota and Y. Yasui (2008), arXiv:0812.1623.
- [36] N. Hamamoto, T. Houri, T. Oota and Y. Yasui, *J. Phys.* **A40**, F177 (2007), hep-th/0611285.
- [37] P. Connell, V. P. Frolov and D. Kubizňák, *Phys. Rev.* **D78**, p. 024042 (2008), arXiv:0803.3259.
- [38] D. Kubiznak, V. Frolov, P. Krtous and P. Connell, *Phys. Rev.* **D79**, p. 024018 (2009), arXiv:0811.0012.
- [39] D. Kubiznak, *Phys. Lett.* **B675**, p. 110 (2009), arXiv:0902.1999.
- [40] V. Apostolov, D. M. J. Calderbank, P. Gauduchon and C. W. Tonneson-Friedman, *J. Diff. Geom.* **68**, 277 (2004), math/0401320.
- [41] W. Chen and H. Lü, *Phys. Lett.* **B658**, 158 (2008), arXiv:0705.4471.
- [42] M. Ortaggio, V. Pravda and A. Pravdova, *Class. Quant. Grav.* **26**, p. 025008 (2009), arXiv:0808.2165.
- [43] J. F. Plebański and M. Demiański, *Ann. Phys. (N.Y.)* **98**, 98 (1976).
- [44] A. Anabalon and H. Maeda (2009), arXiv:0907.0219.
- [45] C. Charmousis, T. Kolyvaris and E. Papantonopoulos, *Class. Quant. Grav.* **26**, p. 175012 (2009), arXiv:0906.5568.
- [46] D. Kubizňák and P. Krtouš, *Phys. Rev.* **D76**, p. 084036 (2007), arXiv:0707.0409.
- [47] M. Durkee, *Class. Quantum Grav.* **26**, p. 085016 (2009), arXiv:0812.0235.
- [48] U. Semmelmann (2002), math/0206117.
- [49] T. Ortin, *Gravity and Strings*, p. 349 (Cambridge Univ. Press., Cambridge, England, 2004).
- [50] V. P. Frolov and A. A. Shoom, *Phys. Rev.* **D79**, p. 104002 (2009), arXiv:0903.2893.
- [51] C. D. Collinson, *Tensor* **28**, p. 173 (1974).
- [52] H. Stephani, *Gen. Rel. Grav.* **9**, 789 (1978).
- [53] E. N. Glass and J. Kress, *J. Math. Phys.* **40**, 309 (1999), gr-qc/9809074.
- [54] S.-Q. Wu (2009), arXiv:0902.2823.
- [55] Z. W. Chong, M. Cvetič, H. Lu and C. N. Pope, *Phys. Rev. Lett.* **95**, p. 161301 (2005), hep-th/0506029.
- [56] D. Kubiznak, H. Kunduri and Y. Yasui, *Phys. Lett.* **B678**, 240 (2009), arXiv:0905.0722.
- [57] P. Davis, H. K. Kunduri and J. Lucietti, *Phys. Lett.* **B628**, 275 (2005), hep-th/0508169.
- [58] S.-Q. Wu (2009), arXiv:0906.2049.
- [59] H. Ahmedov and A. N. Aliev (2008), arXiv:0805.1594.
- [60] H. Lü, J. Mei and C. N. Pope (2008), arXiv:0806.2204.
- [61] H. Ahmedov and A. N. Aliev, *Phys. Lett.* **B679**, p. 396 (2009), arXiv:0907.1804.
- [62] A. N. Aliev and D. K. Ciftci, *Phys. Rev.* **D79**, p. 044004 (2009), arXiv:0811.3948.

MAXIMAL SUPERSYMMETRY AND NON-ABELIAN GAUGE GROUPS

HENNING SAMTLEBEN

*Université de Lyon, Laboratoire de Physique,
Ecole Normale Supérieure de Lyon,
46, allée d'Italie, F-69364 Lyon Cedex 07, France
E-mail: henning.samtleben@ens-lyon.fr*

I review recent progress in the construction and classification of maximally supersymmetric theories with non-abelian gauge groups. The algebraic framework is based on the underlying exceptional symmetry groups. This has applications for supergravity theories describing flux compactifications as well as for the recently constructed three-dimensional superconformal theories describing the dynamics of multiple M2-branes.

Keywords: Supersymmetry; Supergravity; Gauge symmetry

1. Introduction

Supersymmetric theories in four dimensions are invariant under symmetries that close into the superextension of the Poincaré algebra

$$\{Q^i_\alpha, \bar{Q}_{j\beta}\} = \delta^i_j \gamma^\mu_{\alpha\beta} P_\mu, \tag{1}$$

where the indices μ and α, β refer to space-time vectors and spinors, respectively. The indices $i, j = 1, \dots, N$ label the number of supercharges. As a result, the field content of these theories falls into representations of the superalgebra (1). Algebraic consistency restricts the maximal number of supercharges to $N = 4$ and $N = 8$ for global and local supersymmetry, respectively. The resulting supermultiplets are collected in Table 1, the states being organized by the helicity group $SO(2)$.

Table 1. Maximal supermultiplets.

helicity	-2	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$+\frac{1}{2}$	+1	$+\frac{3}{2}$	+2
$N = 4$:			1	4	6	4	1		
$N = 8$:	1	8	28	56	70	56	28	8	1

The $N = 4$ supermultiplet combines 6 scalar fields with their fermionic superpartners and the two helicity states of a massless vector field. The most general

theory with global $N = 4$ supersymmetry combines an arbitrary number n of these multiplets, whose gauge fields realize a gauge group G of dimension n . These are the super Yang-Mills theories [1]. Their classification thus reduces to the classification of Lie algebras $\mathfrak{g}$ of a given dimension, i.e. to the solution of the Jacobi identities

$$f_{[ab}{}^c f_{d]c}{}^e = 0 . \quad (2)$$

The situation is quite different for maximal local supersymmetry, i.e. for theories realizing the $N = 8$ supermultiplet of Table 1. As this multiplet includes the spin-2 states of the space-time metric, only a single copy can be described by an interacting theory. It then becomes a non-trivial question which gauge groups can be realized by the 28 gauge fields while preserving maximal local supersymmetry. Indeed, after the original construction of $N = 8$ supergravity [2] with abelian gauge group $U(1)^{28}$ and the first non-abelian version [3] with gauge group $SO(8)$, a variety of other viable non-compact and non-semisimple gauge groups were discovered [4–6]. A convenient framework to achieve a general classification uses the so-called embedding tensor [7–9]. In the following I will review this formalism which — generalizing (2) — reduces the classification of maximal supergravities to a set of algebraic equations for the embedding tensor. I sketch the emerging structure of non-abelian tensor fields in these theories and finish with an example of a non-trivial application of the formalism also in theories with global supersymmetry.

2. Gauged supergravity

A distinguished property of the abelian $N = 8$ supergravity is the underlying global symmetry group $G_0 = E_{7(7)}$ [2]. Its 70 scalar fields parametrize the coset space manifold $E_{7(7)}/SU(8)$ while its 28 vector fields combine with their magnetic duals into the fundamental 56-dimensional representation of $E_{7(7)}$, which we denote by A_μ^M , ($M = 1, \dots, 56$). On the physical fields of Table 1, this symmetry thus acts in a non-linear and non-local way; infinitesimally, we parametrize this action as

$$\delta_\Lambda \Phi = \Lambda^\alpha t_\alpha \cdot \Phi , \quad (3)$$

for an arbitrary physical field Φ , parameter Λ and with t_α denoting the generators of $\mathfrak{e}_{7(7)} = \text{Lie } E_{7(7)}$ (i.e. $\alpha = 1, \dots, 133$ parametrizing the adjoint representation), closing into

$$[t_\alpha, t_\beta] = f_{\alpha\beta}{}^\gamma t_\gamma . \quad (4)$$

In this version of the theory, none of the matter fields is charged under the $U(1)^{28}$ gauge group. Non-abelian deformations of the theory are obtained by selecting a subset of generators t_α and promoting their action (3) into a local symmetry. The choice of generators is conveniently parametrized as

$$X_M \equiv \Theta_M^\alpha t_\alpha , \quad (5)$$

by a constant *embedding tensor* Θ_M^α , which encodes the embedding of the gauge group into the global $E_{7(7)}$ of the ungauged theory. The action of the generators X_M

can be promoted to a local symmetry by introducing standard covariant derivatives^a

$$D_\mu \equiv \partial_\mu + A_\mu^M X_M . \quad (6)$$

Not every choice of generators (5) leads to a consistent theory. In particular, the generators X_M must close into a subalgebra. Moreover, it is only for very particular subalgebras that the deformation is compatible with maximal supersymmetry. Fortunately, all consistency conditions translate into a simple set of algebraic equations for the embedding tensor Θ_M^α . Closure of the gauge algebra requires

$$\Theta_P^\beta \Theta_N^\alpha (t_\beta)_M{}^N + \Theta_P^\beta \Theta_M^\gamma f_{\beta\gamma}{}^\alpha = 0 , \quad (7)$$

which can be interpreted as generalization of the Jacobi identities (2). Supersymmetry imposes a linear constraint on the embedding tensor

$$\Theta_M^\alpha = (\mathbb{P}_{912} \Theta)_M{}^\alpha , \quad (8)$$

where the projector refers to the tensor product decomposition

$$\mathbf{56} \otimes \mathbf{133} = \mathbf{56} \oplus \mathbf{912} \oplus \mathbf{6480} , \quad (9)$$

of the fundamental and the adjoint representation of $E_{7(7)}$, corresponding to the two indices of the embedding tensor. This leaves 912 free parameters in Θ_M^α , subject to the 8778 quadratic equations (7). It can be shown [9] that every solution to the system of algebraic equations (7), (8) defines a consistent theory with maximal local supersymmetry and gauge group spanned by the generators (5). In particular, the scalar potential that appears in the gauged theory, takes the form

$$V = \Theta_M^\alpha \Theta_N^\beta V^{MN}{}_{\alpha\beta}(\phi) , \\ V^{MN}{}_{\alpha\beta}(\phi) \equiv ((t_\alpha)_P{}^R (t_\beta)_Q{}^S \mathcal{M}^{PQ} \mathcal{M}_{RS} + 7 (t_\alpha)_P{}^Q (t_\beta)_Q{}^P) \mathcal{M}^{MN} , \quad (10)$$

bilinear in the embedding tensor Θ_M^α . The scalar dependence arises through the symmetric positive definite matrices $\mathcal{M} = \mathcal{V} \mathcal{V}^T$, $\mathcal{V}$ denoting the $E_{7(7)}/SU(8)$ coset representative, see [9] for details.

The universal form of the Lagrangian of the gauged theory and in particular of its scalar potential (10) is of particular importance for the study of the low energy effective actions associated to particular flux compactifications of higher dimensional theories. E.g. in flux compactifications of IIA/IIB supergravity on (twisted) tori, the flux parameters show up as deformation parameters among the components of the embedding tensor and can be uniquely identified by purely group-theoretical means, see [11] and references therein. The expression (10) then allows to read off the full scalar potential of the effective theory upon specifying the particular choice of Θ_M^α .

The framework sketched here can equally well be applied to the construction of gauged supergravity theories in other space-time dimensions and with a lower number of supersymmetries, in which case the constraints (7), (8) are replaced by equations covariant under the relevant global symmetry group.

^aTo be as general as possible, we allow for connections combining electric and magnetic gauge fields. It turns out that for consistent choice of Θ_M^α (see below) this gives rise to sensible theories which allow for an off-shell formulation even in the presence of magnetic charges [10].

3. Non-abelian tensor fields

Apart from the minimal couplings induced by the covariant derivatives (6), the field strengths of the vector fields acquire the standard non-abelian modification

$$\mathcal{F}_{\mu\nu}^M = \partial_\mu A_\nu^M - \partial_\nu A_\mu^M + X_{[NP]}^M A_\mu^N A_\nu^P, \quad (11)$$

with $X_{MN}{}^P \equiv \Theta_M{}^\alpha t_{\alpha N}{}^P$. While in standard Yang-Mills theory this defines covariant objects, here the object (11) in general does not transform covariantly under local gauge transformations. Again, the reason is, that vector fields a priori transform in a given fundamental representation of the global symmetry group; as a consequence, the “structure constants” $X_{MN}{}^P$ do not satisfy the standard Jacobi identities. The natural covariant object turns out to be the following combination [9]

$$\mathcal{H}_{\mu\nu}^M = \mathcal{F}_{\mu\nu}^M + Z^{M\alpha} B_{\mu\nu\alpha}, \quad (12)$$

with $Z^{M\alpha} \equiv -\frac{1}{2}\Omega^{MN}\Theta_N{}^\alpha$, upon the introduction of two-form tensor fields $B_{\mu\nu\alpha}$, transforming in the adjoint representation of $E_{7(7)}$. The new field strength $\mathcal{H}_{\mu\nu}^M$ transforms covariantly under the combined set of gauge transformations

$$\begin{aligned} \delta A_\mu^M &= D_\mu \Lambda^M - Z^{M\alpha} \Xi_{\mu\alpha}, \\ \delta B_{\mu\nu\alpha} &= 2 D_{[\mu} \Xi_{\nu]\alpha} + 2(t_\alpha)_M{}^K \Omega_{KN} \left(\Lambda^M \mathcal{H}_{\mu\nu}^N - A_{[\mu}^M \delta A_{\nu]}^N \right), \end{aligned} \quad (13)$$

which exhibit Stückelberg-type couplings between vector fields and antisymmetric two-forms, as is familiar from massive deformations of supergravities, e.g. [12].

Similarly, the construction of a covariant field strength for the two-forms requires the introduction of three-form tensor fields, etc. The final result is a full hierarchy of non-abelian tensor fields which (schematically) extends (13) to

$$\begin{aligned} \delta B_{\mu\nu} &= 2D_{[\mu} \Xi_{\nu]} + \dots - Y \Phi_{\mu\nu}, \\ \delta C_{\mu\nu\rho} &= 3D_{[\mu} \Phi_{\nu\rho]} + \dots - W \Sigma_{\mu\nu\rho}, \\ \delta D_{\mu\nu\rho\sigma} &= 4D_{[\mu} \Sigma_{\nu\rho\sigma]} + \dots, \end{aligned} \quad (14)$$

where we have omitted all $E_{7(7)}$ -indices, see [13, 14] for details. The intertwining tensors Z , Y and W are functions of the embedding tensor and determine the precise $E_{7(7)}$ representation content in which the higher rank p -forms must appear. For the $N = 8$ theory, the required forms B , C and D transform in the **133**, the **912**, and the **133 + 8645**, respectively. This gives agreement with the predictions obtained from analyzing the branchings of the infinite-dimensional representations of the underlying very extended Kac-Moody algebra [15, 16]. Moreover, this representation assignment indicates that the higher rank forms do not represent new propagating degrees of freedom but magnetic duals to the physical field content: the two-forms are dual to the 133 $E_{7(7)}$ Noether currents, while the 912 three-forms are related to the deformation parameters combined in the embedding tensor as

$$D_{[\mu} C_{\nu\rho\sigma]}{}^M{}_\alpha + \dots = \epsilon_{\mu\nu\rho\sigma} V^{MN}{}_{\alpha\beta}(\phi) \Theta_N{}^\beta, \quad (15)$$

with the scalar dependent matrix $V^{MN}{}_{\alpha\beta}(\phi)$ from (10). The 8778 four-forms D in turn can be interpreted as Lagrange multipliers for the quadratic constraints (7) [13].

4. Global supersymmetry in three dimensions

In certain cases, the embedding tensor formalism also provides a valuable framework for theories with global supersymmetry [17, 18]. In three dimensions, maximally supersymmetric Yang-Mills theory is obtained by dimensional reduction from ten dimensions [1]. However, another $N = 8$ theory can be constructed by gauging the free theory of n scalar multiplets $\{\Phi^{Ik}, \Psi^{\dot{A}k}\}$, labeled by k , with I and $\dot{A}$ labeling $SO(8)$ vectors and spinors, respectively. The $SO(n)$ Noether currents

$$J_\mu^{kl} = \Phi^{I[k} \partial_\mu \Phi^{l]I} + \bar{\Psi}^{\dot{A}[k} \gamma_\mu \Psi^{\dot{A}l]} , \quad (16)$$

allow for the introduction of dual vector fields which can be used to gauge (a subgroup of) the global $SO(n)$ group. In three dimensions, the embedding tensor thus is of the form $\Theta_M^\alpha \rightarrow \Theta_{kl,mn}$. The constraints (7), (8) in this case take the form

$$\begin{aligned} \Theta_{kl,mn} &= \Theta_{[kl,mn]} , \\ 0 &= \Theta_{kl,pr} \Theta_{mn,rq} - \Theta_{mn,pr} \Theta_{kl,rq} - \Theta_{kl,mr} \Theta_{nr,pq} + \Theta_{kl,nr} \Theta_{mr,pq} , \end{aligned} \quad (17)$$

which reproduces the equations underlying the three-algebra structure of Bagger and Lambert [19], to which $SO(4)$ is the only solution with $\Theta_{kl,mn} = \epsilon_{klmn}$. For $N < 8$, in contrast the corresponding equations admit a variety of possible solutions.

Acknowledgments

I am most grateful to E. Bergshoeff, B. de Wit, O. Hohm, A. Le Diffon, H. Nicolai, D. Roest, E. Sezgin and M. Trigiante for many discussions and collaboration on the topics presented. This work is supported by the Agence Nationale de la Recherche.

References

- [1] L. Brink, J. H. Schwarz and J. Scherk, *Nucl. Phys.* **B121**, 77 (1977).
- [2] E. Cremmer and B. Julia, *Nucl. Phys.* **B159**, 141 (1979).
- [3] B. de Wit and H. Nicolai, *Nucl. Phys.* **B208**, 323 (1982).
- [4] C. M. Hull, *Phys. Lett.* **B142**, 39 (1984).
- [5] L. Andrianopoli, R. D'Auria, S. Ferrara and M. A. Lledo, *JHEP* **07**, 010 (2002).
- [6] C. M. Hull, *Class. Quant. Grav.* **20**, 5407 (2003).
- [7] H. Nicolai and H. Samtleben, *Phys. Rev. Lett.* **86**, 1686 (2001).
- [8] B. de Wit, H. Samtleben and M. Trigiante, *Nucl. Phys.* **B655**, 93 (2003).
- [9] B. de Wit, H. Samtleben and M. Trigiante, *JHEP* **06**, 049 (2007).
- [10] B. de Wit, H. Samtleben and M. Trigiante, *JHEP* **09**, 016 (2005).
- [11] H. Samtleben, *Class. Quant. Grav.* **25**, 214002 (2008).
- [12] L. J. Romans, *Phys. Lett.* **B169**, 374 (1986).
- [13] B. de Wit, H. Nicolai and H. Samtleben, *JHEP* **02**, 044 (2008).
- [14] E. Bergshoeff, J. Hartong, O. Hohm, M. Hübscher and T. Ortin, *JHEP* **04**, 123 (2009).
- [15] F. Riccioni and P. West, *JHEP* **07**, 063 (2007).
- [16] E. Bergshoeff, I. De Baetselier and T. A. Nutma, *JHEP* **09**, 047 (2007).
- [17] E. Bergshoeff, M. de Roo and O. Hohm, *Class. Quant. Grav.* **25**, 142001 (2008).
- [18] E. Bergshoeff, O. Hohm, D. Roest, H. Samtleben and E. Sezgin, *JHEP* **09**, 101 (2008).
- [19] J. Bagger and N. Lambert, *Phys. Rev.* **D77**, 065008 (2008).

OTHER TALKS AND POSTERS

INVITED TALK

ADS/CFT AND GENERALIZED COMPLEX GEOMETRY

JAMES SPARKS

Mathematical Institute, University of Oxford

E-mail: sparks@maths.ox.ac.uk

We study the most general supersymmetric AdS_5 solutions of type IIB supergravity that are dual to $\mathcal{N} = 1$ superconformal field theories (SCFTs) in $d = 4$. Such solutions have associated six-dimensional geometries that generalize Calabi-Yau cone geometry. We identify generalized vector fields dual to the dilatation and R -symmetry of the dual SCFT and show that they are generalized holomorphic on the cone. We carry out a generalized reduction of the cone to a transverse four-dimensional space and show that this is also generalized Hermitian. When the five-form flux is non-vanishing, the cone is symplectic and we relate this to the generalized geometry. The symplectic structure can be used to obtain Duistermaat-Heckman type integrals for the central charge of the dual SCFT and the conformal dimensions of operators dual to BPS wrapped D3-branes. We illustrate these results using the Pilch-Warner solution.

CONTRIBUTED TALKS

SUPERCONFORMAL CHERN-SIMONS THEORIES AND ADS(4)/CFT(3) CORRESPONDENCE

MARCUS BENNA, IGOR KLEBANOV, THOMAS KLOSE, MIKAEL SMEDBÄCK*

**Joseph Henry Laboratories, Princeton University*

E-mail: smedback@princeton.edu

Recently, a world-volume theory for multiple M2-branes has been proposed by Bagger and Lambert. We introduce this model, and the extension of it given last year by Aharony, Bergman, Jafferis and Maldacena (ABJM). We then investigate the amount of supersymmetry of the ABJM model. We also discuss generalizations of the ABJM

model to field theories which are dual to other M-theory backgrounds, using orbifolding techniques and mass deformations.

DESCRIPTION OF D-BRANES INVARIANT UNDER THE POISSON-LIE T-PLURALITY

LADISLAV HLAVATÝ, LIBOR ŠNOBL

Czech Technical University, Prague

E-mail: Libor.Snobl@fjfi.cvut.cz

We write the boundary conditions for open strings with charged endpoints in the language of gluing matrices. We identify constraints imposed on the gluing matrices that are essential in this setup and investigate the question of their invariance under the Poisson-Lie T-plurality transformations. We show that the chosen set of constraints is equivalent to the statement that the lifts of D-branes into the Drinfel'd double are right cosets with respect to a maximally isotropic subgroup and therefore it is invariant under the Poisson-Lie T-plurality transformations. Published in JHEP07(2008)122.

SUPERCONFORMAL INDICES FOR N=1 THEORIES WITH MULTIPLE DUALS

VYACHESLAV P. SPIRIDONOV, GRIGORY S. VARTANOV

Joint Institute for Nuclear Research, Dubna

E-mail: vartanov@theor.jinr.ru

Recently some of elliptic hypergeometric integrals were interpreted as topological indices in N=1 supersymmetric field theories. Coincidence of superconformal indices in Seiberg dual theories corresponds to nontrivial identities for these integrals. Analysing such relations, from known dual field theories we conjecture many new relations for integrals, and from a number of known integral identities, we come to many new Seiberg dualities. The talk is based on the paper arXiv:0811.1909 [hep-th].

POSTERS

CHANDRASEKHAR MUKKU (International Institute of Information Technology, Gachibowli, Hyderabad): **TRANSMUTING TORSION WITH GAUGE INVARIANCE**

ANATOLIY PRYKARPATSKY (The AGH University of Science and Technology, Krakow), with NIKOLAI N. BOGOLUBOV and UFUK TANERI: **THE RELATIVISTIC ELECTRODYNAMICS LEAST ACTION PRINCIPLES REVISITED: NEW CHARGED POINT PARTICLE AND HADRONIC STRING MODELS ANALYSIS**

HIROSHI UMETSU (Kushiro National College of Technology), with Y. HONMA, S. ISO, Y. SUMITOMO, and S. ZHANG: **GENERALIZED CONFORMAL SYMMETRY AND RECOVERY OF SO(8) IN MULTIPLE M2 AND D2 BRANES**

KUNIHITO UZAWA (Kinki University), with KEI-ICHI MAEDA, NOBUYOSHI OHTA: **DYNAMICS OF INTERSECTING BRANE SYSTEMS**

Quantum Field Theory

(including field theoretic methods
in condensed matter)

Session organizers:

Krzysztof Gawedzki
Manfred Salmhofer

THREE ÉTUDES IN QFT

DIRK KREIMER

Institut des Hautes Études Scientifiques

91440 Bures sur Yvette, France

E-mail: kreimer@ihes.fr

http://math.bu.edu/people/dkreimer

We review QFT from the algebraic structure of its perturbative expansion and point out three new results.

Keywords: Hopf algebras, perturbative QFT

Acknowledgments

It is a pleasure to thank Spencer Bloch for questions which ultimately led to Theorem 3 below. This work was partially supported by grant NSF-DMS/0603781 at Boston University.

1. Algebraic structure in local QFT

In recent years we have collected a considerable amount of algebraic structure underlying the computational practice of quantum field theory, clarifying the mathematical foundations of these computations and coming to new insights in QFT, see [1, 4–7] and references there for more detail and notation. In this section, we give a succinct summary.

1.1. Hopf and Lie algebras of graphs

There is a Hopf and Lie algebras coming with 1PI Feynman graphs:

$$\Delta_{2n}(\Gamma) = \Gamma \otimes \mathbb{I} + \mathbb{I} \otimes \Gamma + \sum_{[\gamma]_{2n}} \gamma \otimes \Gamma/\gamma. \quad (1)$$

The sum is over all proper subsets $[\gamma]_{2n} \subset \Gamma$ which constitute disjoint unions $\gamma = \prod \gamma_i$ of 1PI graphs such that each γ_i fulfills $\omega_{2n}(\gamma_i) \geq 0$. Here, $\omega_{2n}(\Gamma) = 2n|\Gamma| - \sum_{\text{weights } w} w(\Gamma)$, and we sum over edge and vertex weights w .

These Hopf algebras include the Hopf algebra of renormalization for a theory renormalizable at $2n = D$ dimensions which then gives the forest formula of renormalization theory. With an antipode $S(\Gamma) = -\Gamma - \sum S(\gamma)\Gamma/\gamma$ we obtain renormalized Feynman rules Φ^R and counterterms S_R as

$$\Phi^R = m(S_R \otimes \Phi)\Delta, \quad S_R = -R(S_R \otimes \Phi P)\Delta, \quad (2)$$

from unrenormalized Feynman rules Φ and a projection into the augmentation ideal $P : H \rightarrow \text{Aug}$.

For all these Hopf algebras there is an underlying pre-Lie algebra structure:

$$[Z_{\Gamma_1}, Z_{\Gamma_2}] = Z_{\Gamma_1} \otimes Z_{\Gamma_2} - Z_{\Gamma_2} \otimes Z_{\Gamma_1} \quad (3)$$

with

$$[Z_{\Gamma_1}, Z_{\Gamma_2}] = Z_{\Gamma_2 \star \Gamma_1 - \Gamma_1 \star \Gamma_2}. \quad (4)$$

Here, $\Gamma_i \star \Gamma_j$ sums over all ways of gluing Γ_j into Γ_i , which can be written as

$$\Gamma_i \star \Gamma_j = \sum_{\Gamma} n(\Gamma_i, \Gamma_j, \Gamma) \Gamma. \quad (5)$$

For any $\Gamma \in H$, we have a pairing

$$\langle [Z_{\Gamma_1}, Z_{\Gamma_2}], \Gamma \rangle = \langle Z_{\Gamma_1} \otimes Z_{\Gamma_2} - Z_{\Gamma_2} \otimes Z_{\Gamma_1}, \Delta(\Gamma) \rangle, \quad (6)$$

for consistency. With such section coefficients $n(\Gamma_i, \Gamma_j, \Gamma)$ we have

$$\Delta(\Gamma) = \sum_{h,g} n(h, g, \Gamma) g \otimes h. \quad (7)$$

The sum is over all graphs h including the empty graph and over all monomials in graphs g .

Here is an example of the coproduct $\Delta = \Delta_4$ on the sum of the two-loop vertex graphs $c_2^{\bar{\psi} A \psi}$:

$$\Delta \left(\begin{array}{c} \text{diagram 1} + \text{diagram 2} + \text{diagram 3} + \text{diagram 4} + \text{diagram 5} + \text{diagram 6} + \text{diagram 7} \end{array} \right) = 3 \begin{array}{c} \text{diagram 8} \otimes \text{diagram 9} \\ + 2 \text{diagram 10} \otimes \text{diagram 11} + \text{diagram 12} \otimes \text{diagram 13} \end{array}$$

which reveals a sub-Hopf algebra structure.

1.2. Sub-Hopf algebras

Along with such Hopf algebras comes the corresponding Hochschild cohomology and the sub-Hopf algebras H_{grad} generated by the grading and the quantum equations of motion. If we let c_k^r be the sum of all graphs contributing to an amplitude r at k loops

$$c_k^r = \sum_{\text{res}(\Gamma)=r, |\Gamma|=k} \frac{1}{\text{Aut}(\Gamma)} \Gamma, \quad (8)$$

then Δ gives to these generators a sub-Hopf algebra structure: Let $\mathcal{R}$ be the set of amplitudes which need renormalization, and let H_{grad} be the Hopf algebra spanned by generators $\{\mathbb{I}, c_k^r\}$, $k \in \mathbb{N}$, $r \in \mathcal{R}$. The c_k^r form a sub-Hopf algebra:

$$\Delta(c_k^r) = \sum_{j=0}^k P_{j,k}^s \otimes c_{k-j}^r, \quad (9)$$

with $P_{j,k}^s$ a polynomial in generators c_m^t of degree j . These sub-Hopf then allow for multiplicative renormalization at a Lagrangian level.

1.3. *Symmetry*

Co-ideals in the Hopf algebra of perturbation theory correspond to symmetries in the Lagrangian: Internal symmetries are in accordance with the Hochschild cohomology structure of these Hopf algebras. For example the Ward identity in quantum electrodynamics is equivalent to the statement that renormalized Feynman rules can be defined on the quotient H_{grad}/I , where I is the ideal and co-ideal given by $c_k^{\psi A} + c_k^{\bar{\psi} \psi} = 0$. See [6] for further references, in particular for Walter van Suijlekom's work on co-ideals as a reflection of internal symmetry.

1.4. *The renormalization group*

The coradical filtration and Dynkin operators govern the renormalization group and leading log expansion: Indeed, in a leading-log expansion, terms $\sim \ln^j s$ (for s a suitable kinematical parameter) are obtained by the evaluation of

$$\sigma_k = \frac{1}{k!} \Phi^R m^{k-1} [S \otimes Y]^k \Delta^{k-1}, \quad (10)$$

a fact heavily used in the reduction of Dyson–Schwinger equations to ODEs, see [7].

1.5. *Full amplitudes*

There is a semi-direct product structure between the cocommutative Hopf algebra H_{ab} of superficially convergent amplitudes and the Hopf algebra H of amplitudes in $\mathcal{R}$ needing renormalization, which is very handy in the organization of those Dyson–Schwinger equations:

$$H_{\text{full}} = H_{\text{ab}} \times H. \quad (11)$$

1.6. *The core algebra*

Finally the core Hopf algebra based on Δ_∞ has co-ideals [6] leading to recursions à la Britto-Cachazo-Feng-Witten, showing that loops and legs speak to each other in many ways. At the same time it allows an identification of renormalization as a limiting mixed Hodge structure [2].

2. Three new results

In this section we outline three new results with details for each of them to be given elsewhere.

2.1. Generalized Witt algebras

A graded commutative Hopf algebra H can be regarded as the dual of the universal enveloping algebra $U(L)$ of a Lie algebra L . What is the dual Lie algebra L_{grad} such that $U(L_{\text{grad}}) = H_{\text{grad}}$? Again, we need

$$\langle z_m^r \otimes z_n^s - z_n^s \otimes z_m^r, \Delta c_j^t \rangle = \langle [z_n^s, z_m^r], \Delta c_j^t \rangle, \quad (12)$$

$\forall j > 0, t \in \mathcal{R}$.

This certainly implies $t \in \{r, s\}$. Looking first at $r = s$, we get immediately

$$[z_k^s, z_l^s] \sim (l - k)z_{k+l}^s. \quad (13)$$

For the general case we find

$$[z_k^s, z_l^t] = -Q(s)kz_{k+l}^s + Q(t)lz_{k+l}^t. \quad (14)$$

Here, $Q(s)$ are integers determined by the invariant charges of the theory, for example in QED one finds $Q(\bar{\psi}A\psi) = Q(\bar{\psi}\psi) = 2$, $Q(\frac{1}{4}F^2) = 1$. The check of the Jacobi identities is then a straightforward computation.

Theorem 1. *With the above, $\mathcal{L}$ becomes a Lie algebra which is the Milnor-Moore dual to the Hopf algebra above.*

Can we identify this Lie algebra? To this end, let us consider generalized Witt algebras [3]. We take $\mathbb{Q}$ as the underlying field (any field of characteristic zero would be possible) and let $|\mathcal{R}|$ be the number of amplitudes needing renormalization.

For $s \in \mathcal{R}$, consider $D_s := \partial/\partial_{x_s}$ and define

$$W_{\mathcal{R}}^+ := \text{Span}_{\mathbb{Q}}\{x^q D_s | q \in \mathbb{Z}_+^{|\mathcal{R}|}, 1 \leq s \leq |\mathcal{R}|\}. \quad (15)$$

Then, $W_{\mathcal{R}}^+$ is the derivation Lie algebra of the polynomial ring $\mathbb{Q}[x_1, \dots, x_{|\mathcal{R}|}]$ which is a simple Lie algebra of infinite dimensions and a subalgebra of the generalized Witt algebra W .

For integers $Q(t)$ as above, set

$$z_m^s := \left[\prod_{t \in \mathcal{R}} x_t^{Q(t)} \right]^m x_s \partial_{x_s}. \quad (16)$$

This puts $L_{\text{grad}} \subset W^+$. We have no space for a proper discussion of physics aspects here, but mention that we can now augment the algebra W^+ by an element Y such that $[Y, z_1^q] = z_1^q$. We then get with

$$r := Y \otimes z_1^q - z_1^q \otimes Y \quad (17)$$

a R-matrix for L_{grad} . Indeed, one immediately checks that r fulfills the Yang-Baxter equation. As direct consequence, we have Drinfel'd twists and quantized universal enveloping algebras available for future study.

2.2. A massive co-ideal

The Ward identity of quantum electrodynamics is well known: it relates the derivative w.r.t. p_μ of the inverse fermion propagator to the three-point vertex of a spin-one photon coupling at zero momentum to the corresponding pair of fermions: $\partial_{p_\mu} G^{\bar{\psi}\psi}(p^2/\mu^2, m^2) = G^{\bar{\psi}A\psi}(p^2/\mu^2, m^2)$.

The proof is entirely graphical, upon recognizing that we can route the external momentum p through the open fermion line for each graph contributing to $G^{\bar{\psi}\psi}$, and the Leibniz rule ensures that now each internal free propagator obtains a marking corresponding to the zero-momentum coupling of an external spin-one boson. Upon recognizing that this is the sum over all vertex graphs contributing to the described vertex function, we have our Ward identity.

This can be repeated similarly for the derivative $\partial_m G^{\bar{\psi}\psi} = G^{\bar{\psi}\psi\phi}$ which relates the inverse fermion propagator to the coupling at zero-momentum transfer with a spin-zero scalar.

Theorem 2. *This identity corresponds to a co-ideal in the Hopf algebra*

$$c_k^{\bar{\psi}\psi} + c_k^{\bar{\psi}\psi\phi}. \quad (18)$$

This is straightforward to prove, what is interesting here is the interplay of Hopf co-ideals in theories with scalar particle and the interference with the Hopf algebra structure of perturbative gravity [4].

2.3. Kinematics as cohomology

Let us consider a Green function G^r and let us identify a suitable variable s in general position such that

$$G_R^r(g, s, \Theta) = 1 \pm s^{\omega_r} \gamma_j^r(\{g\}, \Theta) \ln^j s, \quad (19)$$

where Θ stands for a set of generalized angles $p_i \cdot p_j/s$. We have

$$G_R^r = \Phi_{s,\Theta}^R(X^r) \quad (20)$$

for renormalized Feynman rules $\Phi_{s,\Theta}^R$ and X^r a series which is a formal fixpoint of the equation

$$X^r = \mathbb{I} \pm \sum_k g^k B_+^{r;k}(X^r Q^k). \quad (21)$$

The following theorem answers how the Hochschild cohomology of the perturbation expansion relates to the variation of physical parameters. We assume R is a kinematical subtraction scheme which renormalizes by subtraction at a fixed kinematical point, for example on shell.

Theorem 3.

$$\Phi_{s_1+s_2,\Theta}^R = \phi_{s_1,\Theta}^R \star \phi_{s_2,\Theta}^R, \quad (22)$$

and

$$\Phi_{s,\Theta'}^R \circ B_+^{r,k} = \Phi_{s,\Theta}^R \circ \left(B_+^{r,k} + L^{r,k} \right), \quad (23)$$

where $L^{r,k}$ is a exact one co-cycle: $L^{r,k} = b\phi^{r,k}$.

b is the Hochschild boundary operator, $b^2 = 0$, and $\phi^{r,k} : H \rightarrow \mathbb{C}$ are suitable functions. Note that this identifies the equivalence class of one-cocycles in the Hochschild cohomology with the space of kinematical variations at fixed scale s .

References

- [1] C. Bergbauer, Dissertation: *Combinatorial and geometric aspects of Feynman graphs and Feynman integrals*, FU Berlin 2009, (http://www.diss.fu-berlin.de/diss/receive/FUDISS-thesis_000000010972).
- [2] S. Bloch and D. Kreimer, *Mixed Hodge structures and renormalization in physics*, Comm. in Number Theory and Physics, Vol. 2.4, p.637-718; arXiv:0804.4399 [hep-th].
- [3] N. Hu, X. Wang, *Quantization of generalized Witt algebra and of Jacobson–Witt algebra in the modular case*, J. of Algebra **312** (2007) 902-929.
- [4] D. Kreimer, *A remark on quantum gravity*, Annals Phys. **323** (2008) 49 [arXiv:0705.3897 [hep-th]].
- [5] D. Kreimer, *Algebra for quantum fields*, arXiv:0906.1851 [hep-th].
- [6] D. Kreimer, W.D. van Suijlekom, *Recursive relations in the core Hopf algebra*, arXiv:0903.2849 [hep-th], Nucl. Phys. **B820** (2009), 682-693, DOI information: 10.1016/j.nuclphysb.2009.04.025.
- [7] D. Kreimer, G. Van Baalen, D. Uminsky, K. Yeats, *The QED beta-function from global solutions to Dyson-Schwinger equations*, Annals of Physics **324**, 205-219 (2008) [arXiv:0805.0826 [hep-th]].

GRAPHENE: RELATIVISTIC TRANSPORT IN A NEARLY PERFECT QUANTUM LIQUID

MARKUS MÜLLER

*The Abdus Salam International Center for Theoretical Physics,
34014 Trieste, Italy
E-mail: markusm@ictp.it
<http://users.ictp.it/~markusm/>*

LARS FRITZ and SUBIR SACHDEV

*Department of Physics, Harvard University,
Cambridge MA 02138, USA*

JÖRG SCHMALIAN

*Ames Laboratory and Department of Physics and Astronomy,
Iowa State University, Ames, IA 50011, USA*

Electrons and holes in clean, charge-neutral graphene behave like a strongly coupled relativistic liquid. The thermo-electric transport properties of the interacting Dirac quasiparticles are rather special, being constrained by an emergent Lorentz covariance at hydrodynamic frequency scales. At small carrier density and high temperatures, graphene exhibits signatures of a quantum critical system with an inelastic scattering rate set only by temperature, a conductivity with a nearly universal value, solely due to electron-hole friction, and a very low viscosity. In this regime one finds pronounced deviations from standard Fermi liquid behavior. These results, obtained by Boltzmann transport theory at weak electron-electron coupling, are fully consistent with the predictions of relativistic hydrodynamics. Interestingly, very analogous behavior is found in certain strongly coupled relativistic liquids, which can be analyzed exactly via the AdS-CFT correspondence, and which had helped identifying and establishing the peculiar properties of graphene.

Keywords: Graphene, relativistic hydrodynamics, viscosity, perfect liquids, AdS-CFT

1. Graphene - a strongly coupled, relativistic electron-hole plasma

Single layer graphene is a zero-gap semiconductor whose low energy quasiparticles obey the massless Dirac equation [1–3]. At charge neutrality, the Fermi surface reduces to two inequivalent Fermi points, forming a non-analyticity in the density of states, which can be viewed as a rather simple quantum critical point [4]. On top of that, however, as a consequence of the linear dispersion of the 2d quasiparticles Coulomb interactions are unusually strong. They are only marginally irrelevant under renormalization, flowing only logarithmically to zero with decreasing temperature T , see [5]. This is reflected, e.g., in the inelastic scattering rate being

proportional to $\alpha^2 T$ ($k_B = \hbar = 1$), where $\alpha = e^2/\kappa v_F$ is the (slowly running) dimensionless “fine structure constant” characterizing the strength of Coulomb interactions, where κ is the dielectric constant of the adjacent medium and v_F is the Fermi velocity of the linearly dispersing quasiparticles. This large scattering rate nearly saturates a Heisenberg uncertainty principle for quasiparticles [6], according to which the scattering rate is conjectured never to exceed significantly the thermal energy scale. Indeed, upon approaching $\alpha \rightarrow O(1)$ one expects a chiral symmetry breaking quantum phase transition towards an insulator [7] with very different low energy excitations. Due to the strong marginal interactions, the neutrality point of graphene is very similar to quantum critical points of more complex, strongly coupled materials [8–11]. In the quantum critical window, i.e., at small chemical potential of the carriers, $|\mu| < T$, the latter form an interacting “hot” electron-hole plasma with rather unusual transport properties which we discuss below.

1.1. Anomalously strong Coulomb scattering at charge neutrality

At finite carrier density, an estimate of the inelastic scattering rate in random phase and Born approximation leads to

$$\hbar\tau_{\text{inel}}^{-1} \sim \max(T, |\mu|) \frac{\alpha^2}{(1 + \alpha|\mu|/T)^2}, \quad (1)$$

where $\alpha \approx \alpha(\epsilon) \sim 4/\log(\Lambda/\epsilon)$ denotes the renormalized strength of Coulomb interactions α at a given energy scale $\epsilon = \max[\mu, T]$, whereby Λ is a UV cutoff. At finite μ , the scattering rate decreases rather quickly according to the familiar law $T^2/|\mu|$, independent of the interaction strength in the ultraviolet. The quantum-critical window is clearly distinguished by its strong inelastic scattering rate $\sim \alpha^2 T$, which has several interesting consequences.

As was first pointed out in the context of the superfluid-insulator quantum phase transition [11] the particle-hole symmetric point $\mu = 0$ exhibits a finite collision-dominated conductivity, even in the absence of impurities. Indeed, the application of an external electrical field induces counter propagating particle and hole currents, and thus no net momentum. The latter is usually the source of infinite current response unless the momentum decays due to impurities. However, in neutral graphene one finds a disorder-independent conductivity which is solely due to electron-hole friction. Scaling arguments based on the Drude formula, the thermal density of carriers $n_{\text{th}} \sim (T/\hbar v_F)^2$, the inelastic scattering rate and a T -dependent “effective mass” $m_{\text{eff}} \sim T/v_F^2$ suggest a conductivity which grows logarithmically with T

$$\sigma(\mu = 0) \sim \frac{e^2 n_{\text{th}} \tau_{\text{inel}}}{m_{\text{eff}}} = \frac{C}{\alpha(T)^2} \frac{e^2}{h}. \quad (2)$$

This is indeed confirmed by a microscopic calculation based on the semiclassical Boltzmann equation, which becomes asymptotically exact for $T \ll \Lambda$ where the coupling is $\alpha \ll 1$, yielding the prefactor $C = 0.760$ [9, 12].

For the same reason as the electrical conductivity remains finite at particle-hole symmetry, the thermal conductivity κ diverges at $\mu = 0$. For the case of relativistically invariant systems this has been shown by Vojta *et al.* [13]. κ describes the heat current response to a thermal gradient in the absence of an electrical current. Usually, the latter forbids the excitation of a finite, non-decaying momentum, and this ensures a finite heat current response. At particle-hole symmetry, however, the momentum created by a thermal gradient does not carry a net current and is thus not affected by the boundary condition. It follows that within the bulk of a sample a thermal gradient cannot be sustained at $\mu = 0$ (see Ref. [14] for a discussion of κ in a sample coupled to leads). For graphene, both relativistic hydrodynamics [15] and Boltzmann theory yield the leading divergence

$$\kappa(\mu \rightarrow 0) = \frac{\sigma(\mu = 0)}{T} \left(\frac{P + \varepsilon}{\rho} \right)^2, \quad (3)$$

P, ε and ρ being the pressure, energy density and charge density of the fluid, respectively. This relation can be interpreted as a relativistic Wiedemann-Franz-like relation between σ and κ .

1.2. Graphene as a nearly perfect liquid

A further consequence of the strong Coulomb coupling in graphene, and more generally, of quantum criticality, is the anomalously low value of the shear viscosity η . Its ratio to the entropy density, η/s is the crucial parameter in the Navier-Stokes equation which controls the occurrence of turbulence via the Reynolds number

$$\text{Re} = \frac{s/k_B}{\eta/\hbar} \times \frac{k_B T}{\hbar v/L} \times \frac{u_{\text{typ}}}{v_F}, \quad (4)$$

where L is a typical length and u_{typ} a typical velocity scale of the electronic current flow. The tendency towards electronic turbulence is stronger the larger is Re. Full-fledged turbulence might require $\text{Re} > 10^4$ in 2d, but interesting, complex flow is already expected at experimentally accessible values $\text{Re} \sim 10^2 - 10^3$ [16].

Viscosity having the units of $\hbar \cdot n$ with n a density the ratio has units of $\hbar/k_B$. For massless fermions or bosons, the coefficient of proportionality is essentially the mean free path divided by the thermal de Broglie wavelength. This ratio is usually large, but becomes of order $O(1)$ when the scattering rate tends to saturate Heisenberg's uncertainty relation. For certain strongly coupled relativistic liquids the low value $1/4\pi$ was obtained via the AdS-CFT correspondence. Interestingly, a similarly low value is found for graphene when the weak coupling result $\eta/s = 0.13\hbar/k_B\alpha^2(T)$ is extrapolated to values of order $\alpha \rightarrow O(1)$. [16]

At charge neutrality, $(\eta/s)^{-1}$ is thus a rather direct measure for the dimensionless interaction strength g (equal to α at $\mu = 0$), and can serve as an experimental indicator for the latter. At finite carrier density, $\mu > T$, the ratio η/s behaves however as $|\mu/T|^3$, and thus does not directly inform about the strength of the

dimensionless coupling g . A better measure is instead furnished by the estimate

$$\frac{1}{g^2} \propto \frac{k_B \eta}{\hbar \rho} \left(\frac{T}{\mu} \right)^2. \quad (5)$$

One may speculate that in this case η is “softly” bounded from below by the restriction $g \lesssim O(1)$.

1.3. Emergent relativistic hydrodynamics in the infrared

While the quasiparticles of graphene have relativistic kinematics and the non-interacting theory enjoys Lorentz-invariance, the instantaneous Coulomb interactions obviously break this symmetry. However, one may expect that at hydrodynamic frequency and length scales, the relativistic invariance re-emerges. This is indeed the case, since the role of interactions mainly consists in establishing local equilibrium, while the relaxation towards global equilibrium is governed by the conservation laws of charge, momentum and energy. The latter essentially remain those of a relativistic fluid, thus fixing the structure of all thermoelectric response functions at small frequencies and wavevectors [15, 17–19].

It can be shown [20] that the specificity of the interactions enters the low-energy hydrodynamic description only via the values of the two dissipative transport coefficients σ and η discussed above for $\mu = 0$. A third coefficient, the bulk viscosity, is very small in graphene due to its near scale invariance.

A microscopic analysis [20] shows that a hydrodynamic description still applies, even when translational invariance and thus momentum conservation is weakly broken by a small magnetic field and dilute disorder. The necessary condition is that the scattering rates associated with those perturbations are significantly smaller than τ_{inel}^{-1} . An interesting prediction of the resulting magneto-hydrodynamics is the presence of a collective cyclotron resonance [17, 19]. Its origin is a collective circulating motion of the electron-hole plasma with a frequency proportional to its net charge density ρ . In contrast to systems with Galilean invariance with a single type of charge carrier, for which Kohn’s theorem ensures a sharp, undamped cyclotron resonance, the cyclotron motion in graphene is damped due to electron-hole friction.

2. Comparison with strongly coupled liquids solved by AdS-CFT

Our investigations of interaction effects in graphene were to a large extent stimulated by very similar physical phenomena in strongly coupled critical systems and ultrarelativistic matter such as the quark-gluon plasma. [17, 21] Indeed, it is interesting to confront and compare the properties of graphene with results obtained for relativistic supersymmetric $SU(N)$ gauge theories in $2 + 1$ dimensions, with which one hopes to catch some of the physically relevant features of the quark gluon plasma. These gauge theories are strongly coupled field theories with an emerging conformal symmetry in the infrared. The $\text{AdS}_3\text{-CFT}_{2+1}$ correspondence stipulates

that this theory is exactly dual to an Einstein-Maxwell gravity theory in $3 + 1$ dimensions, which in the large N limit becomes exactly solvable. Among others, this allows one to compute transport coefficients of the gauge theory [17, 21, 22]

$$\sigma = \left(\frac{2}{9}\right)^{1/2} N^{3/2} \frac{e^2}{h} ; \quad s, \eta \propto N^{3/2} T^2 ; \quad \frac{\eta}{s} = \frac{1}{4\pi} . \quad (6)$$

These results can be physically interpreted as reflecting properties of $N^{3/2}$ effective degrees of freedom [23], whose strong interactions result in an inelastic scattering time which essentially saturates the Heisenberg-like bound $\tau T \geq O(1)$. The AdS-CFT correspondence also confirmed the predictions of relativistic magnetohydrodynamics, including the cyclotron resonance, as discussed above. [17]

3. Conclusion

Graphene exhibits several peculiar transport properties which it shares with other quantum critical systems as well as with ultrarelativistic matter. The crucial ingredients behind these phenomena are anomalously strong Coulomb interactions in 2d and the relativistic kinematics. Experimental verifications at moderately high T should soon be within reach due to the availability of clean suspended graphene.

References

- [1] G. W. Semenoff, Phys. Rev. Lett. **53**, 2449 (1984).
- [2] F. D. M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988).
- [3] S. Y. Zhou, G.-H. Gweon, J. Graf, A. V. Fedorov, C. D. Spataru, R. D. Diehl, Y. Kopelevich, D.-H. Lee, S. G. Louie, and A. Lanzara, Nature Phys. **2**, 595 (2006).
- [4] D. E. Sheehy and J. Schmalian, Phys. Rev. Lett. **99**, 226803 (2007).
- [5] J. Gonz  les, F. Guinea, and M. A. H. Vozmediano, Nucl. Phys. B **424**, 595 (1994); Phys. Rev. B **59**, R2474 (1999).
- [6] S. Sachdev, *Quantum Phase Transitions* Cambridge University Press, Cambridge, (1999).
- [7] J. E. Drut and T. A. L  hde, Phys. Rev. Lett. **102**, 026802 (2009).
- [8] D. E. Sheehy and J. Schmalian, Phys. Rev. Lett. **99**, 226803 (2007).
- [9] L. Fritz, J. Schmalian, M. M  ller, and S. Sachdev, Phys. Rev. B **78**, 085416 (2008).
- [10] S. Sachdev, Phys. Rev. B **57**, 7157 (1998).
- [11] K. Damle and S. Sachdev, Phys. Rev. B **56**, 8714 (1997).
- [12] A. Kashuba, Phys. Rev. B **78**, 085415 (2008).
- [13] M. Vojta, Y. Zhang, and S. Sachdev, Intl. Jour. of Mod. Phys. B **14**, 3719 (2000).
- [14] M. S. Foster and I. L. Aleiner, Phys. Rev. B **77**, 195413 (2008).
- [15] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Butterworth-Heinemann, Oxford (1987), Section 127.
- [16] M. M  ller, J. Schmalian, and L. Fritz, Phys. Rev. Lett. **103**, 025301 (2009).
- [17] S. A. Hartnoll, P. K. Kovtun, M. M  ller, and S. Sachdev, Phys. Rev. B **76**, 144502 (2007).
- [18] L. P. Kadanoff and P. C. Martin, Annals of Physics, **24**, 419 (1963).
- [19] M. M  ller and S. Sachdev, Phys. Rev. B **78**, 115419 (2008).
- [20] M. M  ller, L. Fritz, and S. Sachdev, Phys. Rev. B **78**, 115406 (2008).
- [21] P. Kovtun, D. T. Son and A. Starinets, Phys. Rev. Lett. **94**, 11601 (2005).

- [22] S. A. Hartnoll and C. P. Herzog, Phys. Rev. D **76**, 106012 (2007).
- [23] I. R. Klebanov and A. A. Tseytlin, Nucl. Phys. B **475**, 164 (1996).
- [24] C. P. Herzog, P. Kovtun, S. Sachdev and D. T. Son, Phys. Rev. D **75** 085020, (2007).

DEFECT LINES, DUALITIES AND GENERALISED ORBIFOLDS

JÜRG FRÖHLICH¹, JÜRGEN FUCHS², INGO RUNKEL^{3,4}, and
CHRISTOPH SCHWEIGERT⁴

¹ *Theoretische Physik, ETH Zürich,
8093 Zürich, Switzerland*

² *Teoretisk fysik, Karlstads Universitet,
Universitetsgatan 21, 65188 Karlstad, Sweden*

³ *Department of Mathematics, King's College London,
Strand, London WC2R 2LS, United Kingdom*

⁴ *Department Mathematik, Universität Hamburg,
Bundesstraße 55, 20146 Hamburg, Germany*

Defects are a useful tool in the study of quantum field theories. This is illustrated in the example of two-dimensional conformal field theories. We describe how defect lines and their junction points appear in the description of symmetries and order-disorder dualities, as well as in the orbifold construction and a generalisation thereof that covers exceptional modular invariants.

Keywords: Two-dimensional conformal field theory, defect lines, dualities, orbifolds.

1. Introduction

Defects in a quantum field theory or in a model of statistical mechanics correspond to ‘inhomogeneities’ in correlators localised on hypersurfaces. Expectation values of fields can have discontinuities or singularities at such hypersurfaces. Defects appear naturally in connection with duality symmetries. For example, in electric-magnetic duality of four-dimensional supersymmetric Yang-Mills theory [1, 2], Wilson loop operators get exchanged with ’t Hooft operators, and the latter prescribe a singularity of the gauge field along a line. Another example, more in line with the present exposition, is Kramers-Wannier duality of the two-dimensional Ising model on a square lattice [3, 4]. In that context the correlator of Ising spins at lattice sites i and j at a given temperature is equal to the (suitably normalised) partition function of the Ising model on a lattice with a line of frustration running from i to j , evaluated at the dual temperature. Across the line of frustration, the sign of the interaction of neighbouring spins is inverted.

We will concentrate on one-dimensional defects in two-dimensional conformal field theory (CFT). These describe universality classes of defect lines in lattice models, but they are also a useful tool for investigating properties of the CFT itself.

At first glance, it seems that defect lines will not give more insight than the study of boundary conditions, because by ‘folding’ the surface back onto itself along a defect line, the defect becomes a boundary of the folded model [5]. However, defects have one important property not shared by boundaries, and which cannot be captured by the ‘folding trick’: an arbitrary number of defect lines can join at *defect junctions*. Below we will outline two applications in which defect junctions are important. One is the description of order-disorder dualities via defects, and the other is the construction of orbifold models.

2. Symmetries and order-disorder dualities

At a defect line on the world sheet one must specify boundary conditions for the fields on either side of the defect; we refer to these as *defect conditions*. To a defect with defect condition X one can assign an operator D_X on the space $\mathcal{H}$ of bulk states of the CFT. This operator can be defined by placing a state $\phi \in \mathcal{H}$ at zero and the defect X on the unit circle; the resulting state is $D_X \phi$. There are two compatibility conditions between a defect X and the conformal symmetry. Denoting the modes of the holomorphic and anti-holomorphic component of the stress tensor by L_m and $\bar{L}_m$, we call X *conformal* iff $[L_m - \bar{L}_{-m}, D_X] = 0$ for all $m \in \mathbb{Z}$, and we call it *topological* iff it satisfies the stronger condition $[L_m, D_X] = 0 = [\bar{L}_m, D_X]$ for all $m \in \mathbb{Z}$. Thus a defect is conformal iff the corresponding boundary condition in the folded model is conformal, and it is topological iff it is transparent to all components of the stress tensor. The qualifier ‘topological’ derives from the fact that such defects are tensionless and can be deformed on the world sheet without affecting the value of correlation functions, as long as the defect line is not taken across field insertions or other defect lines.

Finding all conformal defects for a given CFT is a difficult problem. The only CFTs for which all conformal defects are known are the Lee-Yang model of central charge $-\frac{22}{5}$ and the Ising model^a of central charge $\frac{1}{2}$ [6, 7]. Even for a single free boson, only a subset of conformal defects is known [8]. Topological defects preserve more symmetry, and for these the situation is better: one knows all topological defects for the Virasoro minimal models [9] and for the free boson [10].

Consider, for example, the Ising model. It has three elementary topological defects, labelled $\mathbf{1}$, σ , ε ; all other topological defects are superpositions of these. The defect $\mathbf{1}$ is just the trivial defect, and its defect operator $D_{\mathbf{1}}$ is the identity. The two other defect operators have the composition rules $D_{\varepsilon} D_{\varepsilon} = D_{\mathbf{1}}$, $D_{\varepsilon} D_{\sigma} = D_{\sigma} D_{\varepsilon} = D_{\sigma}$, and $D_{\sigma} D_{\sigma} = D_{\mathbf{1}} + D_{\varepsilon}$. The composition of defect operators for defects X and Y can be realised by placing the defect Y on the unit circle, the defect X on a circle with radius $r > 1$ and taking the limit $r \rightarrow 1$. For topological defects, this limit is well-defined (the correlator is independent of r), and in the limit one obtains a new defect on the unit circle, called the *fused defect*, whose

^a Strictly speaking, only conformal defects that have a discrete content of defect fields are known [7, Sect. 4.1].

Thus, to find order-disorder dualities, one does not need to derive the (possibly complicated) rules for taking defects past field insertions, as in (1). Rather, it is enough to know the behaviour of the defect operators under composition.

3. Orbifolds

Denote by G the group formed by the group-like defects of a rational CFT. The group G automatically comes with further data, namely a class $[\psi]$ in $H^3(G, \mathbb{C}^\times)$, the third cohomology of G with values in $\mathbb{C}^\times$ (with trivial G -action). This holds because group-like defects are described by invertible objects in a tensor category [14], and their associator gives rise to a 3-cocycle [15, App. E]. This 3-cocycle can also be seen directly on the level of correlators by employing defect junctions in the following way [16]. For each pair $g, h \in G$ pick a non-zero topological junction field $\varphi_{g,h}$ for a 3-fold junction with defects g and h oriented towards the junction, and defect $g \cdot h$ oriented away from it. The space of such junction fields is one-dimensional [13]. Similarly, for three group-like defects $g, h, k \in G$, the space of topological junction fields for the four-fold junction with three in-coming defects labelled g, h, k , and one out-going defect $g \cdot h \cdot k$ is one-dimensional, too. As a consequence, the following proportionality between two configurations of defect lines and topological junction fields holds:

$$\begin{array}{c} \varphi_{gh,k} \\ \nearrow \quad \searrow \\ \varphi_{g,h} \quad \quad \quad ghk \\ \nearrow \quad \searrow \\ g \quad \quad h \quad \quad k \end{array} = \psi(g, h, k) \cdot \begin{array}{c} ghk \\ \nearrow \quad \searrow \\ \varphi_{g,hk} \quad \quad \varphi_{h,k} \\ \nearrow \quad \searrow \\ g \quad \quad h \quad \quad k \end{array} \quad (2)$$

The pentagon identity for four-to-one junctions built from the $\varphi_{g,h}$ shows that ψ is a three-cocycle. Rescaling the junction fields $\varphi_{g,h}$ modifies ψ by a coboundary. The basis-independent information is thus the class $[\psi] \in H^3(G, \mathbb{C}^\times)$.

The class $[\psi]$ provides an obstruction to orbifolding: one can only orbifold the CFT by subgroups $H \subseteq G$ such that $[\psi|_H] = 1$. The origin of this obstruction can be understood as follows. Fix a subgroup $H \subseteq G$. To define the correlators of the H -orbifold theory, consider the superposition $Q = \sum_{h \in H} h$ of group-like defects. Then $\varphi = \sum_{g,h \in H} \varphi_{g,h}$ is a topological junction field for a two-to-one junction with all three legs labelled by Q . We will also need a topological junction field $\bar{\varphi}$ in the opposite direction, i.e. a one-to-two junction with all legs labelled Q :

$$\begin{array}{c} Q \\ \nearrow \quad \searrow \\ \varphi \\ \nearrow \quad \searrow \\ Q \quad \quad Q \end{array} \quad \text{and} \quad \begin{array}{c} Q \quad \quad Q \\ \nearrow \quad \searrow \\ \bar{\varphi} \\ \nearrow \quad \searrow \\ Q \end{array} \quad (3)$$

Correlators of the orbifold model can be described as correlators of the un-orbifolded model with a sufficiently fine network of Q -defect lines inserted. These defect lines are joined at three-fold junctions and are oriented such that at each junction one can insert either φ or $\bar{\varphi}$.

For example, on the torus one can take

$$Z_{\text{orb}} = \left[\text{Diagram of a square with a defect network inside} \right] . \quad (4)$$

For this description to be independent of the choice of defect network, it is clearly necessary that the defects and junction fields are topological. In addition, we must require invariance of the correlator under the following two local changes of the defect network:

$$\left[\text{Diagram 1} \right] \longleftrightarrow \left[\text{Diagram 2} \right] \quad \text{and} \quad \left[\text{Diagram 3} \right] \longleftrightarrow \left[\text{Diagram 4} \right] \quad (5)$$

A defect network with three-valent vertices is dual to a triangulation of the surface, and the two local moves (5) allow one to transform any dual triangulation into any other. The identities (5) must hold for all allowed orientations of the defect lines. Expanding out the sum hidden in the superposition Q , we see that for three incoming defects and one out-going defect, (2) implies invariance under the first move in (5), provided that $\psi(g, h, k) = 1$ for all $g, h, k \in H$. A choice of $\varphi_{g,h}$ for which this is the case exists iff $[\psi|_H] = 1$. One can convince oneself that then also $\bar{\varphi}$ exists such that (5) holds for all allowed orientations [14].^b

In addition, any two possibilities of choosing $\varphi_{g,h}$ such that $\psi|_H = 1$ are related by a 2-cocycle on H , and if we modify $\varphi_{g,h}$ by a factor that is a coboundary, this modification will cancel from (4), and from all other correlators as well. One finds that different H -orbifolds are in one-to-one correspondence to $H^2(H, \mathbb{C}^\times)$ [14]. This freedom is known as discrete torsion [17].

Recall that the crucial point in the construction of the orbifold correlators is the existence of a topological defect Q with topological junction fields (3), such that the modifications (5) leave correlators invariant; it is not actually necessary for Q to be a superposition of group-like defects. This opens the avenue for a generalisation of the orbifold construction by allowing more general topological defects Q , and thereby leads to the following result:

Every rational CFT that has identical left and right chiral symmetry V , that is well-defined on (oriented, closed) surfaces of genus 0 and 1, and that has a unique vacuum and non-degenerate 2-point correlator is a generalised orbifold of the Cardy case for V (for which in particular the modular invariant torus partition function is given by charge-conjugation).

^bOne also needs that D_g applied to the bulk vacuum $|0\rangle$ gives $|0\rangle$ rather than $-|0\rangle$; this is automatic in unitary models.

To see this, first note that in the TFT formalism [11, 18] a rational CFT with left and right chiral symmetry V is described by a special symmetric Frobenius algebra in the category of V -modules. This Frobenius algebra provides us with the defect Q and the vertices φ and $\bar{\varphi}$. Comparing the construction of the correlator in terms of a Frobenius algebra [18] and in the presence of a defect network [13], one finds that they are exactly identical. Finally, every rational CFT satisfying the above conditions can indeed be described by a Frobenius algebra [19, 20].

Actually, one does not need to use the Cardy case as a starting point. Rather, every rational CFT satisfying the above conditions can be obtained from any other such CFT as a generalised orbifold. In the language of [11, 13, 18], if $\text{CFT}(A)$ and $\text{CFT}(B)$ denote CFTs described by the special symmetric Frobenius algebras A and B , respectively, then $\text{CFT}(B)$ can be obtained as a generalised orbifold from $\text{CFT}(A)$ via the defect^c $Q = A \otimes B \otimes A$.

To summarise: the study of topological defects and defect junctions leads to a generalisation of the concept of an orbifold which provides a uniform description of rational CFTs with left and right chiral symmetry V , including, in particular, exceptional modular invariants.

References

- [1] C. Montonen and D.I. Olive, Phys. Lett. B **72** (1977) 117–120.
- [2] A. Kapustin and E. Witten, hep-th/0604151.
- [3] H.A. Kramers and G.H. Wannier, Phys. Rev. **60** (1941) 252–262.
- [4] R. Savit, Rev. Mod. Phys. **52** (1980) 453–487.
- [5] E. Wong and I. Affleck, Nucl. Phys. B **417** (1994) 403–438.
- [6] M. Oshikawa and I. Affleck, Nucl. Phys. B **495** (1997) 533–582.
- [7] T. Quella, I. Runkel and G.M.T. Watts, JHEP **0704** (2007) 095.
- [8] C. Bachas, J. de Boer, R. Dijkgraaf and H. Ooguri, JHEP **0206** (2002) 027.
- [9] V.B. Petkova and J.-B. Zuber, Phys. Lett. B **504** (2001) 157–164.
- [10] J. Fuchs, M.R. Gaberdiel, I. Runkel and C. Schweigert, J. Phys. A **40** (2007) 11403.
- [11] J. Fuchs, I. Runkel and C. Schweigert, Nucl. Phys. B **646** (2002) 353–497.
- [12] J. Fröhlich, J. Fuchs, I. Runkel and C. Schweigert, Phys. Rev. Lett. **93** (2004) 070601.
- [13] J. Fröhlich, J. Fuchs, I. Runkel and C. Schweigert, Nucl. Phys. B **763** (2007) 354–430.
- [14] J. Fuchs, I. Runkel and C. Schweigert, Nucl. Phys. B **694** (2004) 277–353.
- [15] G.W. Moore and N. Seiberg, Commun. Math. Phys. **123** (1989) 177–254.
- [16] I. Runkel and R.R. Suszek, 0808.1419 [hep-th].
- [17] C. Vafa, Nucl. Phys. B **273** (1986) 592–606.
- [18] J. Fjelstad, J. Fuchs, I. Runkel and C. Schweigert, Theo. Appl. Cat. **16** (2006) 342–433.
- [19] J. Fjelstad, J. Fuchs, I. Runkel and C. Schweigert, Adv. Theor. Math. Phys. **12** (2008) 1283–1375.
- [20] L. Kong and I. Runkel, Commun. Math. Phys. **292** (2009) 871–912.

^c Q is an A -bimodule with left and right action of A given by m_A , and $\varphi = \text{id}_A \otimes (m_B \circ [\text{id}_B \otimes (\varepsilon_A \circ m_A) \otimes \text{id}_B]) \otimes \text{id}_A$, where m denotes the multiplication and ε the counit, etc.

ALGEBRAIC BETHE ANSATZ APPROACH TO CORRELATION FUNCTIONS: FROM EXACT REPRESENTATIONS ON THE LATTICE TO ASYMPTOTIC BEHAVIOR

VÉRONIQUE TERRAS

Laboratoire de Physique, UMR5672,

CNRS & ENS Lyon, France

E-mail: veronique.terras@ens-lyon.fr

We review recent results concerning the computation, from first principles, of correlation functions of quantum integrable systems. On the example of the XXZ Heisenberg chain, we first recall how to obtain, in the algebraic Bethe Ansatz framework, exact representations for the two-point functions in the finite chain. We then explain how to analyze these expressions so as to derive, in the thermodynamic limit and in the massless regime of the chain, the long distance asymptotic behavior of these two-point functions, hence confirming, at leading order, the Luttinger liquid and conformal field theory predictions.

Keywords: Quantum integrable systems; Algebraic Bethe Ansatz; Correlation functions; Asymptotic analysis.

1. Introduction

One of the central question in the theory of quantum integrable models is the exact computation of their correlation functions. In order to briefly review recent advances on this subject [1–6], we consider the simple but representative example of the XXZ spin-1/2 Heisenberg chain with periodic boundary conditions:

$$H = \sum_{m=1}^M \left\{ \sigma_m^x \sigma_{m+1}^x + \sigma_m^y \sigma_{m+1}^y + \Delta (\sigma_m^z \sigma_{m+1}^z - 1) \right\} - h \sum_{m=1}^M \sigma_m^z. \quad (1)$$

Here Δ is an anisotropy parameter, $\sigma_m^{x,y,z}$ are local spin operators (acting as the corresponding Pauli matrices) at site m , and h is an external magnetic field.

There exist two different types of results for the correlation functions of such a model: on the one hand, *exact representations* have been obtained by studying the symmetry of the model on the (finite or infinite) lattice [1–3, 7–10]; on the other hand, for the infinite lattice at zero-temperature and in the massless regime $|\Delta| < 1$, the (power-law) *asymptotic behavior* for the two-point functions at large distances has been predicted from bosonization and Luttinger liquid approximation [11, 12] and conformal field theory analysis [13, 14]. However, until last year, it was still an open problem to connect the two approaches and to confirm the asymptotic predictions from the exact results on the lattice. This is the purpose of this review

to explain how this can be done, starting from the algebraic Bethe Ansatz study of the model on the finite lattice.

2. Correlation Functions in the Algebraic Bethe Ansatz Framework

At zero temperature, correlation functions reduce to mean values $\langle \psi_g | \mathcal{O} | \psi_g \rangle$ of some product $\mathcal{O}$ of local operators in the ground state $|\psi_g\rangle$. The computation of such quantities supposes to be able to solve successively the following problems:

- (i) characterization of the ground state $|\psi_g\rangle$ (or more generally of the eigenstates of the Hamiltonian),
- (ii) evaluation of the action of the product $\mathcal{O}$ of local operators on this ground state, and characterization of the obtained state $|\psi_{\mathcal{O}}\rangle = \mathcal{O}|\psi_g\rangle$,
- (iii) computation of the resulting scalar product $\langle \psi_g | \psi_{\mathcal{O}} \rangle$.

Step (i) can be solved through algebraic Bethe Ansatz [15]. The key point of this approach is the (quadratic) Yang-Baxter algebra, generated by the operator entries $A(\lambda)$, $B(\lambda)$, $C(\lambda)$, $D(\lambda)$ (depending of some spectral parameter λ) of the quantum monodromy matrix. In particular, eigenstates of the Hamiltonian are constructed by multiple action $\prod_i B(\lambda_i)$ on the ferromagnetic state $|0\rangle$ with all spins up, provided the corresponding set of spectral parameters $\{\lambda_i\}$ satisfies the Bethe equations.

The problem is that the creation operators $B(\lambda)$ have a very complicated structure when expressed in terms of the local fields, which makes step (ii) *a priori* highly non-trivial. One of the main ingredients of our method, which enables us to compute the action of local operators on such highly non-local states, is the *solution of the quantum inverse problem* [1, 16], through which local operators are expressed in terms of the generators of the quantum Yang-Baxter algebra. It is then enough to use the quadratic commutation relations of these operators to complete step (ii).

The computation of the resulting scalar product (step (iii)) may also raise some technical difficulties. In the case of the XXZ spin chain, it was obtained as some determinant of usual functions by means of the algebraic structure of the Yang-Baxter algebra [1, 17].

This strategy enabled us to obtain determinant representations for the form factors of local spin operators in the finite chain [1], and multiple integral representations for the elementary building blocks of the correlation functions of the infinite chain at zero-temperature [2].

3. Exact Representation for the Two-point Function

The strategy mentioned above may in principle be applied to the computation of more physical correlation functions, such as the two-point functions $\langle \sigma_1^\alpha \sigma_{m+1}^\beta \rangle \equiv \langle \psi_g | \sigma_1^\alpha \sigma_{m+1}^\beta | \psi_g \rangle$, $\alpha, \beta \in \{+, -, z\}$, which can be experimentally measured by neutron or photon scattering experiments. However, the expressions obtained after steps (ii) and (iii) happen to be highly complicated from a combinatorial viewpoint,

and compact representations can be obtained only after performing some summation, either over elementary blocks or over form factors. It happens in fact that both types of summation result in the same exact representation (called *master equation*) for the two-point function in the finite chain [3].

For example, the *generating function* $\langle e^{\beta \mathcal{Q}_m} \rangle$ of the $\langle \sigma_1^z \sigma_{m+1}^z \rangle$ correlation function^a, with $\mathcal{Q}_m = \frac{1}{2} \sum_{n=1}^m (1 - \sigma_n^z)$, has the following master equation representation:

$$\langle e^{\beta \mathcal{Q}_m} \rangle = \frac{(-1)^N}{N!} \oint_{\Gamma(\{\lambda\})} \prod_{j=1}^N \left(\frac{dz_j}{2\pi i} e^{im[p_0(z_j) - p_0(\lambda_j)]} \frac{d(z_j)}{d(\lambda_j)} \right) \times \frac{\left[\det_N \Omega_\beta(\{z\}, \{\lambda\} | \{z\}) \right]^2}{\prod_{j=1}^N \mathcal{Y}_\beta(z_j | \{z\}) \cdot \det_N \frac{\partial \mathcal{Y}(\lambda_j | \{\lambda\})}{\partial \lambda_k}}. \quad (2)$$

In (2), p_0 stands for the quasi-particle bare momentum, d is the eigenvalue of the operator D on the reference state $|0\rangle$, and the integration contours surround poles at $z_j = \lambda_1, \dots, \lambda_N$ (which appear in $\det \Omega_\beta$), where $\{\lambda\}$ corresponds to the solution of the system of Bethe equations $\mathcal{Y}(\lambda_j, \{\lambda\}) = 0$, $j = 1, \dots, N$, for the ground state. Note that formula (2) can be seen as a sum over some β -deformed form-factors if evaluated by the poles outside the contours: these poles appear when all functions $\mathcal{Y}_\beta = 0$ in the denominator, which corresponds to the solutions of some β -twisted Bethe equations, the scalar product between the corresponding β -deformed Bethe states and the ground state being given by $\det \Omega_\beta$.

In order to take the thermodynamic limit $M \rightarrow \infty$, it is convenient to expand (2) into series. The idea is to single out the poles at $z_j = \lambda_k$ by factorizing Cauchy determinants out of each $\det \Omega_\beta$. The details of this procedure are given in Ref. [5]. In the thermodynamic limit $M \rightarrow \infty$, $N \rightarrow \infty$ with N/M finite, and in the massless regime $|\Delta| < 1$, the ground state Bethe roots $\lambda_1, \dots, \lambda_N$ condensate into some density function $\rho(\lambda)$ on some symmetric interval $[-q, q] \subset \mathbb{R}$, and one obtains

$$\langle e^{\beta \mathcal{Q}_m} \rangle = \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{-q}^q \frac{d^n \lambda}{(2i\pi)^n} \oint_{\Gamma([-q, q])} \frac{d^n z}{(2i\pi)^n} \mathcal{F}_n^{(\beta)}(\{\lambda\} | \{z\}) \times \det_n \left[\frac{e^{im[p_0(z_j) - p_0(\lambda_j)]}}{\sinh(z_j - \lambda_j) \sinh(z_j - \lambda_k)} \right]. \quad (3)$$

Here $\mathcal{F}_n^{(\beta)}$ is a symmetric function in the n variables $\{\lambda\}$ and in the n variables $\{z\}$ separately, satisfying moreover some reduction properties at “ $z_j = \lambda_k$ ”. The contour $\Gamma([-q, q])$ surrounds the interval $[-q, q]$. Note that the Fermi boundary q remains finite in the case of a non-zero magnetic field h , which is the case we consider for our asymptotic study.

^a $\langle \sigma_1^z \sigma_{m+1}^z \rangle$ can be obtained in terms of the second lattice derivative and second derivative with respect to β at $\beta = 0$ of $\langle e^{\beta \mathcal{Q}_m} \rangle$.

4. Asymptotic Analysis of the Two-Point Function

The asymptotic analysis of (3) takes its roots in the following remark: there exists a case in which expressions of the form (3) are simple to analyze in the large m limit, namely when the function $\mathcal{F}_n^{(\beta)}$ is completely decoupled and of the form $\prod_{i=1}^n [\varphi(\lambda_i) e^{g(z_i)}]$, for some functions φ, g holomorphic in some vicinity of $[-q, q]^b$. In that case, it is not difficult to see that the expression (3) corresponds to an expansion of the *Fredholm determinant* of an integral operator $I + V^{(\varphi, g)}$, $V^{(\varphi, g)}$ being a *generalized sine-kernel* depending on the distance m ,

$$V^{(\varphi, g)}(\lambda, \mu) = \varphi(\lambda) e^{g(\lambda)} \frac{\sin \left\{ \frac{m}{2} [p_0(\lambda) - p_0(\mu)] - \frac{i}{2} [g(\lambda) - g(\mu)] \right\}}{\pi \sinh(\lambda - \mu)}. \quad (4)$$

Although non-trivial, the large m asymptotic analysis of such Fredholm determinant is possible, and was performed in Ref. [4] by Matrix Riemann-Hilbert Problems, as a generalization of a procedure by Deift, Its and Zhou for the sine-kernel [18].

In the case of a general anisotropy parameter Δ , the function $\mathcal{F}_n^{(\beta)}$ is not decoupled, and the series (3) cannot be reduced to a single Fredholm determinant. It can nevertheless be rearranged in a specific way, making possible the asymptotic analysis of its separate terms. The idea is to expand the Cauchy determinant appearing in (3) as a sums over permutations and to reorder the variables λ and μ under the integrals using also the symmetry property of the function $\mathcal{F}_n^{(\beta)}$. Such a reordering leads to a series over *cycle integrals* of length s , $1 \leq s \leq n$, of the form

$$\mathcal{I}_s[\mathcal{G}_s] = \oint_{\Gamma([-q, q])} \frac{d^s z}{(2i\pi)^s} \int_{-q}^q \frac{d^s \lambda}{(2i\pi)^s} \mathcal{G}_s \left(\begin{matrix} \{\lambda\} \\ \{z\} \end{matrix} \right) \prod_{j=1}^s \frac{e^{im[p_0(\lambda_j) - p_0(z_j)]}}{\sinh(z_j - \lambda_j) \sinh(z_j - \lambda_{j+1})}, \quad (5)$$

in which $\lambda_{s+1} \equiv \lambda_1$ and $\mathcal{G}_s$ stands for some restriction of the function $\mathcal{F}_n^{(\beta)}$.

The large m asymptotic expansion for each cycle integral follows for the large m asymptotic study of the Fredholm determinant with kernel (4) [4]. The asymptotic summation of the resulting series of cycle integrals is a non-trivial process, which uses in particular the reduction properties of the function $\mathcal{F}_n^{(\beta)}$ at $z_j = \lambda_k$ [5]. Let us merely mention that some part of the series exponentiates, whereas some other part can be summed up as a continuous generalization of the multiple Lagrange series: its sum is expressed in terms of a solution of an integral equation, which corresponds here to the integral equation for the dressed charge $Z(\lambda)$ of the chain.

As a result of this process, we obtain [5]

$$\langle e^{\beta \mathcal{Q}_m} \rangle = \sum_{\sigma=0, \pm} G_m^{(0)}(\beta + 2\pi i \sigma) [1 + o(1)], \quad G_m^{(0)}(\beta) = \mathcal{C}(\beta) e^{\beta m \frac{p_F}{\pi}} m^{\frac{\beta^2 \mathcal{Z}^2}{2\pi^2}}. \quad (6)$$

Here p_F is the Fermi momentum, $\mathcal{Z}$ is the value of the dressed charge at the Fermi boundary, and the constant $\mathcal{C}(\beta)$ can be computed in terms of a ratio of four Fredholm determinants. Taking second lattice derivative and second derivative with

^bThis is in particular the case at the free fermion point $\Delta = 0$, for which $\mathcal{F}_n^{(\beta)}$ is constant.

respect to β at $\beta = 0$ of (6), we get, at the leading order in m ,

$$\langle \sigma_1^z \sigma_{m+1}^z \rangle_{\text{leading}} = \langle \sigma^z \rangle^2 - \frac{2Z^2}{\pi^2 m^2} + 2|F_\sigma|^2 \cdot \frac{\cos(2mp_F)}{m^2 Z^2}, \quad (7)$$

where F_σ is related to the properly normalized form factor of the operator σ^z between the ground state and an excited state containing one particle and one hole sited at the two different boundaries of the Fermi sphere, namely at q and $-q$ [6].

5. Conclusion and Open Problems

The result (7) agrees with bosonization and conformal field theory predictions [11–14]. A problem which is still open is to compare it with existing predictions for the amplitude [19, 20]: this supposes to compute the $\hbar \rightarrow 0$ limit of the constant F_σ , which is not an easy task due to the divergence of the Fredholm determinants in $\mathcal{C}(\beta)$. Another open problem would be to generalize this asymptotic result to other correlation functions, such as $\langle \sigma_1^+ \sigma_{m+1}^- \rangle$, or to more complicated models. Finally, a problem under investigation is to find a simpler method to perform this asymptotic study, by analyzing directly the sum over form factors as suggested by the result.

References

- [1] N. Kitanine, J. M. Maillet and V. Terras, *Nucl. Phys. B* **554**, 647 (1999).
- [2] N. Kitanine, J. M. Maillet and V. Terras, *Nucl. Phys. B* **567**, 554 (2000).
- [3] N. Kitanine, J. M. Maillet, N. A. Slavnov and V. Terras, *Nucl. Phys. B* **712**, 600 (2005).
- [4] N. Kitanine, K. K. Kozłowski, J. M. Maillet, N. A. Slavnov and V. Terras, *Comm. Math. Phys.* **291**, 691 (2009).
- [5] N. Kitanine, K. K. Kozłowski, J. M. Maillet, N. A. Slavnov and V. Terras, *J. Stat. Mech.-Theory Exp.* **P04003** (2009).
- [6] N. Kitanine, K. K. Kozłowski, J. M. Maillet, N. A. Slavnov and V. Terras, *J. Math. Phys.* **50**, 095209 (2009).
- [7] V. E. Korepin, N. M. Bogoliubov and A. G. Izergin, *Quantum inverse scattering method and correlation functions* (Cambridge University Press, 1993).
- [8] M. Jimbo and T. Miwa, *Algebraic analysis of solvable lattice models* (AMS, 1995).
- [9] H. Boos, M. Jimbo, T. Miwa, F. Smirnov and Y. Takeyama, *Comm. Math. Phys.* **261**, 245 (2006).
- [10] F. Göhmann, A. Klümper and A. Seel, *J. Phys. A* **37**, 7625 (2004).
- [11] A. Luther and I. Peschel, *Phys. Rev. B* **12**, 3908 (1975).
- [12] F. D. M. Haldane, *Phys. Lett. A* **81**, 153 (1981).
- [13] J. L. Cardy, *J. Phys. A: Math. Gen.* **17**, L385 (1984).
- [14] I. Affleck, *Phys. Rev. Lett.* **55**, 1355 (1985).
- [15] L. D. Faddeev, E. K. Sklyanin and L. A. Takhtajan, *Theor. Math. Phys.* **40**, 688 (1980), Translated from *Teor. Mat. Fiz.* **40**, 1979.
- [16] J. M. Maillet and V. Terras, *Nucl. Phys. B* **575**, 627 (2000).
- [17] N. A. Slavnov, *Theor. Math. Phys.* **79**, 502 (1989).
- [18] P. A. Deift, A. R. Its and X. Zhou, *Ann. Math.* **146**, 149 (1997).
- [19] S. Lukyanov, *Phys. Rev. B* **59**, 11163 (1999).
- [20] S. Lukyanov and V. Terras, *Nucl. Phys. B* **654**, 323 (2003).

OTHER TALKS

INVITED TALKS

SURPRISING SYMMETRIES OF SCATTERING AMPLITUDES IN GAUGE THEORIES

GREGORY KORCHEMSKY

Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette

E-mail: Gregory.Korchemsky@th.u-psud.fr

I will review a recent progress in computing scattering amplitudes in strongly coupled gauge theories - a fascinating subject which has been recently boosted by the formulation of the gauge/string duality in maximally supersymmetric Yang-Mills theory which is considered nowadays as “the simplest gauge theory” and/or harmonic oscillator of 21st century. In addition to the conventional symmetry of the underlying Lagrangian, the scattering amplitudes in this theory exhibit a new, dual superconformal symmetry. This symmetry is powerful enough to completely determine the scattering amplitudes for arbitrary coupling in a suitably defined limit.

FREE PROBABILITY MEETS SUPERSYMMETRY

STEPHAN MANDT, MARTIN R. ZIRNBAUER

Institut für Theoretische Physik, Universität Köln

E-mail: zirn@thp.uni-koeln.de

The spectral correlation functions of $N \times N$ random matrices (with probability measure μ_N) are expected, under certain conditions, to be universal in the large- N limit. While substantial progress with this conjecture has been made for special situations, the problem remains open for the general case of matrix ensembles which are both non-Gaussian and non-invariant. This talk starts from the known fact that the log-Fourier transform of μ_N relates to the Voiculescu R -transform (a key object in free probability theory) when the Fourier variable is of rank one. Interpreting and extending this relation from the perspective of the method of commuting and commuting variables (known as the supersymmetry method for short) we suggest a new approach to the universality conjecture for non-invariant non-Gaussian ensembles.

CONTRIBUTED TALKS

STUDYING THE BACKREACTION OF QUANTUM SCALAR FIELDS IN A COSMOLOGICAL SCENARIO

CLAUDIO DAPPIAGI

*II. Institut für Theoretische Physik, Universität Hamburg
E-mail: claudio.dappiaggi@desy.de*

We discuss the backreaction properties of a quantum massive scalar field in a Friedmann-Robertson-Walker spacetime with flat spatial section. Particularly we shall prove that, in the framework of the algebraic description of quantum field theory over curved backgrounds, it is possible to show the existence of late time stable de Sitter like solutions of the semiclassical Einstein's equations. The overall construction depends on the possibility to employ, as building blocks, Hadamard states whose existence is proved in a large class of cosmological spacetimes by means of a bulk-to-boundary reconstruction procedure. The talk is based on the following papers:

- [1] C.D., K. Fredenhagen and N. Pinamonti, Phys. Rev. D 77 (2008) 104015,
- [2] C.D., V. Moretti and N. Pinamonti, Commun. Math. Phys. 285 (2009) 1129,
- [3] C.D., V. Moretti and N. Pinamonti, [arXiv:0712.1770 [gr-qc]].

ON FUNCTIONAL DETERMINANTS IN ADS/CFT: A HOLOGRAPHIC FORMULA

DANILO E. DÍAZ

*Universidad Andres Bello, Chile
E-mail: ddiaz@physik.hu-berlin.de*

A 'holographic formula' expressing the functional determinant of the scattering operator in an asymptotically locally anti-de Sitter (ALAdS) space has been proposed in terms of a relative functional determinant of the scalar Laplacian in the bulk. It stems from considerations in AdS/CFT correspondence of a quantum correction to the partition function in the bulk and the corresponding subleading correction at large N on the boundary. As a first application, we extract the type-A conformal anomaly of GJMS operators (conformal powers of the Laplacian). Secondly, we probe the prediction for a class of quotients of hyperbolic space by a discrete subgroup of isometries, which include thermal AdS and BTZ instantons. In all these cases, the functional determinants can be expressed in terms of an associated Selberg's zeta function. In the Lorentzian version, we elucidate the connection with quasi-normal modes.

ON SINGULAR REDUCTION, QUANTIZATION AND THE OBSERVABLE ALGEBRA FOR GAUGE MODELS

GERD RUDOLPH, MATTHIAS SCHMIDT

*Institute for Theoretical Physics, University of Leipzig
E-mail: rudolph@rz.uni-Leipzig.de*

The models under consideration are obtained from lattice approximation of nonabelian gauge theories. First, we formulate them on the classical level as Hamiltonian systems, endowed with a (gauge) symmetry and with a natural momentum map related to the Gauss law. We discuss singular Marsden-Weinstein reduction, which yields the stratified reduced phase space, for two examples [1].

Next, we shortly recall previous results [2] on the structure of the field and observable algebras of these models. The building blocks of the bosonic part of the field algebra are crossed product C^* -algebras yielding a unique (generalized) Schrödinger representation, whereas the matter field component is an ordinary CAR-algebra. The observable algebra is the algebra of gauge invariant operators satisfying the Gauss law. Its irreducible representations are labeled by global colour charge.

Then, we show how to implement the stratified structure on quantum level [3], using the generalized Bargmann-Segal transform for compact Lie groups developed by Hall and the concept of a costratified Hilbert space proposed by Huebschmann. We discuss a simple, exactly solvable example and show, in particular, that the vacuum of the theory concentrates on one of the singular strata.

Finally, we prove for the same toy model [4] that the observable algebra is a C^* -algebra generated by unbounded elements (obtained from geometric quantization of classical invariants) in the sense of Woronowicz.

- [1] E.Fischer, G.Rudolph and M.Schmidt, J. Geom. Phys. 57 (2007) 1193–1213
- [2] J.Kijowski, G.Rudolph, J. Math. Phys. 46, 032303 (2005)
- [3] J.Hübschmann, G.Rudolph and M.Schmidt, Commun.Math.Phys. 286, 459-494 (2009)
- [4] G.Rudolph, M.Schmidt, J. Math. Phys. 50, 052102 (2009)

RADIATIVE CORRECTIONS TO THE MASSES OF COMPOUND PARTICLES IN THE ISING FIELD THEORY

SERGEI B. RUTKEVICH

Institute of Solid State and Semiconductor Physics, Minsk

E-mail: rut@ifttp.bas-net.by

The mass spectrum $M_n(h)$ of elementary excitations in the two-dimensional ferromagnetic Ising field theory in a weak external magnetic field h is studied. In the leading order in h , these excitations can be viewed as bound states of two confined fermions, which attract one another with a long-range linear potential. Multi-fermion fluctuations with four, six, ... fermions also contribute to the elementary excitation wave function leading to the corrections to the bound-state masses $M_n(h)$ in the higher orders in h . We calculate such multi-fermion contributions to the masses $M_n(h)$, which arise from the regular radiative correction to the kernel of the Bethe-Salpeter equation in the third order in the magnetic field h .

POSTERS

IVAN AVRAMIDI (New Mexico Institute of Mining and Technology): **HEAT KERNEL ON HOMOGENEOUS BUNDLES OVER SYMMETRIC SPACES**

BINDU ANUBHA BAMBAH (School of Physics, University of Hyderabad), with K.V.S. SHIV CHAITANYA: **PARAMETRIC AMPLIFICATION OF QUANTUM FLUCTUATIONS AND THE BARYON ASYMMETRY OF THE UNIVERSE**

HENNING BOSTELMANN (University of York, Department of Mathematics), with CHRISTOPHER J. FEWSTER: **QUANTUM INEQUALITIES FROM OPERATOR PRODUCT EXPANSIONS**

HERMANN FREIRE (Federal University of Goias), with EBERTH CORREA and ALVARO FERRAZ: **BREAKDOWN OF THE FERMI-LIQUID REGIME IN THE 2D HUBBARD MODEL FROM A TWO-LOOP FIELD-THEORETICAL RENORMALIZATION GROUP APPROACH**

OLGA YU. KHETSELIUS (Department of Applied Mathematics, Odessa State University): **QUANTIZATION OF STATES OF THE DIRAC EQUATION WITH NON-SINGULAR POTENTIAL AND HYPERFINE STRUCTURE OF SPECTRUM AND PARITY NON-CONSERVING EFFECT**

OLGA YU. KHETSELIUS (Department of Applied Mathematics, Odessa State University): **GREEN'S FUNCTION OF THE DIRAC EQUATION WITH COMPLEX ENERGY AND NON-SINGULAR CENTRAL POTENTIAL**

SHIGEAKI NAGAMACHI (Department of Mathematics, University of Tokushima), with ERWIN BRÜNING: **EDGE OF THE WEDGE THEOREM FOR ULTRA-HYPERFUNCTIONS AND ITS APPLICATION TO QUANTUM FIELD THEORY WITH A FUNDAMENTAL LENGTH**

ZALIHE ÖZÇAMAKLI (Anadolu University), with CEM YÜCE: **ROTATING CAVITY IN DYNAMICAL CASIMIR EFFECT**

JULIEN QUEVA (Université Diderot–Paris 7), with S. FACI, E. HUGUET, and J. RENAUD: **A CONFORMALLY COVARIANT QUANTIZATION OF THE MAXWELL FIELD IN DE SITTER SPACETIME**

AKIFUMI SAKO (Kushiro National College of Technology): **NONCOMMUTATIVE DEFORMATION OF INSTANTONS, INSTANTON NUMBERS AND ADHM CONSTRUCTION**

ALEXANDER J. SILENKO (Belarusian State University): **FOLDY-WOUTHYUSEN TRANSFORMATION AND SEMICLASSICAL LIMIT FOR RELATIVISTIC QUANTUM MECHANICS**

GRIGORY VARTANOV (Joint Institute for Nuclear Research, Dubna), with L. BORK, D. KAZAKOV, and A. ZHIBOEDOV: **INFRARED-FINITE OBSERVABLES IN $N = 4$ SYM THEORY**

NELSON YOKOMIZO (Universidade de São Paulo), with JOÃO C.A. BARATA: **MULTIPLE CLASSICAL LIMITS IN RELATIVISTIC QUANTUM MECHANICS**

YURI M. ZINOVIEV (Steklov Mathematical Institute, Moscow): **VACUUM EXPECTATION VALUES OF QUANTUM FIELDS**

Many-Body Quantum Mechanics

(condensed matter, nanoscale, elasticity)

Session organizers:

Gian Michele Graf
Jakob Yngvason

STRENGTH OF SUPERCONDUCTIVITY CLOSE TO CRITICAL MAGNETIC FIELD

SØREN FOURNAIS* and AYMAN KACHMAR

*Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, Building 1530,
DK-8000 Århus C, Denmark*

**E-mail: fournais@imf.au.dk*

We present recent results on superconductivity in the Ginzburg-Landau model when the strength of the external magnetic field is close to the second critical field. We determine for what field strength the bulk component of the energy contributes to leading order. Also we study the magnitude of the order parameter away from the boundary.

1. Introduction

The mathematics of the (stationary) Ginzburg-Landau model of superconductivity has seen much progress over the last decade. The methods of analysis and the results obtained vary strongly depending on the magnitude of the parameter measuring the strength of an external magnetic field. When the external magnetic field is not too strong, the minimizers of the model exhibit vortices—an overview of the analysis in this case can be found in [15]. When the magnetic field is much stronger, minimizers are essentially localized to a narrow region around the boundary, and techniques from semi-classical analysis turn out to be efficient in the analysis of the model. A presentation of this situation can be found in [9]. For an introduction to the physics of superconductors, see the classical texts [6, 14].

The intermediate zone is much less understood. One considers the loosely defined critical field denoted by H_{C_2} as being the magnetic field strength at which superconductivity becomes weak in the bulk region. The results described in [15] concern the region (far) below H_{C_2} and the semiclassical results the region (far) above H_{C_2} . Near, but below, H_{C_2} vortices are supposed to organize in regular lattices [1] but this remains a challenging question (see [2, 3] for indications in this direction). In this text we will describe some results from [11] concerning the transition that takes place at H_{C_2} .

1.1. The model

We study the standard Ginzburg-Landau functional with magnetic field. For a smooth domain $\Omega \subset \mathbb{R}^2$ (the cross section of a superconducting wire), one defines

$$\begin{aligned} \mathcal{E}_{\kappa,H}(\psi, \mathbf{A}; \Omega) &= \int_{\Omega} e_{\kappa,H}(\psi, \mathbf{A}) \, dx \\ &= \int_{\Omega} \left(|(\nabla - i\kappa H \mathbf{A})\psi|^2 - \kappa^2 |\psi|^2 + \frac{\kappa^2}{2} |\psi|^4 + (\kappa H)^2 |\operatorname{curl}(\mathbf{A} - \mathbf{F})|^2 \right) dx. \end{aligned} \quad (1)$$

Here $\psi \in W^{1,2}(\Omega)$ is a complex valued wave function (called order parameter in this context), $\mathbf{A} \in W^{1,2}(\Omega, \mathbb{R}^2)$ a vector potential (the vector potential of the induced magnetic field), κ the Ginzburg-Landau parameter (a material parameter which is temperature independent), and H measures the strength of the applied magnetic field. The potential $\mathbf{F} : \Omega \rightarrow \mathbb{R}^2$ is the unique vector field satisfying,

$$\operatorname{curl} \mathbf{F} = 1, \quad \operatorname{div} \mathbf{F} = 0 \quad \text{in } \Omega, \quad \nu \cdot \mathbf{F} = 0 \quad \text{on } \partial\Omega, \quad (2)$$

where ν is the unit inward normal vector of $\partial\Omega$.

The quantity $|\psi(x)|$ is a measure of the superconducting state of the material near the point x . If $|\psi(x)| \approx 1$, then the material is superconducting near x . If $|\psi(x)| \approx 0$ then superconductivity is weak near the given point.

The equilibrium state of the material is given by a minimizer of the functional $\mathcal{E}$. This, and more general stationary points, will solve the associated Euler-Lagrange equations, the Ginzburg-Landau equations,

$$\begin{cases} -(\nabla - i\kappa H \mathbf{A})^2 \psi = \kappa^2 (1 - |\psi|^2) \psi, \\ -\nabla^\perp \operatorname{curl} \mathbf{A} = (\kappa H)^{-1} \operatorname{Im}(\overline{\psi} (\nabla - i\kappa H \mathbf{A}) \psi), & \text{in } \Omega, \\ \nu \cdot (\nabla - i\kappa H \mathbf{A}) \psi = 0, \quad \operatorname{curl} \mathbf{A} = 1, & \text{on } \partial\Omega. \end{cases} \quad (3)$$

We will only consider the situation where $\kappa \gg 1$, corresponding to superconductors of Type II. With these conventions

- (1) The first vortex appears at $H = H_{C_1} \approx \frac{\log \kappa}{2\kappa}$ (as $\kappa \rightarrow \infty$).
- (2) Superconductivity becomes weak in the interior of Ω at $H = H_{C_2}$. This takes place for magnetic field strength $H \approx \kappa$. This is the situation we will describe in the present text.
- (3) The functional ceases to have non-trivial stationary points at $H = H_{C_3} \approx \Theta_0^{-1} \kappa$ (as $\kappa \rightarrow \infty$), where $\Theta_0 \approx 0.59$ is a universal constant.

We have not given a precise definition of the critical field H_{C_2} , however the results below will indicate that a change from 'bulk' to 'surface' takes place when the magnetic field strength is varied over a narrow interval around $H = \kappa$.

2. Results

2.1. Above H_{C_2} : Decay in the normal variable

When $H > \kappa$ the superconducting order parameter ψ is expected to decay with the distance to the boundary. Notice though, that ψ is expected to be of order unity in a boundary layer of thickness κ^{-1} . The strongest result in this direction is the following result due to Almog [4].

Theorem 2.1. *There exist constants $\lambda_0, \beta, \kappa_0, C > 0$, such that if $\kappa \geq \kappa_0$, and*

$$H \geq \kappa + \lambda_0 \frac{\log \kappa}{\kappa}, \quad (4)$$

then for all $x \in \Omega$ and all solutions $(\psi, \mathbf{A})$ to (3),

$$|\psi(x)| \leq C \exp \left(-\beta \text{dist}(x, \partial\Omega) \sqrt{\kappa(H - \kappa)} \right). \quad (5)$$

2.2. Uniform decay in the bulk

The following uniform bound gives the magnitude of $|\psi(x)|$ for $\text{dist}(x, \partial\Omega) \gg \kappa^{-1}$. Theorem 2.2 is a sharpened version of a result of [7]. Together, these results give an affirmative answer to a conjecture of [2] and [16].

Theorem 2.2 (cf. [11]).

Let $\delta \in (0, 1)$ and $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a function such that $g(\kappa) = o(\kappa)$ as $\kappa \rightarrow \infty$. Then there exists a constant $C > 0$ such that if $H = \kappa + \nu$, where $|\nu| \leq g(\kappa)$, then

$$\|\psi\|_{L^\infty(\omega_\kappa)} \leq C\lambda(\kappa), \quad (6)$$

for all critical points $(\psi, \mathbf{A})$ of (1). Here

$$\omega_\kappa := \{x \in \Omega \mid \text{dist}(x, \partial\Omega) \geq \kappa^{-1+\delta}\}, \quad (7)$$

and

$$\lambda(\kappa) := \max \left\{ \left| \frac{\kappa}{H} - 1 \right|^{1/2}, \kappa^{-1+\delta} \right\}. \quad (8)$$

Furthermore, by the following theorem, the L^∞ bound of Theorem 2.2 is effective.

Theorem 2.3 (cf. [11]).

There exists a universal constant E_2 such that if $H(\kappa)$ satisfies

$$\frac{H(\kappa)}{\kappa} \rightarrow 1, \quad \text{as } \kappa \rightarrow \infty, \quad (9)$$

Then, for all minimizers $(\psi, \mathbf{A})$ of $\mathcal{E}_{\kappa, H}$, and all open smooth domains D with $\overline{D} \subset \Omega$, we have

$$\int_D |\psi|^4 dx = 2E_2 |D| \left| \frac{\kappa}{H} - 1 \right|_+^2 + o \left(\max(\kappa^{-1}, \left| \frac{\kappa}{H} - 1 \right|_+^2) \right) \quad \text{as } \kappa \rightarrow \infty. \quad (10)$$

2.3. Energy separation

Our final result gives a description of H_{C_2} in terms of energy. It is when $H = \kappa + o(\sqrt{\kappa})$ that the bulk ceases to contribute to the leading order energy of minimizers.

Theorem 2.4 (cf. [11]).

There exist universal constants E_1 and E_2 such that the following is true. If the magnetic field satisfies,

$$H = \kappa - \mu(\kappa)\sqrt{\kappa} \quad \text{with} \quad \lim_{\kappa \rightarrow \infty} \mu(\kappa) = \mu_0, \quad \mu_0 \in \mathbb{R}.$$

Then for any minimizer $(\psi, \mathbf{A})$ of (1) and any open smooth domain $D \subset \Omega$, the following asymptotic expansion holds

$$\mathcal{E}(\psi, \mathbf{A}; D \cap \Omega) = -(E_1 |\overline{D} \cap \partial\Omega| + [\mu_0]_+^2 E_2 |D \cap \Omega|) \kappa + o(\kappa), \quad \text{as } \kappa \rightarrow \infty. \quad (11)$$

In particular, defining $\mu_\kappa = -\frac{e_{\kappa, H}(\psi, \mathbf{A}) dx}{\kappa}$, then

$$\mu_\kappa \rightharpoonup E_1 \mathcal{H}_{\partial\Omega}^1 + [\mu_0]_+^2 E_2 dx$$

where $\mathcal{H}_{\partial\Omega}^1$ is the 1-dimensional Hausdorff measure restricted on $\partial\Omega$ and dx is the Lebesgue measure in Ω .

References

- [1] A.A. Abrikosov. On the magnetic properties of superconductors of the second group. *J. Exp. Theor. Phys.* 5 (1957), p. 1174-1182.
- [2] A. Aftalion, S. Serfaty. Lowest Landau level approach in superconductivity for the Abrikosov lattice close to H_{C_2} . *Selecta Math. (NS)*. **13**, no.2 (2007), 183–202.
- [3] Y. Almog. Abrikosov lattices in finite domains. *Commun. Math. Phys.* **262** (2006), 677-702.
- [4] Y. Almog. Non-linear surface superconductivity in the large κ limit. *Rev. Math. Phys.* **16** (2004), 961-976.
- [5] Y. Almog, B. Helffer. The distribution of surface superconductivity along the boundary: on a conjecture of X. B. Pan. *SIAM J. Math. Anal.* **38** (2007), no. 6, 1715–1732.
- [6] P.G. de Gennes. *Superconductivity of metals and alloys*. Benjamin (1966).
- [7] S. Fournais, B. Helffer. Bulk superconductivity in type II superconductors near the second critical field. To appear in *J. Europ. Math. Soc. (JEMS)* (2008).
- [8] S. Fournais, B. Helffer. Energy asymptotics for type II superconductors. *Calc. Var. Partial Differential Equations* **24** (2005), no. 3, 341–376.
- [9] S. Fournais, B. Helffer. *Spectral Methods in surface superconductivity*. Monograph submitted.
- [10] S. Fournais, B. Helffer. Optimal uniform elliptic estimates for the Ginzburg-Landau system. *Adventures in Mathematical Physics. Contemp. Math.* **447** (2007), 83-102.
- [11] S. Fournais, A. Kachmar. Nucleation of bulk superconductivity close to critical magnetic field. In preparation.
- [12] K. Lu, X.B. Pan. Estimates of the upper critical field for the Ginzburg-Landau equations of superconductivity. *Physica D* **127** (1999), 73-104.
- [13] X.B. Pan. Surface superconductivity in applied magnetic fields above H_{C_2} . *Commun. Math. Phys.* **228** (2002), 228-370.
- [14] D. Saint-James, G. Sarma, E.J. Thomas. *Type II Superconductivity*. Pergamon, Oxford 1969.

- [15] E. Sandier, S. Serfaty. *Vortices for the Magnetic Ginzburg-Landau Model*. Birkhuser. (2008).
- [16] E. Sandier, S. Serfaty. The decrease of bulk superconductivity close to the second critical field in the Ginzburg-Landau model. *SIAM. J. Math. Anal.* **34** No. 4 (2003), 939-956.

THE 2D HUBBARD MODEL ON THE HONEYCOMB LATTICE

ALESSANDRO GIULIANI

Dipartimento di Matematica

Università di Roma Tre

L.go S. L. Murialdo 1

00146 Roma - Italy

E-mail: giuliani@mat.uniroma3.it

We consider the 2D Hubbard model on the honeycomb lattice, as a model for a single layer graphene sheet in the presence of screened Coulomb interactions. At half filling and weak enough coupling, we compute the free energy, the ground state energy and we construct the correlation functions up to zero temperature in terms of convergent series; analyticity is proved by making use of constructive fermionic renormalization group methods. We show that the interaction produces a modification of the Fermi velocity and of the wave function renormalization without changing the asymptotic infrared properties of the model with respect to the unperturbed non-interacting case; this rules out the possibility of superconducting or magnetic instabilities in the ground state.

Keywords: Hubbard model; honeycomb lattice; graphene; renormalization group; ground state correlations

1. Introduction

The recent experimental realization of a monocrystalline graphitic film, known as *graphene* [1], revived the interest in the low temperature physics of two-dimensional electron systems on the honeycomb lattice, which is the typical underlying structure displayed by single-layer graphene sheets.

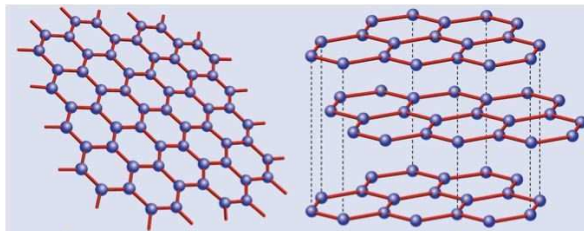


Fig. 1. A sketch of the crystal structure of graphene and of graphite. Graphene (left) is a honeycomb lattice of carbon atoms. Graphite can be viewed as a stack of graphene layers.

Graphene is quite different from most conventional quasi-two dimensional electron gases, because of the peculiar quasi-particles dispersion relation, which closely resembles the one of massless Dirac fermions in $2 + 1$ dimensions. This was already pointed out in [2] and further exploited in [3], where the analogy between graphene and $2 + 1$ -dimensional quantum electrodynamics (QED) was made explicit, and used to predict a condensed-matter analogue of the axial anomaly in QED. From this point of view, graphene can be considered as a sort of testing bench to investigate the properties of infrared QED in $2 + 1$ dimensions. Recently, the experimental observation of graphene greatly enhanced the study of the anomalous effects induced by the pseudo-relativistic dispersion relation of its quasi particles, see [4] for an up-to-date description of the state of art.

The main difference with respect to usual 2D Fermi systems lies in the geometry of the Fermi surface, which at half filling is not given by a curve, as in usual 2D Fermi systems, but is completely degenerate: it consists of *two isolated points*, as in one dimensional Fermi systems. From a theoretical point of view, this fact completely changes the infrared scaling properties of the propagator: in particular, in the case of short-range electron-electron interactions, it makes all the operators with four or more fermionic fields irrelevant in a Renormalization Group (RG) sense (contrary to the case of the usual 2D Fermi gas, where local quartic operators are marginal, i.e., they give rise to logarithmic divergences in perturbation theory).

In this talk I will present the first rigorous construction of the low temperature and ground state properties of the 2D Hubbard model on the honeycomb lattice with weak local interactions; this is achieved by rewriting the correlation functions in terms of resummed series, convergent uniformly in the temperature up to zero temperature, as we prove by making use of constructive fermionic renormalization group. These results are based on joint work with V. Mastropietro; the details of the proofs can be found in [5]. The resulting picture is that the interaction does not change the asymptotic infrared properties of the model with respect to the unperturbed non-interacting case, but it produces a renormalization of the Fermi velocity and of the wave function. Our results rule out the presence of superconducting or magnetic instabilities at weak coupling; this is in striking contrast with the behavior of the 2D Hubbard model on the square lattice, where quantum instabilities (corresponding to the magnetic or superconducting long range order that are presumably present in the ground state) prevent the convergence of the perturbative expansion in U for low enough temperatures.

2. The model

The grandcanonical Hamiltonian of the 2D Hubbard model on the honeycomb lattice at half filling in second quantized form is given by:

$$H_\Lambda = - \sum_{\vec{x} \in \Lambda} \left[\sum_{\substack{\sigma=\uparrow\downarrow \\ i=1,2,3}} \left(a_{\vec{x},\sigma}^+ b_{\vec{x}+\vec{\delta}_i,\sigma}^- + b_{\vec{x}+\vec{\delta}_i,\sigma}^+ a_{\vec{x},\sigma}^- \right) + U \left(a_{\vec{x},\uparrow}^+ a_{\vec{x},\uparrow}^- - \frac{1}{2} \right) \left(a_{\vec{x},\downarrow}^+ a_{\vec{x},\downarrow}^- - \frac{1}{2} \right) \right. \\ \left. + U \left(b_{\vec{x}+\vec{\delta}_1,\uparrow}^+ b_{\vec{x}+\vec{\delta}_1,\uparrow}^- - \frac{1}{2} \right) \left(b_{\vec{x}+\vec{\delta}_1,\downarrow}^+ b_{\vec{x}+\vec{\delta}_1,\downarrow}^- - \frac{1}{2} \right) \right] \quad (1)$$

where: (i) Λ is a periodic triangular lattice, with basis $\vec{a}_1 = \frac{1}{2}(3, \sqrt{3})$, $\vec{a}_2 = \frac{1}{2}(3, -\sqrt{3})$ and period L both in direction $\vec{a}_1$ and in direction $\vec{a}_2$; (ii) the nearest neighbor vectors $\vec{\delta}_i$ are defined as $\vec{\delta}_1 = (1, 0)$, $\vec{\delta}_2 = \frac{1}{2}(-1, \sqrt{3})$, $\vec{\delta}_3 = \frac{1}{2}(-1, -\sqrt{3})$; (iii) $a_{\vec{x},\sigma}^\pm$ and $b_{\vec{y},\sigma}^\pm$ are creation or annihilation fermionic operators with spin index $\sigma = \uparrow\downarrow$ and site index $\vec{x} \in \Lambda_A := \Lambda$ and $\vec{y} \in \Lambda_B := \Lambda + \vec{\delta}_1$, respectively, satisfying periodic boundary conditions; (iv) U is the strength of the on-site density-density interaction, which can be either positive or negative. See Fig. 2

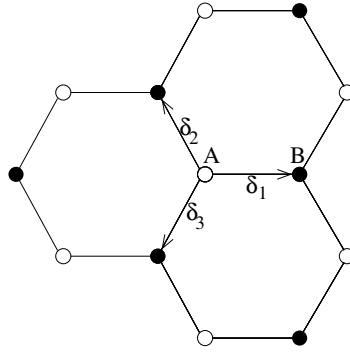


Fig. 2. The honeycomb lattice and its two triangular sublattices Λ_A and Λ_B , corresponding in the picture to the white and black sites, respectively. $a_{\vec{x},\sigma}^\pm$ and $b_{\vec{y},\sigma}^\pm$ are creation/annihilation of electrons localized at the sites of the A and B sublattices, respectively.

Note that Hamiltonian (1) is hole-particle symmetric, i.e., it is invariant under the exchange $a_{\vec{x},\sigma}^\pm \longleftrightarrow a_{\vec{x},\sigma}^\mp$, $b_{\vec{x}+\vec{\delta}_1,\sigma}^\pm \longleftrightarrow -b_{\vec{x}+\vec{\delta}_1,\sigma}^\mp$, which implies that the average particle density is exactly 1, for any $|\Lambda|$ and any β .

Our goal is to characterize the low and zero temperature properties of the system described by (1), by computing thermodynamic functions and a complete set of correlations at low or zero temperatures. To this purpose it is convenient to introduce the notions of specific free and ground state energies:

$$f_\beta(U) = -\frac{1}{\beta} \lim_{|\Lambda| \rightarrow \infty} |\Lambda|^{-1} \log \text{Tr} \{ e^{-\beta H_\Lambda} \}, \quad e(U) = \lim_{\beta \rightarrow \infty} f_\beta(U), \quad (2)$$

and the notion of Schwinger functions, defined as follows. Let us introduce the two component fermionic operators $\Psi_{\mathbf{x},\sigma}^{\pm} = e^{Hx_0} (a_{\vec{x},\sigma}^{\pm}, b_{\vec{x}+\vec{\delta}_1,\sigma}^{\pm}) \Psi_{\vec{x},\sigma}^{\pm} e^{-Hx_0}$, with $\mathbf{x} = (x_0, \vec{x})$ and $x_0 \in [0, \beta]$, for some $\beta > 0$; we shall call x_0 the time variable. We shall write $\Psi_{\mathbf{x},\sigma,1}^{\pm} = a_{\mathbf{x},\sigma}^{\pm}$ and $\Psi_{\mathbf{x},\sigma,2}^{\pm} = b_{\mathbf{x}+\delta_1,\sigma}^{\pm}$, with $\delta_1 = (0, \vec{\delta}_1)$. We define

$$S_n^{\beta,\Lambda}(\mathbf{x}_1, \varepsilon_1, \sigma_1, \rho_1; \dots; \mathbf{x}_n, \varepsilon_n, \sigma_n, \rho_n) = \langle \mathbf{T} \{ \Psi_{\mathbf{x}_1, \sigma_1, \rho_1}^{\varepsilon_1} \cdots \Psi_{\mathbf{x}_n, \sigma_n, \rho_n}^{\varepsilon_n} \} \rangle_{\beta, \Lambda} \quad (3)$$

where: $\mathbf{x}_i \in [0, \beta] \times \Lambda$, $\sigma_i = \uparrow \downarrow$, $\varepsilon_i = \pm$, $\rho_i = 1, 2$ and $\mathbf{T}$ is the operator of fermionic time ordering, acting on a product of fermionic fields as:

$$\mathbf{T}(\Psi_{\mathbf{x}_1, \sigma_1, \rho_1}^{\varepsilon_1} \cdots \Psi_{\mathbf{x}_n, \sigma_n, \rho_n}^{\varepsilon_n}) = (-1)^{\pi} \Psi_{\mathbf{x}_{\pi(1)}, \sigma_{\pi(1)}, \rho_{\pi(1)}}^{\varepsilon_{\pi(1)}} \cdots \Psi_{\mathbf{x}_{\pi(n)}, \sigma_{\pi(n)}, \rho_{\pi(n)}}^{\varepsilon_{\pi(n)}} \quad (4)$$

where π is a permutation of $\{1, \dots, n\}$, chosen in such a way that $x_{\pi(1)0} \geq \dots \geq x_{\pi(n)0}$, and $(-1)^{\pi}$ is its sign. At coinciding times, $\mathbf{T}$ is defined as the normal ordering operator. Taking the limit $\Lambda \rightarrow \infty$ in (3) we get the finite temperature n -point Schwinger functions, denoted by $S_n^{\beta}(\mathbf{x}_1, \varepsilon_1, \sigma_1, \rho_1; \dots; \mathbf{x}_n, \varepsilon_n, \sigma_n, \rho_n)$, which describe the properties of the infinite volume system at finite temperature. Taking the $\beta \rightarrow \infty$ limit of the finite temperature Schwinger functions, we get the zero temperature Schwinger functions, simply denoted by $S_n(\mathbf{x}_1, \varepsilon_1, \sigma_1, \rho_1; \dots; \mathbf{x}_n, \varepsilon_n, \sigma_n, \rho_n)$, which define the properties of the thermal ground state of (1) in the thermodynamic limit.

2.1. The non interacting case

In the non-interacting case $U = 0$ the Schwinger functions of any order n can be exactly computed as linear combinations of products of two-point Schwinger functions, via the well-known *Wick rule*. The two-point Schwinger function itself, also called the *free propagator*, for $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{x} - \mathbf{y} \neq (\pm\beta, \vec{0})$, is equal to

$$S_2^{\beta,\Lambda}(\mathbf{x}, \sigma, -, \rho; \mathbf{y}, \sigma, +, \rho') \Big|_{U=0} = \frac{1}{\beta|\Lambda|} \sum_{\mathbf{k} \in \mathcal{D}_{\beta,L}} \frac{e^{-i\mathbf{k} \cdot (\mathbf{x}-\mathbf{y})}}{k_0^2 + |v(\vec{k})|^2} \begin{pmatrix} ik_0 & -v^*(\vec{k}) \\ -v(\vec{k}) & ik_0 \end{pmatrix}_{\rho, \rho'} \quad (5)$$

where: (i) $\mathbf{k} = (k_0, \vec{k})$ and $\mathcal{D}_{\beta,L} = \mathcal{D}_{\beta} \times \mathcal{D}_L$, with $\mathcal{D}_{\beta} = \{k_0 = \frac{2\pi}{\beta}(n_0 + \frac{1}{2}) : n_0 \in \mathbb{Z}\}$ and $\mathcal{D}_L = \{\vec{k} = \frac{n_1}{L}\vec{b}_1 + \frac{n_2}{L}\vec{b}_2 : 0 \leq n_1, n_2 \leq L-1\}$ (here $\vec{b}_i = \frac{2\pi}{3}(1, (-1)^{i-1}\sqrt{3})$, $i = 1, 2$, form a basis of the dual lattice Λ^*); (iii) $v(\vec{k}) = \sum_{i=1}^3 e^{i\vec{k}(\vec{\delta}_i - \vec{\delta}_1)} = 1 + 2e^{-i3/2k_1} \cos \frac{\sqrt{3}}{2}k_2$.

If we take $\beta, L \rightarrow \infty$, the limiting propagator becomes singular at $\{k_0 = 0\} \times \{\vec{k} = \vec{p}_F^{\pm}\}$, where $\vec{p}_F^{\pm} := (\frac{2\pi}{3}, \pm \frac{2\pi}{3\sqrt{3}})$ are the *Fermi points* or *Dirac points*. The asymptotic behavior of $v(\vec{k})$ close to the Fermi points is given by $v(\vec{p}_F^{\pm} + \vec{k}') \simeq \frac{3}{2}(ik'_1 \pm k'_2)$. In particular, if $\omega = \pm$, the Fourier transform of the 2-point Schwinger function close to the singularity can be rewritten in the form:

$$\hat{S}_2^{(U=0)}(k_0, \vec{p}_F^{\omega} + \vec{k}') = \begin{pmatrix} -ik_0 & -\frac{3}{2}(-ik'_1 + \omega k'_2) + r_{\omega}(\vec{k}') \\ -\frac{3}{2}(ik'_1 + \omega k'_2) + r_{\omega}^*(\vec{k}') & -ik_0 \end{pmatrix}^{-1}, \quad (6)$$

where $|r_{\omega}(\vec{k}')| \leq C|\vec{k}'|^2$, for small values of $\vec{k}'$ and for some positive constant C .

2.2. The interacting case

We are now interested in what happens by adding a local interaction. In the case $U \neq 0$, the Schwinger functions are not exactly computable anymore. It is well-known that they can be written as formal power series in U , constructed in terms of *Feynmann diagrams*, using as free propagator the function $S_2^{\beta, \Lambda}|_{U=0}$ in (5). Our main result consists in a proof of convergence of this formal expansion for U small enough, after the implementation of suitable resummations of the original power series.

Theorem 2.1 (Giuliani and Mastropietro, [5]). *Let us consider the 2D Hubbard model on the honeycomb lattice at half filling, defined by (1). There exist a constant $U_0 > 0$ such that, if $|U| \leq U_0$, the specific free energy $f_\beta(U)$ and the finite temperature Schwinger functions are analytic functions of U , uniformly in β as $\beta \rightarrow \infty$, and so are the specific ground state energy $e(U)$ and the zero temperature Schwinger functions. The Fourier transform of the zero temperature two point Schwinger function $S(\mathbf{x})_{\rho, \rho'} := S_2(\mathbf{x}, \sigma, -, \rho; \mathbf{0}, \sigma, +, \rho')$, denoted by $\hat{S}(\mathbf{k})$, is singular only at the Fermi points $\mathbf{k} = \mathbf{p}_F^\pm = (0, \vec{p}_F^\pm)$ and, close to the singularities, if $\omega = \pm$, it can be written as*

$$\hat{S}(k_0, \vec{p}_F^\omega + \vec{k}') = \frac{1}{Z} \begin{pmatrix} -ik_0 & -v_F(-ik_1' + \omega k_2') \\ -v_F(ik_1' + \omega k_2') & -ik_0 \end{pmatrix}^{-1} \left(1 + R(k_0, \vec{k}')\right), \quad (7)$$

with $Z = 1 + O(U^2)$ and $v_F = \frac{3}{2} + O(U^2)$ two analytic functions of U . Moreover the matrix $R(\mathbf{k}')$ satisfies $\|R(\mathbf{k}')\| \leq C|\mathbf{k}'|^\vartheta$ for some constants $C, \vartheta > 0$ and for $|\mathbf{k}'|$ small enough.

Theorem 2.1 says that the location of the singularity does not change in the presence of interaction; however, the wave function renormalization Z and Fermi velocity v_F are modified by the interaction. The resulting theory is not quasi-free: the Wick rule is not valid anymore in the presence of interactions. However, the long distance asymptotics of the n -points Schwinger functions is the same suggested by the Wick rule. The fact that the interacting correlations decay as in the non-interacting case implies in particular the absence of long range order at zero temperature, e.g., the absence of Néel or superconducting order in the thermal ground state at weak coupling (in fact, spin-spin correlations and Cooper pairs correlations decay to zero as (distance) $^{-4}$, which is integrable in two dimensions).

References

- [1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov, *Science* **306**, 666 (2004).
- [2] P. R. Wallace, *Phys. Rev.* **71**, 622–634 (1947).
- [3] G. W. Semenoff, *Phys. Rev. Lett.* **53**, 2449–2452 (1984).
- [4] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, *Rev. Mod. Phys.* **81**, 109–162 (2009).
- [5] A. Giuliani and V. Mastropietro, *Comm. Math. Phys.*, **293**, 301–346 (2010).

QUANTUM PHASES OF SUPERSYMMETRIC LATTICE MODELS

LIZA HUIJSE* and KARELJAN SCHOUTENS

*Institute for Theoretical Physics, University of Amsterdam,
Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands*

**E-mail: L.Huijse@uva.nl*

We review recent results on lattice models for spin-less fermions with strong repulsive interactions. A judicious tuning of kinetic and interaction terms leads to a model possessing supersymmetry. In the 1D case, this model displays critical behavior described by superconformal field theory. On 2D lattices we generically find superfrustration, characterized by an extensive ground state entropy. For certain 2D lattices analytical results on the ground state structure reveal yet another quantum phase, which we tentatively call 'supertopological'.

Keywords: Strongly correlated lattice fermions, supersymmetry, topological phases

1. Introduction

A long standing challenge in condensed matter physics is to understand the properties of strongly correlated electron systems. While it is relatively easy to formulate model descriptions, it has proved exceedingly difficult to arrive at exact results for these in spatial dimensions $D > 1$, in particular in regimes where interaction and kinetic terms in the relevant hamiltonian are of comparable strength. In a series of papers, initiated by one of the authors together with P. Fendley and J. de Boer [1], a specific model for lattice fermions was explored, precisely in this non-perturbative regime. The key property of this model is **supersymmetry**. This gives a subtle balance between kinetic and interaction terms, leading to remarkable features such as, in particular, large degeneracies of quantum ground states. At the same time, supersymmetry provides a rich mathematical structure that can be employed to derive rigorous results for some of the key features of the model.

The degrees of freedom are spin-less fermionic particles living on a given lattice. A fermion at site i is created by the operator $c_i^\dagger$ with $\{c_i, c_j^\dagger\} = \delta_{ij}$. The fermions have a *hard core*, meaning that the presence of a fermion on a given site excludes the occupation of all adjacent sites. With this, the fermion creation operator becomes $d_i^\dagger = c_i^\dagger \mathcal{P}_{< i >}$, where

$$\mathcal{P}_{< i >} = \prod_{j \text{ next to } i} (1 - c_j^\dagger c_j) \quad (1)$$

is zero if any site next to i is occupied. The hamiltonian is defined in terms of the

supercharge $Q = \sum_i d_i^\dagger$:

$$H = \{Q, Q^\dagger\} = \sum_{\langle ij \rangle} d_i^\dagger d_j + \sum_i \mathcal{P}_{\langle i \rangle}. \quad (2)$$

The model is supersymmetric because $Q^2 = (Q^\dagger)^2 = 0$, which then implies that $[Q, H] = [Q^\dagger, H] = 0$. The second term in the hamiltonian combines a chemical potential and a repulsive next-nearest-neighbor potential. The details of this term depend on the lattice we choose.

A powerful analytical tool is the Witten index

$$W = \text{Tr}[(-1)^F] \quad (3)$$

with F the operator counting the number of fermions in a given state. One easily shows that $W = N_b - N_f$, where N_b is the number of bosonic ground states (those with an even number of fermions), and N_f is the number of fermionic ground states (those with an odd number of fermions) [2]. By this result, $|W|$ is a lower bound on the number of ground states. In analyzing the supersymmetric model we also used techniques from cohomology theory, relying on a one-to-one correspondence between quantum ground states and cohomology classes of the complex associated with the supercharge Q .

The papers [1, 3–7] explored this model on a variety of lattices and revealed a number of remarkable quantum phases. While some of the specifics of these phases require the fine-tuning set by supersymmetry, many qualitative features are expected to survive in models that are sufficiently close to the supersymmetric point.

On 1D lattices (quantum chains), the model turns out to be quantum critical, with the critical behavior fully described by the first ($k = 1$) unitary minimal model of $N = 2$ supersymmetric Conformal Field Theory (SCFT) [1].

On generic 2D and 3D lattices, the supersymmetric lattice model displays a phenomenon we call **superfrustration**. This term denotes a strong form of quantum charge frustration, with the number of quantum ground states growing exponentially with the volume of the system, implying extensive ground state entropy [5, 6]. For an elaborate account we refer the reader to an earlier review [8].

In this paper we focus on 2D lattices which exhibit yet another type of anomalous behavior, with the number of quantum ground states growing exponentially with the *linear* dimensions of the system (sub-extensive ground state entropy). One example is the 2D square lattice, where this property is established via a rigorous theorem relating the number of quantum ground states to specific rhombus tilings of the plane. A second example is the octagon-square lattice. In this latter case the ground state structure is less involved, allowing us to delve a bit deeper and to extend the analysis to the presence of defects. With some of the features observed (torus degeneracies, presence of edge modes) being reminiscent of those of topological phases of 2D matter, we tentatively refer to these phases as ‘supertopological’.

Table 1. Witten Index for $M \times N$ square lattice.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	-1	1	3	1	-1	1	3	1	-1	1	3	1	-1	1
3	1	1	4	1	1	4	1	1	4	1	1	4	1	1	4
4	1	3	1	7	1	3	1	7	1	3	1	7	1	3	1
5	1	1	1	1	-9	1	1	1	1	11	1	1	1	1	-9
6	1	-1	4	3	1	14	1	3	4	-1	1	18	1	-1	4
7	1	1	1	1	1	1	1	1	1	1	1	1	1	-27	1
8	1	3	1	7	1	3	1	7	1	43	1	7	1	3	1
9	1	1	4	1	1	4	1	1	40	1	1	4	1	1	4
10	1	-1	1	3	11	-1	1	43	1	9	1	3	1	69	11
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	3	4	7	1	18	1	7	4	3	1	166	1	3	4

2. Square lattice

Where numerical studies of the Witten index [5, 6] showed superfrustration for a wide range of 2D lattices, they revealed a very different behavior for the square lattice wrapped around the torus (see Table 1). At first glance one notices that the index does not grow exponentially with the system size as it does for the superfrustrated systems. Inspired by the curiosities of this table and two conjectures [9] on its structure, Jonsson [10] proved a general expression for the Witten index $W_{u,v}$ of the square lattice with periodic boundary conditions given by the vectors $\vec{u} = (u_1, u_2)$ and $\vec{v} = (v_1, v_2)$. He showed that $W_{u,v}$ is simply related to tilings constructed from the rhombuses pictured in Fig. 1. Precisely, let t_b (t_f) be the number of ways of tiling the torus with these four rhombus types, so that there are an even (odd) number of rhombuses.

Theorem 2.1 (Jonsson, 2006). *The expression for the Witten index reads*

$$W_{u,v} = N_b - N_f = t_b - t_f - (-1)^d \theta_d \theta_{d^*}, \quad (4)$$

where $d \equiv \gcd(u_1 - u_2, v_1 - v_2)$, $d^* \equiv \gcd(u_1 + u_2, v_1 + v_2)$ and

$$\theta_d \equiv \begin{cases} 2 & \text{if } d = 3k, \text{ with } k \text{ integer} \\ -1 & \text{otherwise.} \end{cases} \quad (5)$$

A natural extension of Jonsson's theorem is to relate the *total* number of ground states to rhombus tilings. Exploiting the one-to-one correspondence between ground states and elements of the cohomology of the supercharge Q , we were able to prove this relation explicitly when $\vec{u} = (m, -m)$ and $v_1 + v_2 = 3p$ [7, 11].

Theorem 2.2 (Fendley, Huijse and Schoutens, 2009). *The total number of ground states reads*

$$N_b + N_f = t_b + t_f + \Delta, \quad (6)$$

where $|\Delta| = |\theta_d \theta_{d^*}|$. For $\vec{u} = (m, -m)$ and $v_1 + v_2 = 3p$ we find $\Delta = -(-1)^{(\theta_{m+1})^p \theta_d \theta_{d^*}}$.

In this correspondence, the number of fermions in a given ground state matches the number of tiles in the corresponding tiling. Although the proof is restricted to

a certain set of periodicities, there is strong evidence that the theorem holds for general $\vec{u}$ and $\vec{v}$. We explicitly checked the result for a variety of small systems.

The relation between tilings and ground states implies that for large enough systems ground states exist at all rational fillings (particles per site) F/N between $1/5$ and $1/4$ (see also [12] for an alternative proof). One can show that for $\vec{v} = (n, n)$ and $m = 3p$, $n = 3q$, the number of tilings grows as $4^{p+q}/\sqrt{pq}$, thus establishing sub-extensive ground state entropy [13].

For free boundary conditions in either one or both of the diagonal directions along the square lattice ($(m, -m)$ and (n, n)) the number of ground states reduces dramatically [7, 14]. One finds that it is either one or zero, except for the cylindrical case periodic in the $(m, -m)$ -direction with $m = 3p$ and $n = 3q + 2$ or $n = 3q + 3$. In that case the number of ground states is $4^{(q+1)}$.

With the above results in place, the ground state counting problem for the square lattice has been completely settled. Further pressing questions concern the nature of these ground states and of the excited state spectrum. Some progress on these matters was provided in [7], where we presented numerical results strongly indicating the presence of critical modes in ladder versions of the 2D lattice. We then argued that the full 2D lattice with (diagonal) open boundary conditions supports edge modes described by $N = 2$ superconformal field theory.

While the physical understanding of the quantum phase on the square lattice remains far from complete, one is led to a picture where the ground state corresponding to a given tiling has fermions that are confined to the area set by an individual tile, but quantum fluctuating within that space. For the particular case of a ladder with periodicity vectors $(1, 2)$ and $(L, 0)$, closed form expressions for the ground states at filling $1/4$ confirm this picture; for the more general case such explicit expressions are not available. The tiling based physical picture of the ground state wavefunctions is reminiscent of electrons in a filled magnetic Landau level, each of them effectively occupying an area set by the strength of the magnetic field. Critical edge modes naturally fit into a picture of this sort.

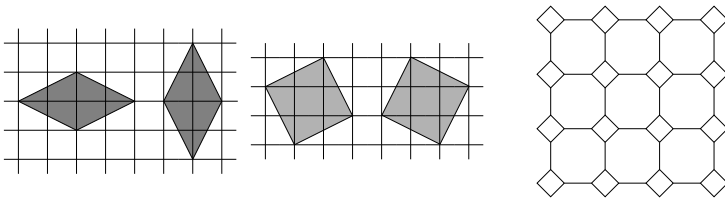


Fig. 1. On the left the four rhombuses on the square lattice, on the right the octagon-square lattice of size 4×4 .

3. Octagon-square lattice

A second 2D lattice where the supersymmetric model displays sub-extensive ground state entropy is the octagon-square lattice (Fig. 1). The growth behavior of the numbers of ground states on the plane, cylinder and torus is similar to that of the square lattice. A big difference, however, is that here all ground states reside at $1/4$ filling. This hugely simplifies the computation of the degeneracies. For the plane we find that the ground state is unique. For the cylinder with $M \times L$ square plaquettes, where M is the number of square plaquettes along the periodic, horizontal direction and L along the open, vertical direction, the number of ground states is 2^L . Finally, for the torus the number of ground states is $2^M + 2^L - 1$ [5].

There is again a relatively simple physical picture which we propose as a basis for further analysis of physical properties. This picture reflects the systematics uncovered by the analysis of the associated cohomology problem as well as results for small system sizes. The basic building block of the many-body ground states is the 1-fermion ground state on an isolated square plaquette. The unique many-body ground state on the plane essentially has individual fermions occupying this lowest 1-plaquette orbital, again allowing the analogy with a filled Landau level. Closing boundaries leads to the possibility that electrons on horizontal or vertical rows of plaquettes ‘shift’ into a second 1-fermion state, this way building up the total of $2^M + 2^L - 1$ ground states. This picture can be further substantiated by allowing defects, which we bring in by adding diagonal connections in individual plaquettes.

Among the key issues that are presently on the agenda for further study are: the existence of energy gaps, the presence of bulk or edge critical modes, and interactions and braiding properties of defects. We are confident that the constraints set by the supersymmetry, which have been instrumental in the progress made so far, will allow further progress in the analysis of the remarkable ‘supertopological’ phases on the square and octagon-square lattices.

References

- [1] P. Fendley, K. Schoutens and J. de Boer, *Phys. Rev. Lett.* **90**, 120402 (2003).
- [2] E. Witten, *Nuclear Physics B* **202**, p. 253 (1982).
- [3] P. Fendley, B. Nienhuis and K. Schoutens, *J. Phys. A* **36**, p. 12399 (2003).
- [4] M. Beccaria and G. F. De Angelis, *Phys. Rev. Lett.* **94**, 100401 (2005).
- [5] P. Fendley and K. Schoutens, *Phys. Rev. Lett.* **95**, 046403 (2005).
- [6] H. van Eerten, *J. Math. Phys.* **46**, 123302 (2005).
- [7] L. Huijse et al., *Phys. Rev. Lett.* **101**, p. 146406 (2008).
- [8] L. Huijse and K. Schoutens, *Eur. Phys. J. B* **64**, p. 543 (2008).
- [9] P. Fendley, K. Schoutens and H. van Eerten, *J. Phys. A* **38**, p. 315 (2005).
- [10] J. Jonsson, *Electronic Journal of Combinatorics* **13(1)** (2006).
- [11] L. Huijse and K. Schoutens, preprint [arxiv.org:0903.0784], (2009).
- [12] J. Jonsson, Certain homology cycles of the independence complex of grid graphs, (2005).
- [13] J. Jonsson, Hard squares on grids with diagonal boundary conditions, (2006).
- [14] M. Bousquet-Mélou, S. Linusson and E. Nevo, *J. Algebraic Comb.* **27**, 423 (2008).

OTHER TALKS AND POSTERS

INVITED TALKS

SYMMETRY BREAKING IN LAUGHLIN'S STATE ON A CYLINDER

SABINE JANSEN*, ELLIOTT H. LIEB, RUEDI SEILER

**Princeton University*

E-mail: jansen@princeton.edu

In this talk we present exact results on a many-body wave function proposed by R.B. Laughlin as an approximate ground state for electrons in the fractional quantum Hall effect. When the wave function is adapted to a cylinder geometry, the correlation functions can be expressed in terms of a discrete one-dimensional polymer system, and the normalization satisfies a simple recurrence relation. This allows us to prove that on sufficiently thin cylinders, the state is periodic with respect to translations along the cylinder axis: at filling factor $1/p$, the period is p times the period of the filled Landau level.

THE LEE-HUANG-YANG FORMULA FOR DILUTE BOSE GASES

ELLIOTT H. LIEB, JAN PHILIP SOLOVEJ*

**University of Copenhagen*

E-mail: solovej@math.ku.dk

I will discuss the ground state energy of a Bose gas in the dilute limit. I will consider the usual model where the Bose particles are interacting through a two-body potential. In standard Bose gas experiments the Bose particles are atoms and it is in this case of course an idealization to model the interaction by two-body forces. For the model in question it is believed that there is at least a two term expansion of the energy in the dilute limit, where the potential appears only through its scattering length. The leading term has been known rigorously for some time. The next to leading term often referred to as the Lee-Huang-Yang term is still an open problem. There has recently been some progress in understanding this term and I will review it in this talk. The problem is interesting not only for establishing the formula, but also because it will validate aspects of Bogolubov's theory for Bose gases.

ENTANGLEMENT IN QUANTUM SPIN SYSTEM

FRANK VERSTRAETE

University of Vienna

E-mail: fverstraete@gmail.com

The field of quantum information theory has recently led to a renewed interest into the field of quantum spin systems. This is due to the fact that strongly correlated quantum spin systems exhibit large amounts of entanglement, the main resource for doing quantum computation. We will review how the entanglement in those systems can be quantified using area laws, and how entanglement theory yields new insights into the structure of the associated low-energy wavefunctions. We will also discuss connections with computational complexity and the renormalization group.

CONTRIBUTED TALKS

SHARP TRACE ASYMPTOTICS FOR A CLASS OF 2D MAGNETIC OPERATORS

HORIA D. CORNEAN*, SØEREN FOURNAIS, RUPERT FRANK, BERNARD HELFFER

**Department of Mathematical Sciences, Aalborg University*

E-mail: cornean@math.aau.dk

Our work is motivated by a paper by H. Kunz in which he studied (among other things) the boundary correction for the grand-canonical pressure and density of a Fermi gas confined to a large 2D box submitted to a constant magnetic field, and with Dirichlet boundary conditions.

Our main theorem is a sharp semiclassical trace estimate which provides a rigorous proof to the formulas announced by Kunz, and extends them to the case of Neumann boundary conditions. Moreover, the same theorem provides several other results on the integrated density of states for operators of the type $(-i\hbar\nabla - \tilde{\mu}\mathbf{A})^2$ in $L^2(\Omega)$ with Dirichlet and Neumann boundary conditions.

TWO INTEGRABLE QUANTUM IMPURITY MODELS

HANS-PETER ECKLE

University of Ulm

E-mail: hans-peter.eckle@gmx.net

We report exact results for two related integrable quantum impurity models. The impurity, a localized magnetic moment, is coupled to a mesoscopic ring. A magnetic flux through the ring induces a persistent current via the Aharonov-Bohm effect. This persistent current, a genuine finite-size effect, is calculated exactly by a finite-size Bethe Ansatz analysis for the two models. Remarkably, we find that the persistent current is

insensitive to the presence of the quantum impurity. This protection of the persistent current from the quantum impurity only exploits the generic symmetry of the Bethe Ansatz equations. We therefore argue that this protection is a universal property of a large class of integrable quantum impurity models. Therefore it constitutes a further characteristic of integrable quantum impurity models, similar to the absence of backscattering at an integrable impurity.

THE ATOMIC PHOTOEFFECT IN NON-RELATIVISTIC QED

MARCEL GRIESEMER, HERIBERT ZENK

Universität Stuttgart

E-mail: marcel@mathematik.uni-stuttgart.de

In this talk I present a mathematical analysis of the photoelectric effect for one-electron atoms in the framework of non-relativistic QED. Photo-ionization is treated as a scattering process, where in the remote past an atom in its ground state is targeted by one or several photons, while in the distant future the atom is ionized and the electron escapes to spacial infinity. The main result shows, that the ionization probability, to leading order in the fine-structure constant, α , is correctly given by formal time-dependent perturbation theory, and, moreover, that the dipole approximation produces an error of only sub-leading order in α . In this sense the dipole approximation is rigorously justified.

TOPOLOGICAL ORDER IN QUANTUM LATTICE MODELS

JIRÍ VALA, GRAHAM KELLS, AHMET BOLUKBASI, NIALL MORAN

National University of Ireland Maynooth

E-mail: graham.kells@gmail.com

We investigate quantum lattice models that are believed to form topologically ordered states known also as topological phases. Our particular focus is on the Kitaev honeycomb spin-1/2 lattice model [1]. In the absence of external magnetic field, the model is exactly solvable and its phase diagram exhibits a gapless phase and an abelian topological phase whose effective description is given by the $\mathbb{Z}_2 \times \mathbb{Z}_2$ topological field theory. As known from the perturbation theory, a weak magnetic field has no dramatic effect on the abelian topological phase but turns the gapless phase into a non-abelian topological phase whose effective description is given by the Ising topological field theory. The quasiparticle excitations of this phase are nonabelian anyons satisfying the Ising fusion rules.

We particularly study the Kitaev honeycomb lattice on torus [2]. We describe symmetries of this model and review the perturbative mapping of its abelian topological phase onto the $\mathbb{Z}_2 \times \mathbb{Z}_2$ square lattice model known as the toric code. We provide the classification of finite size effects on the model low-energy spectral properties [3]. In this context, special attention is given to the thin-torus limit and related conformal field theory data. We then investigate properties of the model's vortex excitations and complete with discussion of topological degeneracy of the model and the effect of the magnetic field [2,4].

We conclude with a brief review of other lattice models whose low energy spectra provide realization of topological field theories and outline applications of topologically ordered systems for quantum information processing.

- [1] A. Kitaev, Ann. Phys. 321, 2 (2006).
- [2] G. Kells et al., Phys. Rev. Lett. 101, 240404 (2008).
- [3] G. Kells et al., J. Stat. Mech. - Theory Exp. P03006 (2009).
- [4] V. Lahtinen et al, Ann. Phys. 323, 2286 (2008).

POSTERS

BHAMIDIPATI CHANDRASEKHAR (Instituto de Fisica, Universidade de Sao Paulo), with PRASANTA K. PANIGRAHI and VIVEK M. VYAS: **UNITARY FERMİ GAS: SCALING SYMMETRIES AND EXACT MAP**

NATALIE GILKA (Department of Mathematics, University of Copenhagen): **BOUNDS ON DIATOMIC MOLECULES IN A RELATIVISTIC MODE**

GRAHAM KELLS (National University of Ireland, Maynooth), with JOOST K. SLINGERLAND and JIŘÍ VALA: **A DESCRIPTION OF KITAEV'S HONEYCOMB MODEL WITH TORIC-CODE STABILIZERS**

Z. ÖZTAŞ (Anadolu University), with CEM YÜCE: **STABILITY OF A DIPOLAR BOSE-EINSTEIN CONDENSATION**

DENIS O. POLISHCHUK (Shevchenko Nat University, Kyiv), with V.I. GERASIMENKO: **ON EVOLUTION OF CORRELATIONS OF BOSE AND FERMI MANY-PARTICLE SYSTEMS**

VIACHESLAV O. SHTYK (Institute for Theoretical Physics, Kyiv): **ON MEAN-FIELD ASYMPTOTIC OF SOLUTION OF QUANTUM BBGKY HIERARCHY**

FRANÇOISE TRUC (Inst Fourier, Université Grenoble 1), with YVES COLIN DE VERDIÈRE: **CONFINING QUANTUM PARTICLES WITH A PURELY MAGNETIC FIELD**

New Frontiers in Mathematical Physics

Session organizers:

Joseph Avron

Ruth Durrer

Israel Michael Sigal

GLIMPSES INTO THE EARLY UNIVERSE

DAVID LANGLOIS

*APC (Astroparticules et Cosmologie), CNRS-Université Paris 7
10, rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France
E-mail: langlois@apc.univ-paris7.fr*

This contribution gives a brief overview of the theoretical ideas underlying our current understanding of the early Universe. Confronting the predictions of the early Universe models with cosmological observations, in particular of the cosmic microwave background fluctuations, will improve our knowledge about the physics of the primordial Universe.

Keywords: Cosmology; general relativity; inflation.

1. Introduction

Inflation is today the main theoretical framework to describe the early Universe. In thirty years of existence, inflation, in contrast with earlier competitors, has survived the confrontation with cosmological data, which have tremendously improved over the years. Indeed, the fluctuations of the Cosmic Microwave Background (CMB) had not yet been measured when inflation was invented, whereas they give us today a remarkable picture of the cosmological perturbations in the early Universe. In the future, one can hope that even more precise observations will allow us to test inflation further, and also to discriminate between the many different possible realizations of inflation.

This contribution discusses the basic ideas underlying inflation (many more details can be found in e.g. [1]) and some more recent results.

2. Cosmological evolution

Modern cosmology is based on the theory of general relativity, according to which our Universe is described by a four-dimensional geometry $g_{\mu\nu}$ that satisfies Einstein's equations

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (1)$$

where $R_{\mu\nu}$ is the Ricci tensor, $R \equiv g^{\mu\nu} R_{\mu\nu}$ the scalar curvature and $T_{\mu\nu}$ the energy-momentum tensor of the matter distribution.

The basic assumption of cosmology, which has been confirmed by observations so far, is to consider, as a first approximation, the universe as being homogeneous

and isotropic. This leads to the FLRW (Friedmann-Lemaître-Robertson-Walker) spacetimes, with metric

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - \kappa r^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (2)$$

where $\kappa = 0, -1, 1$ determines the curvature of spatial hypersurfaces: respectively flat, elliptic or hyperbolic. Moreover, the matter content compatible with homogeneity and isotropy is necessarily characterized by an energy-momentum tensor of the form

$$T^\mu_\nu = \text{Diag} [-\rho(t), P(t), P(t), P(t)] \quad (3)$$

where ρ corresponds to an energy density and P to a pressure.

Substituting the metric (2) and the energy-momentum tensor (3) into Einstein's equations (1) gives the Friedmann equations

$$H^2 \equiv \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{\kappa}{a^2}, \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3P), \quad (4)$$

which govern the time evolution of the scale factor $a(t)$. There are several types of matter in the Universe, in particular pressureless matter (baryonic matter and the mysterious dark matter) and a gas of cosmological photons, characterized by a temperature T , which scales like $1/a(t)$. Going backwards in time, radiation dominates, with higher and higher temperatures in the early Universe.

From an observational point of view, the two most important events in cosmological history are : i) nucleosynthesis ($T \sim 0.1$ MeV), when the lightest nuclei were formed; ii) last scattering ($T \sim 3000$ K), when the Universe became quasi-transparent (due to the sudden suppression of interactions between photons and matter, as nuclei and electrons combined into neutral atoms). Photons that were emitted at that epoch are observed in the CMB radiation, discovered in 1964 by Penzias and Wilson. In 1992, the COBE satellite detected its anisotropies at a level of 10^{-5} . Since then, these fluctuations have been measured with increasing precision, lately by the WMAP satellite and in the near future, by the Planck satellite. One of the main goals of primordial cosmology is to explain the origin of these primordial fluctuations.

3. Inflation

Inflation is a phase of accelerated expansion, i.e. $\ddot{a} > 0$, in the early Universe. Initially, inflation was introduced to provide an explanation for several puzzles of the standard hot Big Bang model, in particular the flatness problem (or why the present spatial geometry is so close to Euclidean geometry) and the horizon problem (or why the CMB sky is so homogeneous on scales larger than the causal horizon at the time of last scattering, as defined in the standard Big Bang model). But it was realized, soon after, that inflation also gives a very natural explanation for the origin of primordial perturbations: they simply arise from quantum vacuum

fluctuations that got amplified when their wavelength, proportional to the scale factor, is stretched out beyond the Hubble radius H^{-1} .

The simplest way to get inflation is to assume the existence of a scalar field, governed by the action

$$S_\phi = \int d^4x \sqrt{-g} \left(-\frac{1}{2} \partial^\mu \phi \partial_\mu \phi - V(\phi) \right), \quad (5)$$

where $g \equiv \det(g_{\mu\nu})$ and $V(\phi)$ is the potential of the scalar field. In a FLRW space-time, the energy density and pressure are respectively given by

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad P = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (6)$$

Whenever the kinetic energy is negligible with respect to the potential energy, the equation of state is effectively $P \simeq -\rho$, which leads to an acceleration, according to the second Friedmann equation in (4). This can happen in the so-called slow-roll regime, for sufficiently flat potentials.

4. Cosmological perturbations

Let us now consider the perturbations during inflation. In addition to the scalar field perturbation $\delta\phi$, one must take into account the metric perturbations as well and introduce the (scalarly) perturbed metric

$$ds^2 = -(1 + 2A)dt^2 + 2a(t)\partial_i B dx^i dt + a^2(t) [(1 - 2\psi)\delta_{ij} + 2\partial_i \partial_j E] dx^i dx^j. \quad (7)$$

Using Einstein's equations and coordinate freedom, it can be shown that there is in fact a single dynamical scalar degree of freedom,

$$Q \equiv \delta\phi + \frac{\dot{\phi}}{H} \psi \equiv \frac{\dot{\phi}}{H} \mathcal{R}. \quad (8)$$

It can be seen either as a pure scalar field perturbation (in a coordinate system such that $\psi = 0$) or as a pure metric perturbation $\mathcal{R}$ (up to the factor $\dot{\phi}/H$) if one chooses the uniform scalar field hypersurfaces as constant time hypersurfaces. Its dynamics is governed by the action

$$S = \frac{1}{2} \int d\tau d^3x \left[u'^2 + \partial_i u \partial^i u + \frac{z''}{z} u^2 \right], \quad z \equiv a \frac{\dot{\phi}}{H} \quad (9)$$

where it is convenient to use the new variable $u \equiv aQ$ and the conformal time τ ($d\tau = dt/a$).

Let us now quantize u by following the standard procedure of quantum field theory. One treats u as a quantum field denoted $\hat{u}$, which can be expanded in Fourier space as

$$\hat{u}(\tau, \vec{x}) = \frac{1}{(2\pi)^{3/2}} \int d^3k \left\{ \hat{a}_{\vec{k}} u_k(\tau) e^{i\vec{k} \cdot \vec{x}} + \hat{a}_{\vec{k}}^\dagger u_k^*(\tau) e^{-i\vec{k} \cdot \vec{x}} \right\}, \quad (10)$$

where the $\hat{a}_{\vec{k}}^\dagger$ and $\hat{a}_{\vec{k}}$ are creation and annihilation operators that satisfy the usual commutation rules. In the slow-roll regime, the expansion is quasi-de Sitter (i.e. such

that $\dot{H} \ll H^2$) with $a \simeq -1/H\tau$, and $z''/z \simeq a''/a \simeq 2/\tau^2$ (where τ grows from $-\infty$ to 0). The most natural choice for the mode function $u_k(\tau)$ is the particular solution (of the classical equation of motion)

$$u_k = \sqrt{\frac{\hbar}{2k}} e^{-ik\tau} \left(1 - \frac{i}{k\tau}\right), \quad (11)$$

which means that each Fourier mode is initially in the usual Minkowski vacuum, when $k|\tau| \gg 1$, i.e. when its wavelength is smaller than the Hubble radius. Later, its wavelength is stretched on super Hubble scales, i.e. $k|\tau| \ll 1$, and the mode undergoes an amplification (the last term in the parentheses blows up). The resulting power spectrum (i.e. the Fourier transform of the correlation function) is

$$\mathcal{P}_Q \equiv \frac{k^3}{2\pi^2} \frac{|u_k|^2}{a^2} \simeq \hbar \left(\frac{H}{2\pi}\right)^2, \quad (k \ll aH) \quad (12)$$

which corresponds to a curvature fluctuation $\mathcal{P}_{\mathcal{R}}^{1/2} = H^2/(2\pi\dot{\phi})$ (with $\hbar = 1$).

It is easy to relate these fluctuations generated *during* inflation to the fluctuations of ordinary matter in the subsequent radiation and matter dominated eras, by invoking a conservation law. Indeed, the conservation of the energy-momentum tensor for any perfect fluid, characterized by the energy density ρ , the pressure p and the four-velocity u^a , leads to the *exact* relation [2, 3]

$$\dot{\zeta}_a \equiv \mathcal{L}_u \zeta_a = -\frac{\Theta}{3(\rho + p)} \left(\nabla_a p - \frac{\dot{p}}{\dot{\rho}} \nabla_a \rho \right), \quad (13)$$

where we have defined

$$\zeta_a \equiv \nabla_a \alpha - \frac{\dot{\alpha}}{\dot{\rho}} \nabla_a \rho, \quad \Theta = \nabla_a u^a, \quad \alpha = \frac{1}{3} \int d\tau \Theta, \quad (14)$$

and where a dot denotes a Lie derivative with respect to u^a (which reduces to a derivative along u^a for scalars, e.g. $\dot{\rho} \equiv u^a \nabla_a \rho$). The quantity e^α can be interpreted as an inhomogeneous generalization of the scale factor, as defined by an observer following the fluid. For linear perturbations, the identity (13) implies that the quantity

$$\zeta \equiv -\psi - \frac{H}{\dot{\rho}} \delta\rho \quad (15)$$

is conserved on super-Hubble scales ($k \ll aH$) for adiabatic perturbations, i.e. such that $\delta p - (\dot{p}/\dot{\rho})\delta\rho = 0$. Moreover, it can be shown that $-\zeta$ coincides with $\mathcal{R}$ on super-Hubble scales. For single field inflation, this conservation law holds and the curvature perturbation remains unchanged until the perturbation reenters the Hubble radius, much later during the radiation era.

5. Beyond the simplest models

So far, the simplest models of inflation are compatible with observational data (see [4]) but it is worth studying more refined models for at least two reasons.

First, models inspired by high energy physics are usually more complicated than the simplest phenomenological inflationary models. Second, exploring larger classes of inflation models and identifying their specific observational features is a useful preparation for the interpretation of future data. At present, two types of extensions beyond the simplest scenarios have been mainly studied: models with non standard kinetic terms and/or with multiple scalar fields.

Scenarios involving several scalar fields include models with *multiple inflatons*, where several scalar fields affect directly the inflationary evolution, but also models where the extra scalar field(s) plays a rôle only later. In all cases, the crucial novelty is the generation during inflation of extra perturbations, usually called entropy modes, in addition to the adiabatic mode (corresponding to fluctuations along the inflation trajectory). The entropy fluctuations can be transferred into the final curvature perturbation, during or after inflation. This means that the quantities $\mathcal{R}$ or ζ are *a priori* no longer conserved in a multi-field set-up, as first pointed out in [5] (see also [6]).

In some of these more sophisticated models, the primordial perturbations exhibit non-Gaussianities that could be detectable in future observations, whereas the simplest single field models predict an undetectable level of non-Gaussianities. In this context, an important observable is the three-point function or its Fourier transform, often written as

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = \frac{6}{5} f_{\text{NL}}(k_1, k_2, k_3) [P_\zeta(k_1)P_\zeta(k_2) + 2 \text{ perms}] (2\pi)^3 \delta^{(3)}\left(\sum_i \mathbf{k}_i\right) \quad (16)$$

where $P_\zeta(k)$ is defined by $\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle = P_\zeta(k)(2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2)$. The three-point function can be computed for any model of inflation by using the (non-linear) relation between the curvature perturbation ζ and the various scalar field fluctuations generated during inflation. Intrinsic non-Gaussianities of the scalar fields, i.e. non-vanishing three-point functions for the scalar field perturbations, lead to non-Gaussianities of *equilateral* shape (where the signal peaks at $k_1 \sim k_2 \sim k_3$) whereas a non-linear (classical) relation between ζ and the scalar field fluctuations lead to non-Gaussianities of *local* shape (which peaks at $k_1 \ll k_2, k_3$).

An analysis of general multi-field models with an action of the form

$$S = \int d^4x \sqrt{-g} \left[\frac{R}{16\pi G} + P(X^{IJ}, \phi^K) \right], \quad X^{IJ} = -\frac{1}{2} \nabla_\mu \phi^I \nabla^\mu \phi^J, \quad (17)$$

where P is an arbitrary function, can be found in [7] (see also [8] for a more restricted class of models).

An illustrative example combining both non-Gaussianities and multi-field effects is Dirac-Born-Infeld (DBI) inflation [9, 10], which arises from the motion of a $D3$ -brane in an internal six-dimensional compact space, as one can encounter in the context of string theory compactifications. In this case, the function P is given by

$$P = -\frac{1}{f(\phi^I)} \left[\sqrt{\det(\delta_\nu^\mu + f G_{IJ} \partial^\mu \phi^I \partial_\nu^J)} - 1 \right] - V(\phi^I), \quad (18)$$

where the scalar fields ϕ^I are directly related to the compact coordinates of the brane, and where the functions f and G_{IJ} , which depends on the ϕ^I , are determined by the 10-dimensional geometry (bulk forms can also be included [11]).

During multi-field DBI inflation, both adiabatic and entropic modes can be generated, with power spectra $\mathcal{P}_{Q_\sigma} = (H/2\pi)^2$ and $\mathcal{P}_{Q_s} = (H/2\pi c_s)^2$ respectively [12]. The entropy power spectrum is thus enhanced by its dependence on the sound speed $c_s = \sqrt{1 - f\dot{\sigma}^2} < 1$, and the final curvature is

$$\mathcal{R} = \frac{H}{\dot{\sigma}} (Q_\sigma + T_{\mathcal{RS}} c_s Q_s), \quad \dot{\sigma}^2 \equiv G_{IJ} \dot{\phi}^I \dot{\phi}^J, \quad (19)$$

where $T_{\mathcal{RS}}$ parametrizes the transfer from entropy modes into the curvature perturbation.

By expanding the action beyond the second order, one can also determine the non-Gaussianities generated in this class of models [12]. This leads to a bispectrum of equilateral shape with

$$f_{NL}^{(3)} = -\frac{35}{108} \frac{1}{c_s^2} \frac{1}{1 + T_{\mathcal{RS}}^2}, \quad (20)$$

which corresponds to the single-field result but with a suppression due to the entropy-curvature transfer. At the next order, in the trispectrum, multi-field effects induce a shape of non-Gaussianities that differs from the single-field case [13].

References

- [1] D. Langlois, “Inflation, quantum fluctuations and cosmological perturbations,” in *Cargese 2003, Particle physics and cosmology*, p. 235-278 [arXiv:hep-th/0405053].
- [2] D. Langlois and F. Vernizzi, Phys. Rev. Lett. **95**, 091303 (2005) [arXiv:astro-ph/0503416].
- [3] D. Langlois and F. Vernizzi, Phys. Rev. D **72**, 103501 (2005) [arXiv:astro-ph/0509078].
- [4] E. Komatsu *et al.* [WMAP Collaboration], Astrophys. J. Suppl. **180**, 330 (2009) [arXiv:0803.0547 [astro-ph]].
- [5] A. A. Starobinsky and J. Yokoyama, “Density fluctuations in Brans-Dicke inflation”, gr-qc/9502002
- [6] Z. Lalak, D. Langlois, S. Pokorski and K. Turzyski, JCAP **0707**, 014 (2007) [arXiv:0704.0212 [hep-th]].
- [7] D. Langlois, S. Renaux-Petel, D. A. Steer and T. Tanaka, Phys. Rev. D **78**, 063523 (2008) [arXiv:0806.0336 [hep-th]].
- [8] D. Langlois and S. Renaux-Petel, JCAP **0804**, 017 (2008) [arXiv:0801.1085 [hep-th]].
- [9] E. Silverstein and D. Tong, Phys. Rev. D **70**, 103505 (2004) [arXiv:hep-th/0310221];
- [10] M. Alishahiha, E. Silverstein and D. Tong, Phys. Rev. D **70**, 123505 (2004) [arXiv:hep-th/0404084].
- [11] D. Langlois, S. Renaux-Petel and D. A. Steer, JCAP **0904**, 021 (2009) [arXiv:0902.2941 [hep-th]].
- [12] D. Langlois, S. Renaux-Petel, D. A. Steer and T. Tanaka, Phys. Rev. Lett. **101**, 061301 (2008) [arXiv:0804.3139 [hep-th]].
- [13] S. Mizuno, F. Arroja, K. Koyama and T. Tanaka, Phys. Rev. D **80**, 023530 (2009) [arXiv:0905.4557 [hep-th]].

SPATIOTEMPORAL DYNAMICS IN ECOLOGY: INSIGHTS FROM PHYSICS

JONATHAN A. SHERRATT

*Department of Mathematics and Maxwell Institute for Mathematical Sciences,
Heriot-Watt University, Edinburgh EH14 4AS, UK
E-mail: jas@ma.hw.ac.uk*

MATTHEW J. SMITH

*Microsoft Research,
7 J.J. Thompson Avenue, Cambridge, CB3 0FB, UK
E-mail: Matthew.Smith@microsoft.com*

JENS D. M. RADEMACHER

*National Research Institute for Mathematics and Computer Science (CWI),
Science Park 123, 1098 XG Amsterdam, The Netherlands
E-mail: Jens.Rademacher@cwi.nl*

Simulations of invasion in cyclic predator-prey systems show plane waves behind the invasion front. When the selected plane wave is unstable, there is a band of plane waves of constant width, followed by spatiotemporal chaos. We describe a new method for calculating the width of this band, based on the absolute stability of plane waves in moving frames of reference. This calculation shows that the band width can be very sensitive to changes in parameters, and we discuss the ecological implications of this result.

Keywords: Plane wave; Absolute Stability; complex Ginzburg-Landau equation

1. Invasions in Cyclic Predator-Prey Systems

Many natural populations exhibit multi-year cycles in abundance. These cycles arise from a variety of different mechanisms, but in some cases there is strong evidence that the underlying cause is the interaction between a predator population and its prey. For example, vole populations in Fennoscandia cycle with a period of 4 or 5 years [1]. The main terrestrial predator of voles is weasels, and when these are removed from a region of habitat via live trapping, the multi-year cycles in vole density disappear [2]. Predator-prey cycles have been studied mathematically for nearly 100 years, and a wide variety of different models have been proposed, including a large number consisting of two coupled *odes* (see the books [3, 4] for review). A typical example is the Rosenzweig-MacArthur model [5], which has the

dimensionless form

$$\boxed{\text{predators}} \quad dp/dt = \overbrace{ACph/B(1+Ch)}^{\text{benefit from predation}} - \overbrace{p/AB}^{\text{death}} \quad (1)$$

$$\boxed{\text{prey}} \quad dh/dt = \underbrace{h(1-h)}_{\text{intrinsic birth \& death}} - \underbrace{Cph/(1+Ch)}_{\text{predation}}. \quad (2)$$

Here p and h are the densities of predators and prey at time t , and A , B and C are positive parameters. Eqs. (1,2) have a unique “coexistence” steady state in which h and p are both non-zero, and as C increases through $(A+1)/(A-1)$, this steady state becomes unstable via undergo a standard supercritical Hopf bifurcation, leading to a stable limit cycle.

It is increasingly common for field studies of cyclic populations to record abundances at a number of different locations. In many cases this data reveals that the population is not spatially homogeneous; rather, there is a one-dimensional gradient in the phase of the cycles, so that the overall dynamics have the form of a plane wave [6]. By this, we mean that the population density is a periodic function of $x - at$, where x is a one-dimensional spatial coordinate, t is time, and a is the (phase) velocity. Note that the terms “plane wave”, “wavetrain”, and “periodic travelling wave” are synonymous. The model (1,2) can easily be adapted to study spatiotemporal behaviour, by adding diffusion terms to both equations, representing unbiased local dispersal. For simplicity, we assume that the populations have the same diffusion coefficient (which we arbitrarily set to 1); however, the phenomena that we will report also occur when the two coefficients are different [7] or even density-dependent [8]. Standard theory shows that oscillatory reaction-diffusion equations have a one-parameter family of plane wave solutions; some waves in the family are stable as solutions of the *pdes*, but others are unstable [9].

For predator-prey systems, a particularly important spatiotemporal scenario is the invasion of prey by predators. We simulate this by solving our *pde* model on a large spatial domain with zero Neumann boundary conditions, and with initial conditions consisting of a uniform prey density, with predators zero except near to one boundary. The *pde* solutions exhibit an advancing wave of predators and a corresponding receding wave of prey. For C above the threshold for population cycles, the behaviour behind the invasion front consists of spatiotemporal oscillations [10–12]. In some cases, these are stable plane waves. However when the appropriate plane wave is unstable, more complicated dynamics occur. Immediately behind the invasion front, the population densities are almost constant, at their (unstable) coexistence steady state. Behind this, there is a plane wave band whose width remains constant as the invasion progresses, and which is followed by spatiotemporal chaos [11, 12, Fig. 1]. The width of the plane wave band is the key determinant of the behaviour observed in practice behind the invasion. We have recently developed a method that enables direct calculation of this width, for the first time [13].

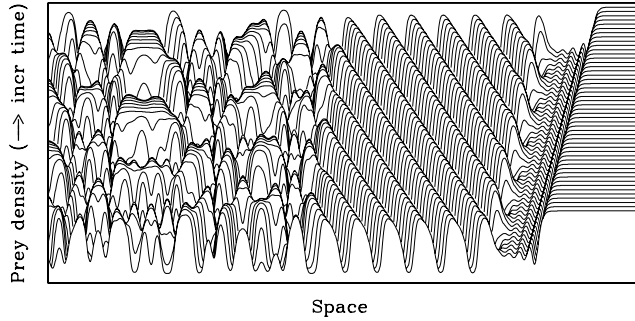


Fig. 1. Plane wave generation in a simulated invasion of a prey population by predators, using Eqs. (1,2) augmented by diffusion terms for p and h , with both diffusion coefficients set to 1. We plot prey density h as a function of space and time. There is a receding invasive wave front, behind which the solution settles at the (unstable) coexistence steady state. A plane wave band then develops; the plane wave is unstable, and destabilises to give irregular spatiotemporal oscillations. The predator dynamics are directly analogous, with the oscillations of the two populations being out of phase. The initial conditions correspond to a prey-only state everywhere, except for a small non-zero predator density at the left hand boundary. The parameters are $A = 1.3$, $B = 1.2$, $C = 11.0$. The domain length is 1000, and the solution is plotted for $1155 \leq t \leq 1284$. The equations were solved using a semi-implicit finite difference method, with a grid spacing of 0.5 and a time step of 3×10^{-5} .

2. Amplitude Equation and Band Width Calculation

Detailed calculations show that the plane waves behind the invasion arise via the invasion front leaving the system at the coexistence steady state and also initiating an invasion of that state, in the opposite direction [14]. This means that close to Hopf bifurcation in the kinetics, the plane wave band can be studied using an amplitude equation. Since the Hopf bifurcation is supercritical and the diffusion coefficients of predators and prey are equal, the appropriate amplitude equation is the cubic complex Ginzburg-Landau equation with zero linear dispersion:

$$A_t = A_{xx} + A - (1 + ic)|A|^2 A.$$

The relevant solution of this equation is a propagating front, with $A = 0$ ahead of the front and a plane wave behind it [15, 16]. When the selected plane wave is unstable, the plane waves occur in a band that is followed by spatiotemporal chaos. The width of the band increases (from zero) in the early stages of the invasion, but at large times it is constant.

The key to understanding the width of the plane wave band is to consider absolute stability [17, 18] when viewed in a frame of reference moving with a fixed, arbitrary velocity V . That is, we consider whether perturbations to the plane wave grow or decay over time when viewed at a fixed point travelling with velocity V . We denote by $\lambda_{max}(V)$ the growth rate of the most unstable linear mode, with $\nu_{max}(V)$ being the corresponding spatial eigenvalue. Using general theory of absolute stability [17, 18], we reduce the calculation of these quantities to the numerical tracking of

solutions of a quartic polynomial as parameters in the coefficients vary; details are given in [13]. There is a range of velocities (V_L, V_R) for which $\lambda_{max} > 0$; all perturbations decay in frames of reference moving with velocities outside this range.

Having calculated $\lambda_{max}(V)$, we can address the width of the plane wave band. For this, we require a precise definition of its left-hand edge, which we take as the first point at which the perturbations to the plane wave that are present immediately behind the invasion front become amplified by an arbitrary factor $\mathcal{F}$. We show in [13] that the dependence of the plane wave band width on $\mathcal{F}$ and c decouples, having the form $\log(\mathcal{F})\mathcal{W}(c)$. We refer to $\mathcal{W}(c)$ as the “band width coefficient”; it contains all of the dependence of the band width on ecological parameters. Let (x^*, t^*) be a point on the invasion front. We make the generic assumption that the most unstable linear modes are present in the perturbation given to the plane wave at such a point. As t increases above t^* , these perturbations will spread out in space and time, growing along all rays $x = x^* + (t - t^*)V$ with $V \in (V_L, V_R)$. On any such ray, the initial perturbation becomes amplified by the factor $\mathcal{F}$ at time $t_{crit}(V) = t^* + \log(\mathcal{F})/\text{Re}[\lambda_{max}(V)]$, at location $x_{crit}(V) = x^* + V \log(\mathcal{F})/\text{Re}[\lambda_{max}(V)]$. The left-hand edge of the plane wave band occurs at the point on the locus (as V varies) of $(x_{crit}(V), t_{crit}(V))$ that is closest to the invasion front. We show in [13] that this implies $\mathcal{W} = 1/\text{Re}[\nu_{max}(V_{band})]$, where

$$(V_{band} - c_{inv})\text{Re}[\nu_{max}(V_{band})] = \text{Re}[\lambda_{max}(V_{band})]. \quad (3)$$

It is of course entirely expected that $\mathcal{W}$ is the reciprocal of the real part of the spatial eigenvalue; our key result is Eq. (3) for the frame velocity V_{band} at which this eigenvalue should be calculated. Note that in general V_{band} is different from the velocity that maximises $\lambda_{max}(V)$.

3. Band Width Sensitivity and Ecological Implications

The parameter c in the complex Ginzburg-Landau equation can be related to ecological parameters via the standard process of reduction to normal form [19, 20]. Therefore our formula for the band width coefficient enables precise predictions of the variation in the width of the plane wave band as ecological parameters are altered. In [13] we show a contour plot of the width in a key ecological parameter plane. An important conclusion from such plots is that the width of the plane wave band can be very sensitive to parameters. For example, in the case of parameter estimates for the interaction between field voles and weasels [20], a 5% increase in vole birth rate increases the band width by 22%. An even more striking example of this sensitivity is provided by the interaction between the zooplankton *Daphnia pulex* and the phytoplankton *Chlamydomonas reinhardtii*. Using estimates of the relevant parameters [21], we predict that the width of a plane wave band behind invasion would be extremely sensitive to the estimate of zooplankton birth rate, with a reduction of only 5.2% sufficient to double the band width. Such sensitivity is particularly important in the context of a changing climate, which is both increasing the frequency of ecological invasions [22] and also changing the parameters

of many oscillatory ecological systems [23, 24]. The sensitivity of the band width to such parameter changes suggests a very profound effect of climate change on the spatiotemporal dynamics of cyclic populations.

Acknowledgments

We thank Björn Sandstede (Brown University) and Leonid Brevdo (Louis Pasteur University) for helpful discussions. JAS was supported in part by a Research Fellowship from the Leverhulme Trust. MJS was supported by Microsoft Research Cambridge (MSRC) and thanks the MSRC Tools and Technology Group (Computational Science) for technical assistance. JDMR acknowledges support by the NWO program NDSN+.

References

- [1] P. Turchin, *Complex Population Dynamics: a Theoretical/Empirical Synthesis* (Princeton University Press, 2003).
- [2] T. Klemola, M. Koivula, E. Korpimäki and K. Norrdahl, *J. Animal Ecol.* **66**, 607 (1997).
- [3] R. M. May, *Stability and Complexity in Model Ecosystems* (Princeton University Press, 1973).
- [4] J.D. Murray, *Mathematical Biology I: an Introduction* (Springer-Verlag, New York, 2002).
- [5] M. L. Rosenzweig and R. H. MacArthur, *Am. Nat.* **97**, 209 (1963).
- [6] J. A. Sherratt and M. J. Smith, *J. R. Soc. Interface* **5**, 483 (2008).
- [7] M. J. Smith and J. A. Sherratt, *Physica D* **236**, 90 (2007).
- [8] M. J. Smith, J. A. Sherratt and X. Lambin, *J. Theor. Biol.* **254**, 264 (2008).
- [9] N. Kopell and L. N. Howard, *Stud. Appl. Math.* **52**, 291 (1973).
- [10] M. R. Garvie, *Bull. Math. Biol.* **69**, 931 (2007).
- [11] J. A. Sherratt, M. A. Lewis and A. C. Fowler, *Proc. Natl. Acad. Sci. USA* **92**, 2524 (1995).
- [12] S. V. Petrovskii and H. Malchow, *Math. Comp. Modelling* **29**, 49 (1999).
- [13] J. A. Sherratt, M. J. Smith and J. D. M. Rademacher, *Proc. Natl. Acad. Sci. USA* **106**, 10890 (2009).
- [14] J. A. Sherratt, *Physica D* **117**, 145 (1998).
- [15] K. Nozaki and N. Bekki, *Phys. Rev. Lett.* **51**, 2171 (1983).
- [16] W. van Saarloos, *Phys. Rep.* **386**, 29 (2003).
- [17] B. Sandstede and A. Scheel, *Physica D* **145**, 233 (2000).
- [18] J. D. M. Rademacher, B. Sandstede and A. Scheel, *Physica D* **229**, 166 (2007).
- [19] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer-Verlag, Berlin, 1983).
- [20] J. A. Sherratt, *Ecology Lett.* **4**, 30 (2001).
- [21] R. M. Nisbet, E. McCauley, A. M. De Roos, W. W. Murdoch and W. S. C. Gurney, *Theor. Pop. Biol.* **40**, 125 (1991).
- [22] J. J. Hellmann, J. E. Byers, B. G. Bierwagen and J. S. Dukes, *Conserv. Biol.* **22**, 534 (2008).
- [23] R. A. Ims, J.-A. Henden and S. T. Killengreen, *TREE* **23**, 79 (2008).
- [24] K. L. Kausrud, A. Myrsterud, H. Steen, J. O. Vik, E. Østbye, B. Cazelles, E. Framstad, A. M. Eikeset, I. Myrsterud, T. Solhoy and N. C. Stenseth, *Nature* **456**, 93 (2008).

OTHER TALKS

INVITED TALKS

QUANTUM HAMILTONIAN COMPLEXITY: WHAT QUANTUM INFORMATION TELLS US ABOUT CONDENSED MATTER PHYSICS

DORIT AHARONOV*, ITAI ARAD, ZEPH LANDAU, UMESH VAZIRANI,
DANIEL GOTTESMAN, JULIA KEMPE, SANDY IRANI

** Hebrew University, Jerusalem*
E-mail: dorit.aharonov@gmail.com

The central problem in the area of condensed matter physics: understanding ground states of many body local Hamiltonians, turns out to be the quantum analogue of the major problem in computer science – the problem of satisfying local constraints over many variables. Quantum Hamiltonian complexity is a new and fast growing subfield of quantum information science which combines these two seemingly unrelated questions; it studies ground states, their structure and complexity, from a computational point of view. Over the past few years exciting new insights and intuitions arose from this marriage, and in this talk I will try to explain some of those. In particular I will focus on the recent result that 1 dimensional systems are as hard – computationally – as higher dimensional systems, as well as on new insights about the entanglement structure of ground states of general local Hamiltonians. The talk will be scattered with the many intriguing open problems in this new area.

QUANTUM QUENCHES: CURRENT SURVIVAL AND OTHER PROBLEMS

ISRAEL KLICH

University of Virginia
E-mail: klich@virginia.edu

A quantum quench is a rapid change of a quantum system from one phase to another at zero temperature. Such quenches have been the subject of growing interest in recent years due to possible realization in cold atom systems. In the talk I will describe some of the questions raised in this subject. I will study in detail the survival of super-currents in a system of impenetrable bosons subject to a quantum quench from critical super fluid phase to an insulating phase. I will describe the evolution of the current and it's

oscillations when the quench follows a Rosen-Zener profile, which is exactly solvable. (Collaboration with G. Refael and C. Lannert)

MEASUREMENT-BASED QUANTUM COMPUTATION WITH CLUSTER STATES

ROBERT RAUSSENDORF

University of British Columbia, Vancouver

E-mail: raussen@phas.ubc.ca

I give an introduction to cluster state quantum computation [1], state the main open problems and describe recent developments in the field which are interesting both from the viewpoint of Physics and Mathematics.

Cluster state quantum computation is not driven by unitary evolution but rather by projective local measurements. The scheme uses a highly entangled multi-qubit cluster state as resource for universal computation. As the computation proceeds the entanglement in the resource state is progressively destroyed. Any quantum algorithm can be implemented on a sufficiently large cluster state, by appropriate choice of the local measurement bases. The result of the computation is revealed in correlations of the individually random measurement outcomes.

From the viewpoint of physics, interesting is the fact that quantum correlations emerge as a central notion for the processing of information with cluster states [1,2]. Further, connections with fundamental aspects of quantum mechanics arise such as the Einstein-Podolsky-Rosen paradox and the Kochen-Specker Theorem.

From the viewpoint of Mathematics, to date connections with two areas of Mathematics have been established, namely Graph Theory and Topology. Graph theory arises in entanglement-based criteria for the hardness of classical simulation of cluster state quantum computation [3], and topology in fault-tolerance [4].

[1] R. Raussendorf and H.J. Briegel, PRL 86, 5188 (2001).

[2] J. Anders and D.E. Browne, Phys. Rev. Lett. 102, 050502 (2009).

[3] M. Van den Nest, W. Duer, G. Vidal, H. J. Briegel, PRA 75, 012337 (2007).

[4] R. Raussendorf, J. Harrington, PRL 98, 190504 (2007).

CLOAKING AND TRANSFORMATION OPTICS

ALLAN GREENLEAF, YAROSLAV KURYLEV, MATTI LASSAS, GÜNTHER UHLMANN*

** University of Washington*

E-mail: gunther@math.washington.edu

We describe recent theoretical and experimental progress on making objects invisible to detection by electromagnetic waves, acoustic waves and quantum waves. Maxwell's equations have transformation laws that allow for design of electromagnetic materials that steer light around a hidden region, returning it to its original path on the far side. Not only would observers be unaware of the contents of the hidden region, they would not even be aware that something was being hidden. The object, which would have no shadow, is said to be cloaked. We recount the recent history of the subject and discuss

some of the mathematical and physical issues involved.

CONTRIBUTED TALKS

PISOT NUMBERS AND QUASICRYSTALS

LUIS J. BOYA

University of Zaragoza

E-mail: luisjo@unizar.es

Quasicrystals live in one, two or three dimensions, but the most interesting are in one. In the talk I shall explain the relation of Pisot numbers (a kind of integer algebraic numbers) with one-dimensional quasicrystals. I shall also address the recent proposal of FJ Dyson to relate quasicrystals to the Riemann Hypothesis.

THE LATTICE SPIN MODELS AND HECKE GRAPHS

MICHAEL MONASTYRSKY

ITEP Moscow

E-mail: monastyrsky@itep.ru

In this talk I discuss two topics. 1. The Kramers-Wannier Duality to Spin systems with non-abelian groups of symmetry. This transformation is well known for Potts models with abelian symmetry. The complete solution of this problem for finite and compact groups was found only recently by the author and V. Buchstaber. It solved a longstanding problem and opened new perspectives. For instance, I present a new approach to McKay correspondence for the discrete subgroups of $SL(2, R)$. This class of group relates with so called Hecke groups, discrete subgroups G of $SL(2, R)$ with finite volume of the fundamental domain of G . Another application of graphs generated by Hecke groups relates with the famous Beraha conjecture. The behavior of zeroes of chromatic polynomials is determined by the partition function of Potts model on Hecke graph.

SPECTRAL SINGULARITIES AND THEIR PHYSICAL APPLICATIONS

ALI MOSTAFAZADEH

Koç University

E-mail: amostafazadeh@ku.edu.tr

A spectral singularity is an obstruction to the completeness of the eigenfunctions of the Schrödinger operator with a complex scattering potential. We offer a simple physical interpretation of spectral singularities as the real energy values for which both the reflection and transmission coefficients diverge. This in turn allows for identifying spectral singularities with resonances having a zero width. We describe the spectral singularities

of a class of point interactions as well as a piecewise constant imaginary potential. The latter arises in modeling a certain type of electromagnetic wave guides. At the frequency of a spectral singularity the wave guide operators as a resonator. The resonance phenomenon associated with the presence of spectral singularities may be used as a method of amplifying guided electromagnetic waves.

TOPOLOGY BOUNDS THE ENERGY OF KNOTS AND LINKS

RENZO L. RICCA

Mathematics Department, Università di Milano-Bicocca
E-mail: renzo.ricca@unimib.it

In this talk we present new results on the minimal energy of magnetic knots and links in ideal magnetohydrodynamics. By using classical results of Arnold (1974), Moffatt (1990) and Freedman and He (1991), we prove that the topology of magnetic knots and links provides a lower bound on the magnetic energy of the system [2]. By relying on standard relaxation techniques of the magnetic field, the groundstate energy of tight knots up to 10 crossings is determined [1], and the relationship between this energy and topological crossing number, ropelength and internal twist is examined.

These results find useful applications in the context of astrophysical flows, and provide further grounds to establish a mathematical foundation for the classification of physical knots and links based on a one-to-one correspondence between energy and topology [3].

[1] Maggioni F. & Ricca, R.L. On the groundstate energy of tight knots. Proc. R. Soc. A. Submitted (2009).

[2] Ricca, R.L. Topology bounds energy of knots and links. Proc. R. Soc. A 464, 293-300 (2008).

[3] Ricca, R.L. (Ed.) Lectures on Topological Fluid Mechanics. Lecture Notes in Mathematics vol. 1973, Springer-Verlag, Heidelberg (2009).

QUANTUM CORRECTIONS TO THE ENTROPY OF DILATON-AXION BLACK HOLES

KHALID SAIFULLAH

Department of Mathematics, Quaid-i-Azam University, Islamabad
E-mail: saifullah@qau.edu.pk

Axially symmetric Einstein-Maxwell black holes with dilaton axion charges are studied. We are particularly interested in the thermodynamics of these objects. We investigate the relationship of the Einstein field equations with the first law of thermodynamics. We also study the entropy and Bekenstein-Hawking area law for these black holes.

GEOMETRIC FOURIER ANALYSIS IN COGNITIVE VISION

JACEK TURSKI

University of Houston-Downtown
E-mail: turskij@uhd.edu

The group $SL(2, C)$ occupies a truly remarkable position in mathematics and sciences. It is inherently relevant to non-Euclidean geometries, modern complex analysis (Teichmüller spaces), and special theory of relativity. We demonstrate that $SL(2, C)$ provides geometrical and numerical framework for computational vision, including visual neuroscience and machine vision. In the conformal camera, which models eye's imaging functions, image projective transformations are given by $SL(2, C)$ acting on the camera's image plane by linear-fractional mappings. The conformal camera has its own projective Fourier analysis constructed in the framework of representation theory of $SL(2, C)$. Projective Fourier transform (PFT) provides image representation well adapted to both perspective transformations of retinal images and the retinotopy of the brain visual pathways. We use the conformal camera to process visual information during fast scanning movements called saccades. Despite these incisive eye movements, we perceive a stable world. This visual constancy is maintained by neuronal receptive fields shifts prior to saccade onset in various retinotopically organized cortical areas. These shifts may integrate visual information across saccades and eliminate the need for starting image processing anew three times per second at each fixation. However, this remapping is not perfect; around the time of saccades the perceptual space is compressed—the phenomenon called perisaccadic mislocalization. In our modeling of trans-saccadic perception, we utilize basic properties of PFT. First, the inverse PFT can be efficiently computed by FFT in logarithmic coordinates that approximate the retinotopic mappings. Second, a simple translation in retinotopic (logarithmic) coordinates, modeled by the standard shift property of Fourier transform, remaps the presaccadic scene into a postsaccadic reference frame. This shift accounts for the perisaccadic mislocalization.

POSTERS

ENZO BONACCI (L.S.S. “G.B. Grassi”): **COULD MASS BE AN ELECTRO-MAGNETIC HELIX?**

ARKADIUSZ JADCZYK (Institute of Mathematics, Toulouse): **BEYOND QUANTUM**

MIROSLAV KORBELÁŘ (Math and Physics, Charles University, Prague): **SYMMETRIES AND FINE GRADINGS OF $SL(N, \mathbb{C})$ GENERATED BY TENSOR PRODUCT OF GENERALIZED PAULI MATRICES**

EVGENY LAKHSTANOV (University of Aveiro): **EXAMPLES OF ACCEPTABLE SIMPLIFICATION OF MATHEMATICAL MODELS**

ZUZANA MASÁKOVÁ (Department of Mathematics & Doppler Institute, Czech Technical University), with MARK GRUNDLAND, JIŘÍ PATERA, NEIL A. DODGSON: **IMAGE SAMPLING WITH QUASICRYSTALS**

VLASTA PEŘINOVÁ (Palacký University, Olomouc), with ANTONÍN LUKŠ: **INVARIANT SUBSPACE METHOD IN QUANTUM OPTICS**

SEVERÍN POŠTA (Czech Technical University), with MILOSLAV HAVLÍČEK: **POLYNOMIAL IDENTITIES IN NONSTANDARD DEFORMATIONS**

VASILE POSTOLICA (Bacau State University): **EFFICIENCY AND RECENT APPLICATIONS**

ALBERTO RUBIO PONCE (Universidad Autónoma Metropolitana – Azcapotzalco, Mexico), with JOSE JUAN Peña and J. MORALES: **GENERALIZED THERMODYNAMIC PROPERTIES**

RATHINDRA NATH SEN (Ben-Gurion University, Beer-Sheva): **CAUSALITY AND THE DIFFERENTIABLE STRUCTURE OF SPACE-TIME**

PART C

Supplementary Programme

Young Researcher Symposium

Czech Technical University

July 31 – August 1, 2009

YRS PLENARY TALKS

ISOPERIMETRIC INEQUALITIES FOR EIGENVALUES OF THE LAPLACIAN

RAFAEL BENGURIA

Facultad de Física, P. Universidad Católica de Chile
E-mail: rbenguri@fis.puc.cl

Isoperimetric Inequalities have a long tradition in Mathematical Physics. There are well known isoperimetric inequalities for many physical quantities (e.g., the electrostatic capacity, the torsional rigidity, the principal eigenfrequency of a membrane, etc.) For the derivation of these inequalities several new tools of mathematical analysis and geometric measure theory had to be introduced during the last century. In this talk I will start with a quick review of some classical isoperimetric inequalities of mathematical physics, and of the main tools used in their proof. In the main part of my talk I will concentrate on isoperimetric inequalities for the eigenvalues of the Dirichlet and Neumann Laplacians, and their applications. Finally I will present a selection of open problems in this field.

ANALYTICAL FOUNDATIONS OF SYMPLECTIC FIELD THEORY

HELMUT HOFER

School of Mathematics, Institute for Advanced Study, Princeton
E-mail: hofer@math.ias.edu

Symplectic Field Theory (SFT) is the study of holomorphic curves in symplectic cobordisms and contains Gromov-Witten theory and symplectic Floer theory as special cases. The study of holomorphic curves is very complicated due to compactness and transversality issues. The algebraic invariants of SFT are obtained by a simultaneous study of infinitely interdependent first order elliptic systems. A treatment of SFT with classical (nonlinear) Fredholm theory, though possible, would be extremely cumbersome. This led to the development of a new generalized Fredholm theory in a new classes of general spaces called polyfolds.

MACROSCOPIC PHENOMENA AND RANDOMNESS

ANNA DE MASI

Dipartimento di Matematica, Università di L'Aquila

E-mail: demasi@univaq.it

Macroscopic phenomena are usually described by deterministic non linear PDEs but there are several cases where randomness plays a significant role and stochastic effects cannot be ignored, as for instance in spinodal decomposition, interface fluctuations, metastability, tunneling and hysteresis. I shall illustrate such phenomena by discussing simplified mathematical models where a rigorous analysis can be carried through, at least partially.

As the appearance of deterministic equations is related to the validity of a law of large numbers in the macroscopic limit, the presence of stochastic effects must be ascribed to a failure of such a law. It will clearly appear in the models I shall present, that this is due to some instabilities in the deterministic equations which amplify the small fluctuations intrinsically present when reducing to macroscopic variables. As we shall see spinodal decomposition is related to a linear instability, metastability, tunnelling and hysteresis to the existence of several local minimizers, interface fluctuations to the existence of neutral directions. The mathematical analysis of these phenomena is also very rich, non linear PDE's, variational problems, large deviations will in fact be extensively used.

RENORMALIZATION GROUP METHOD AND SINGULARITIES OF COMPLEX-VALUED SOLUTIONS OF EQUATIONS OF FLUID-DYNAMICS

YAKOV G. SINAI

Mathematics Department, Princeton University

E-mail: sinai@math.princeton.edu

I shall explain the method of constructing solutions with singularities developed by Dong Li and myself in the case of two-dimensional Burgers system.

RESONANT PERTURBATIONS OF HAMILTONIAN SYSTEMS IN INFINITE DIMENSIONS

WEI-MIN WANG

Mathématiques, Université Paris-Sud, Orsay

E-mail: wei-min.wang@math.u-psud.fr

We develop a resonant perturbation theory for Hamiltonian PDE's, which include both the linear and nonlinear Schrödinger equations. In this lecture, we focus on the linear theory, where we prove eigenfunction localization (in the Fourier space) for the 2D periodic Schrödinger operator on the square torus, solving a basic problem in spectral theory.

We will also give an indication of the construction of time quasi-periodic solutions for the nonlinear Schrödinger equations.

TOWARDS A QUANTUM CHURCH-TURING THEOREM

REINHARD F. WERNER

School of Mathematics, Institute for Mathematical Physics, TU Braunschweig

E-mail: r.werner@tu-bs.de

There are many theoretical models for quantum computation now, all of which appear to be equivalent in that they can simulate each other with polynomial overheads. In this talk I would like to address the question whether thereby we have already found the most general way of making quanta compute. In the traditional phrasing of the Church-Turing thesis: Can any reasonable quantum mechanical computation process be simulated efficiently by one of the standard quantum computational models? In a suitably restricted quantum theoretical setting this should be a provable statement, and I will give examples of such Theorems, particularly for lattice systems.

YRS SEMINAR TALKS

- MARTÍN PATRICIO ARCIGA ALEJANDRE (Instituto de Matemáticas, UNAM, Mexico), with ELENA KAIKINA: **OTT-SUDAN-OSTROVSKIY EQUATION ON A HALF- LINE**
- PETR AMBROŽ (Doppler Inst. & Dept of Mathematics, Czech Technical University): **ON THE NUMBER OF FACTORS IN CODINGS OF THREE INTERVAL EXCHANGE**
- ANDREI BAGROV (Steklov Mathematical Institute, Moscow), with IRINA AREFEVA and E.A. GUSEVA: **N-FIELD APPROACH TO BLACK HOLES PRODUCTION**
- L'UBOMÍRA BALKOVÁ (Department of Mathematics, Czech Technical University, Prague): **SCHRÖDINGER OPERATORS ASSOCIATED WITH PARRY NUMBERS**
- RAM BAND (Weizmann Institute, Rehovot), with GREGORY BERKOLAIKO and UZY SMILANSKY: **COUNTING NODAL DOMAINS ON QUANTUM GRAPHS**
- MIROSLAV BULÍČEK (Charles University, Prague), with PIOTR GWIAZDA, JOSEF MÁLEK, and AGNIESZKA SWIERCZEWSKA-GWIAZDA: **ON SCALAR HYPERBOLIC LAWS WITH DISCONTINUOUS FLUX**
- RAFAELLE CARLONE (Doppler Institute, Prague), with CLAUDIO CACCIAPUOTI, GIANFAUSTO DELL' ANTONIO, RODOLFO FIGARI, DOMENICO FINCO, and ALESSANDRO TETA: **THE MOTT'S CONJECTURE**
- BHAMIDIPATI CHANDRASEKHAR (Instituto de Física, Universidade de Sao Paulo), with PRASANTA K. PANIGRAHI and VIVEK M. VYAS: **UNITARY FERMION GAS: SCALING SYMMETRIES AND EXACT MAP**
- AURÉL GÁBRIS (Doppler Institute, Prague), with VÁCLAV POTOČEK, TAMÁS KISS, and IGOR JEX: **OPTIMIZING THE QUANTUM RANDOM WALK SEARCH ALGORITHM ON THE HYPERCUBE**
- GRAHAM KELLS (National University of Ireland, Maynooth), with JOOST K. SLINGERLAND and JIŘÍ VALA: **A DESCRIPTION OF KITAEV'S HONEYCOMB MODEL WITH TORIC-CODE STABILIZERS**
- JIŘÍ LIPOVSKÝ (Doppler Institute and Charles University, Prague), with PAVEL EXNER: **RESONANCES FROM PERTURBATIONS OF QUANTUM GRAPHS WITH RATIONALLY RELATED EDGES**
- TAKUYA MINE (Kyoto Institute of Technology): **AHARONOV-BOHM SOLENOIDS ON THE HYPERBOLIC PLANE**
- MARCELLA PALESE (University of Torino), with EKKEHART WINTERROTH: **NON-LINEAR OPTICAL MODELS AND ALGEBRAIC SKELETONS**
- THOMAS QUELLA (University of Amsterdam), with VLADIMIR MITEV and VOLKER SCHOMERUS: **CONFORMAL SUPERSPACE SIGMA-MODELS**
- JULIEN QUEVA (Université Diderot-Paris 7), with S. FACI, E. HUGUET, and J. RENAUD: **A CONFORMALLY COVARIANT QUANTIZATION OF THE MAXWELL FIELD IN DE SITTER SPACETIME**
- EVELINA SHAMAROVA (University of Porto), with ANA BELA CRUZEIRO: **NAVIER-STOKES EQUATIONS AND FORWARD-BACKWARD SDES ON THE GROUP OF DIFFEOMORPHISMS OF A TORUS**

VIACHESLAV O. SHTYK (Institute for Theoretical Physics, Kiyv): **ON MEAN-FIELD ASYMPTOTIC OF SOLUTION OF QUANTUM BBGKY HIERARCHY**

PETR SIEGL (Doppler Institute, Prague): **PT-SYMMETRIC MODELS IN CURVED MANIFOLDSE**

MARTIN ŠTEFANAŇÁK (Czech Technical University, Prague), with IGOR JEX and TAMÁS KISS: **RECURRENCES IN QUANTUM WALKS**

PAULINA SUCHANEK (Jagiellonian University, Krakow): **4-POINT CORRELATION FUNCTIONS IN 2 DIM N=1 SCFT**

MATEJ TUŠEK (Czech Technical University, Prague), with PIERRE DUCLOS and PAVEL ŠŤOVÍČEK: **ON A HYDROGEN ATOM IN A THIN SLAB**

EKKEHART WINTERROTH (Univ. of Torino), with MARCELLA PALESE: **INVARIANT VARIATIONAL PROBLEMS AND BUNDLES OF CARTAN CONNECTIONS**

Round Table

**“Mathematical Physics:
Future Challenges”**

Moderated by:

Joel Lebowitz

FAST AND SLOW

MICHAEL BERRY

Physics Department, University of Bristol

E-mail: tracie.anderson@bristol.ac.uk

In classical or quantum systems with time scales widely separated by a small slowness parameter, the fast system is driven by the slow and reacts back on it. The asymptotics of the driven fast motion has been much studied, but understanding the series of reaction forces in high orders of slowness is harder: only the lowest order (Born-Oppenheimer) and next order (geometric magnetism) have been fully explored. If the series of reactions converged, this would restrict the coupled motion to a slow manifold on which there are no fast oscillations. This hope is frustrated when, as is typical, the series diverges factorially. In a spinning-particle classical hamiltonian nonresonant model, where the asymptotics can be studied in analytical and numerical detail, the inevitability of exponentially weak fast oscillations arises from a Stokes phenomenon in the series of reaction forces.

HISTORY VERSUS FUNCTION: THE CASE OF PROTEINS

STANISLAS LEIBLER

The Rockefeller University and Institute for Advanced Study

E-mail: kmclean@mail.rockefeller.edu

Some of the concepts one encounters while studying systems under (natural) selection are proper to biology. An important example is the concept of function, fundamental for the description of biological systems. Similarly, evolution is largely an historical process and many similarities between different organisms originate in their common past. Therefore, one of the challenges for future mathematical modeling of biological systems is to develop a theoretical framework for the interplay between history and function. To illustrate this point, I will describe a recent piece of work on the structure of proteins, done in collaboration with N. Halabi, O. Rivoire and R. Ranganathan. By comparing hundreds of sequences of proteins belonging to the same family (of serine proteases), we were able to separate the functional and historical correlations between protein residues. This allowed us to uncover the functional structure of these proteins, consisting of several coexisting, quasi-independent, 3-d “sectors”, each associated with a different function. The detailed physical nature of the sectors needs to be elucidated.

FAST SCRAMBLERS

LEONARD SUSSKIND

Stanford University

E-mail: susskind@stanford.edu

I will consider the problem of how fast a quantum system can scramble (thermalize) information, given that the interactions are between bounded clusters of degrees of freedom; pairwise interactions would be an example. Based on previous work, we conjecture: 1) The most rapid scramblers take a time logarithmic in the number of degrees of freedom.

2) Matrix quantum mechanics (systems whose degrees of freedom are n by n matrices) saturate the bound. 3) Black holes are the fastest scramblers in nature. The conjectures are based on two sources, one from quantum information theory, and the other from the study of black holes in String Theory.

The above three opening talks were followed by a general discussion.

Human Rights and Social Responsibility

Moderated by:

Joel Lebowitz

MATHEMATICAL PHYSICS FROM AFRICA: STATUS AND PROGRESS

NORBERT M. HOUNKONNOU

International Chair in Mathematical Physics, 072 BP 50 Cotonou, Republic of Benin
E-mail: hounkonnou@yahoo.fr

This talk addresses the status of Mathematical Physics by people living and working in Africa and discusses the conditions of its advancement.

HUMAN RIGHTS AND SOCIAL RESPONSIBILITY OF SCIENTISTS

JOEL LEBOWITZ

Department of Mathematics, Rutgers University
E-mail: lebowitz@math.rutgers.edu

I will take as my theme an editorial which appeared in the 12 October 2006 issue of *Nature* under the title of “Forgotten Plights”. It begins with a quote from Martin Niemöller who was later murdered by the Nazis. “First they came for the Socialists, and I didn’t speak up, because I wasn’t a Socialist... Then they came for the Jews, and I didn’t speak up, because I wasn’t a Jew. Then they came for me, and there was no one left to speak up for me.” The editorial goes on to say: “Most readers of *Nature* take it for granted that they can travel to work each day, free to enquire, express opinions and criticize government policy, without fear of intimidation or reprisals - let alone imprisonment or torture. Sadly, these freedoms can only be dreamt of in many countries of the world, where academics must live with, and often suffer directly, human-rights abuses. Their plight is our business”. It concludes with: “All scientists can contribute, by making themselves aware of current cases of human-rights abuses and by lending their support to campaigns against them”.

I will describe briefly a few current cases of colleagues (interpreted broadly) whose denial of human rights is being protested by the Committee of Concerned Scientists and other organizations.

FREEDOM AND DEMOCRACY ARE NOT SELF-EVIDENT

ŠIMON PÁNEK

People in Need, Prague
E-mail: simon.panek@peopleinneed.cz

Freedom and Democracy — in the central Europe where this congress convenes we regard it as a dream came through twenty years ago and as a reason why we have to carry on in countries which still do not enjoy it and their people remain deprived of basic rights and freedoms.

Satellite Meetings

CONSTRUCTIVE AND MULTISCALE METHODS IN QUANTUM THEORY

HEIDELBERG UNIVERSITY, JULY 28-30, 2009

Organizers: Volker Bach, Christoph Kopper, Manfred Salmhofer

Webpage: <http://www.thphys.uni-heidelberg.de/~salmhof/satellite09.html>

ALGEBRAIC QFT – THE FIRST 50 YEARS

GÖTTINGEN UNIVERSITY, JULY 29-31, 2009

Organizers: Dorothea Bahns, Detlev Buchholz, Pierre Martinetti, Karl-Henning Rehren

Webpage: <http://www.uni-math.gwdg.de/aqft/>

MATHEMATICAL ASPECTS OF QUANTUM TRANSPORT AND APPLICATIONS IN NANOPHYSICS

AALBORG UNIVERSITY, AUGUST 9-14, 2009

Organizer: Horia Cornean

Webpage: http://people.math.aau.dk/~cornean/ICMP09_satellite/

14TH SUMMER SCHOOL IN GLOBAL ANALYSIS AND MATHEMATICAL PHYSICS

PALACKÝ UNIVERSITY OLOMOUC, AUGUST 10-14, 2009

Organizers: Demeter Krupka, Olga Krupková, David J. Saunders

Webpage: <http://globalanal.upol.cz/summer2009/index.html>

XVII INTERNATIONAL FALL WORKSHOP IN GEOMETRY AND PHYSICS

BENASQUE, SEPTEMBER 6-10, 2009

Organizers: J.F. Cariñena, E. Martínez, J. Clemente-Gallardo,

J.N. da Costa, D. Martín de Diego

Webpage: <http://benasque.ecm.ub.es/2009gph/>

LIST OF REGISTERED PARTICIPANTS

Abdesselam A.	University of Virginia, Charlottesville VA, United States
Adamová D.	Nuclear Physics Institute ASCR, Řež, Czech Republic
Adams S.	University of Warwick, Coventry, United Kingdom
Aharonov D.	Hebrew University, Jerusalem, Israel
Aizenman M.	Princeton University, United States
Ajanki O.	University of Helsinki, Finland
Albeverio S.	Bonn Universitaet, Bonn, Germany
Aleksenko A.	University of Aveiro, Portugal
Alexander K.	University of Southern California, Los Angeles, CA, United States
Alicki R.	Gdansk, Poland
Ambrož P.	FNSPE, Czech Technical University in Prague, Czech Republic
Amornwichet S.	Thammasat University, Pathumthani, Thailand
Andrié M.	Bonn, Germany
Araki H.	Kyoto, Japan
Arciga Alejandre M. P.	Universidad Michoacana de San Nicolas de Hidalgo, Mexico
Argerami M.	University of Regina, Canada
Aschbacher W.	Garching, Germany
Avramidi I.	New Mexico Tech, Socorro, NM, United States
Avron J.	technion, Haifa, Israel
Bach V.	University of Mainz, Germany
Bachmann S.	ETHZ, Zuerich, Switzerland
Bagrov A.	Steklov Mathematical Institute, Moscow, Russia
Baker M.	University of Oxford, United Kingdom
Balint P.	Budapest University of Technology and Economics, Hungary
Balkova L.	FNSPE, Czech Technical University in Prague, Czech Republic
Ballesteros M.	INRIA, Paris, France and UNAM, Mexico
Bambah B. A.	University of Hyderabad, Hyderabad, India
Band R.	Weizmann Institute of Science, Rehovot, Israel
Barata J.	Universidade de Sao Paulo, Brazil
Baumgartner B.	University of Vienna, Austria
Bazhanov V.	Australian National University, Canberra ACT, Australia
Beau M.	Université Aix-Marseille II, Marseille, France
Beceanu M.	EHESS, Iasi, Romania
Bednář J.	Univerzita Karlova v Praze, Praha, Czech Republic
Beiglboeck W.	Springer, Heidelberg, Germany
Belyakov N.	Belfort, France
Benguria R. D.	Pontificia Universidad Católica de Chile, Santiago, Chile
Bernabei M. S.	University of Camerino, Camerino, Italy
Berry M.	Bristol, United Kingdom
Betz V.	University of Warwick, Coventry, United Kingdom
Bhamidipati C.	Instituto de Fisica Teorica, Sao Paulo, Brazil
Bičák J.	Charles University in Prague, Czech Republic
Bila H.	ÚJF AV ČR, Praha, Czech Republic
Biskup M.	UCLA & JCU, Los Angeles, United States
Blinova I.	St.-Petersburg State Univ. of Information Technologies, Russia
Bonacci E.	L.S.S. "G.B. Grassi", Latina, Italy
Borisov D.	Bashkir State Pedagogical University, Ufa, Russia
Borodin A.	California Institute of Technology, Pasadena, United States
Borsuk M.	University of Warmia and Mazury in Olsztyn, Poland
Borthwick K.	Springer, Heidelberg, Germany
Bossard G.	Max-Planck-Institut, Golm, Germany
Bostelmann H.	University of York, United Kingdom
Boya L. J.	Universidad de Zaragoza, Spain
Brandon L.	Springer, Heidelberg, Germany
Brasche J.	TU Clausthal, Clausthal-Zellerfeld, Germany
Bregman Y.	Soreq NRC, Yavne, Israel

- Brougham T.
 Bruzon M.
 Brydges D.
 Buchholz D.
 Bulíček M.
 Burgstaller B.
 Burke B.
 Buscemi F.
 Byrne C.
 Cacciapuoti C.
 Cachafeiro M. A.
 Cadamuro D.
 Camassa P.
 Campbell C.
 Capelin S.
 Carbone R.
 Casado-Díaz J.
 Cassar M.
 Ceballos M.
 Cech M.
 Cegla W.
 Cellarosi F.
 Chadzitaskos G.
 Cheng C.
 Chirou V.
 Choi Q. H.
 Choquard P.
 Christodoulou D.
 Chruscinski D.
 Cieplý A.
 Clark J.
 Cornean H.
 Cortés E.
 Corwin I.
 Costa E.
 Crawford N.
 Cruzeiro A. B.
 Dafermos M.
 Damianou P.
 Damron M.
 Danielsen T. H.
 Dappiaggi C.
 Datta A.
 Datta N.
 De Lellis C.
 De Masi A.
 De Melo W.
 De Nittis G.
 De Roeck W.
 de Siqueira Pedra W.
 Delfino G.
 Demirel S.
 Dereziński J.
 Díaz D.
 Dimock J.
 Disertori M.
 Dito G.
 Dittrich J.
 Dobeš J.
 Dobrev V.
 Dorey P.
 Downard T.
 Downes J.
 Dragomirescu F. I.
 Dudnikova T.
 Dumitracu A.
 Duncan M.
 Dunlop F.
 Duplantier B.
 Czech Technical University in Prague, Czech Republic
 University of Cadiz, Puerto Real, Cadiz, Spain
 University of British Columbia, Vancouver, Canada
 Georg-August Universität Göttingen, Germany
 Charles University, Prague, Czech Republic
 Doppler Institute, FNSPE CTU, Prague, Czech Republic
 American Institute of Physics, Melville, NY, United States
 University of Cambridge, United Kingdom
 Springer, Heidelberg, Germany
 Hausdorff Center for Mathematics, Bonn, Germany
 Universidad de Vigo, Spain
 University of Goettingen, Germany
 II University of Rome "Tor Vergata", Milan, Italy
 University of Queensland, Brisbane, Australia
 Cambridge University Press, Cambridge, United Kingdom
 Doppler Institute, Rez near Prague, Czech Republic
 University of Seville, Spain
 American Institute of Physics, Melville, NY, United States
 Universidad de Sevilla, Spain
 ČVUT, FJFI, Praha, Czech Republic
 University of Wrocław, Poland
 Princeton University, United States
 FNSPE Czech Technical University in Prague, Czech Republic
 Nanjing University, China
 Institute of Solid Mechanics of Romanian Academy, Bucharest
 Inha University, Incheon, South Korea
 EPFL, Lausanne, Switzerland
 ETH-Zurich, Switzerland
 Nicolaus Copernicus University, Torun, Poland
 Nuclear Physics Institute ASCR, Řež, Czech Republic
 K.U. Leuven, Heverlee, Belgium
 Aalborg University, Denmark
 Universidad Autónoma Metropolitana, Mexico
 Courant Institute, New York, USA
 SISSA/ISAS, Trieste, Italy
 The Technion, Essex, CT, USA
 IST, Lisboa, Portugal
 University of Cambridge, United Kingdom
 University of Cyprus, Nicosia, Cyprus
 Princeton University, USA
 University of Copenhagen, Karlslunde, Denmark
 Hamburg Universität, Hamburg, Germany
 Imperial College London, United Kingdom
 University of Cambridge, United Kingdom
 University of Zuerich, Switzerland
 University of L'Aquila, Italy
 Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil
 SISSA, Trieste, Italy
 ETH Zurich- KULeuven, Antwerpen, Belgium
 University of Mainz, Germany
 SISSA, Trieste, Italy
 University of Stuttgart, Germany
 University of Warsaw, Poland
 Universidad Andres Bello, Santiago de Chile, Chile
 SUNY at Buffalo, NY, United States
 Université de Rouen, Saint Etienne du Rouvray, France
 Université de Bourgogne, Dijon cedex, France
 Nuclear Physics Institute ASCR, Rez, Czech Republic
 Ústav jaderné fyziky AV ČR, Řež, Czech Republic
 Bulgarian Academy of Sciences, Bulgaria
 Durham University, UK
 University of Illinois at Springfield, Weselberg, Germany
 Technische Universität Berlin, Germany
 University "Politehnica" of Timisoara, Romania
 Elektrostal Polytechnical Institute, Elektrostal, Russia
 Inst. of Math. Statistics and Appl. Math., Bucharest, Romania
 Oxford University, Oxford, United Kingdom
 Université de Cergy-Pontoise, France
 CEA, Gif-sur-Yvette Cedex, France

- Dybalski W. Technische Universitaet Muenchen, Germany
 Eckle H. University of Ulm, Germany
 Egli D. ETH Zuerich, Switzerland
 El-Said El-Nadi K. Faculty of Science, Alexandria University, Egypt
 Elon Y. Weizmann Institute of Science, Rehovot, Israel
 Elton D. Lancaster University, Lancaster, United Kingdom
 Enciso A. ETH Zurich, Switzerland
 Endres S. Universität Ulm, Germany
 Eppelbaum L. Tel Aviv University, Israel
 Erdos L. Ludwig-Maximilian University Munich, Germany
 Esteban M. J. CNRS & University Paris-Dauphine, France
 Exner P. Nuclear Physics Institute, ASCR, Řež, Czech Republic
 Faci S. APC, Université de Paris Diderot, France
 Family F. Emory University, Atlanta, USA
 Fedorchuk V. Inst. of Mathematics, Pedagogical University, Krakow, Poland
 Finch P. University of Queensland, Brisbane, QLD, Australia
 Finco D. "La Sapienza" University, Roma, Italy
 Fisher M. E. University of Maryland, United States
 Fleischhack C. Hamburg University, Hamburg, Germany
 Fournais S. University of Aarhus, Denmark
 Fraas M. Technion, Haifa, Israel
 Frank R. Princeton University, United States
 Fredenhagen K. University of Hamburg, Germany
 Freiberg U. University of Hamburg, Germany
 Freiji A. University of Alabama at Birmingham, USA
 Freire H. Universidade Federal de Goiás, Goiânia - GO, Brazil
 Freixas J. Technical University of Catalonia, Manresa, Spain
 Froehlich J. ETH Zurich, Switzerland
 Frunzulica F. Univesity Politechnica of Bucharest, Romania
 Fuentes R. Santiago, Chile
 Fujimoto M. Nara Medical University, Kashihara, Nara, Japan
 Fukuda M. University of California, Davis, USA
 Fulling S. Texas A&M University, College Station, United States
 Furuichi S. Nihon Univ., College of Humanities and Sciences, Tokyo, Japan
 Gábris A. Czech Technical University in Prague, Czech Republic
 Gandarias M. L. University of Cadiz, Puerto Real, Spain
 Garay Ó. J. University of the Basque Country, Bilbao, Spain
 Garcilazo Botello E. UNAM, Paris, France
 Gazda D. Nuclear Physics Institute ASCR, Řež, Czech Republic
 Geisinger L. Universität Stuttgart, Germany
 Gelfand S. American Mathematical Society, Providence, RI, United States
 Giancesello C. Centre de Physique théorique, Marseille, France
 Gianfreda M. University of Salento, Lecce, Italy
 Gilka N. University of Copenhagen, Denmark
 Gillan R. IOP Publishing, Bristol, United Kingdom
 Giuliani A. Univ. di Roma Tre, Roma, Italy
 Golenia S. Universitaet Erlangen-Nuernberg, Erlangen, Germany
 Golse F. Ecole polytechnique, Palaiseau cedex, France
 Gonçalves Silveira de Serpa M. C. Faculty of Sciences, University of Lisbon, Portugal
 Graf G. M. ETH Zurich, Switzerland
 Grech P. ETH Zurich, Switzerland
 Griesemer M. Universitaet Stuttgart, Germany
 Grossman P. Vanderbilt University, Nashville, TN, USA
 Gruber M. J. TU Clausthal, Clausthal-Zellerfeld, Germany
 Grummt R. Ludwig-Maxmilians University, Munich, Germany
 Guenther U. Research Center Dresden-Rossendorf, Dresden, Germany
 Guido D. Univ. Roma Tor Vergata, Roma, Italy
 Guionnet A. CNRS, Lyon, France
 Hadjiivanov L. Inst. for Nuclear Research and Nuclear Energy, Sofia, Bulgaria
 Hammond A. New York University, NY, United States
 Hampl V. Universita Karlova v Praze, Praha, Czech Republic
 Hamza E. Michigan State University, East Lansing, USA
 Hansen F. Copenhagen University, Denmark
 Hantsch F. Universität Stuttgart, Germany
 Harrell E. Georgia Institute of Technology, Atlanta, GA, United States
 Harrow A. University of Bristol, United Kingdom
 Hasler D. College of William & Mary, Williamsburg, VA, United States
 Hatzinikitas A. University of Aegean, Karlovasi, Samos, Greece

- Havel I.
 Havlíček M.
 Havlíček V.
 Hejcik P.
 Hempel R.
 Hempfling T.
 Henrici A.
 Hepp K.
 Herbst I.
 Herdegen A.
 Hintermann T.
 Hirokawa M.
 Hlavatý L.
 Hofer H.
 Holmes P.
 Hořejší J.
 Hošek J.
 Hounkonnou M. N.
 Hrdlicka F.
 Ianiro N.
 Ichinose T.
 Ichinose W.
 Imaykin V.
- Ionescu A.
 Ito K.
 Ivanshin P.
 Izumi M.
 Jadczyk A.
 Jaffe A.
 Jansen S.
 Jaskolski Z.
 Jellal A.
 Jensen A.
 Jex I.
 Joita M.
 Joye A.
 Julia B.
 Jung T.
 Kaiser G.
 Kaloshin V.
 Kalvoda T.
 Kang Y.
 Karaliolios N.
 Kawahigashi Y.
 Kells G.
 Keskin Y.
 Khatiazhvili N.
 Khetselius O.
 Khorunzhiy O.
 Kirsch W.
 Kiyohara K.
 Klein A.
 Klich I.
 Klouda K.
 Knowles A.
 Koc A. B.
 Kochubei A.
 Koh S.-E.
 Kolesár M.
 Komech A.
 Komech S.
 Kopsky G. K.
 Kopylova E.
 Korbelař M.
 Korchemsky G.
 Kosmann-Schwarzbach Y.
 Kosower D. A.
 Kostenko A.
- CTS, Prague, Czech Republic
 FNSPE, Czech Technical University in Prague, Czech Republic
 Czech Technical University in Prague, Czech Republic
 University of Hradec Kralove, Czech Republic
 TU Braunschweig, Germany
 Birkhäuser, Basel, Switzerland
 Universität Zürich, Switzerland
 ETH Zürich, Switzerland
 University of Virginia, Charlottesville, United States
 Jagiellonian University, Kraków, Poland
 EMS Publishing House, Zürich, Switzerland
 Okayama University, Japan
 Czech Technical University in Prague, Czech Republic
 New York University, NY, United States
 Princeton University, United States
 Charles University, Prague, Czech Republic
 Nuclear Physics Institute, Rez, Czech Republic
 University of Abomey-Calavi, Cotonou, Benin
 Czech Technical University in Prague, Czech Republic
 Università di Roma "La Sapienza", Italy
 Kanazawa University, Kanazawa, Japan
 Shinshu University, Matsumoto, Japan
 Scientific Research Institute of Innovative Strategies in the General Education
 Development, Moscow, Russia
 University of Wisconsin - Madison, Madison, WI, United States
 Setsunan University, Neyagawa, Japan
 Kazan State University, Kazan, Russia
 Kyoto University, Kyoto, Japan
 University Paul Sabatier, Toulouse, France
 Harvard University, Cambridge, USA
 Princeton University, United States
 Wroclaw University, Poland
 Chouaib Doukkali University, El Jadida, Morocco
 Aalborg University, Denmark
 Czech technical University in Prague, Czech Republic
 University of Bucharest, Romania
 Université de Grenoble 1, Saint-Martin d'Heres, France
 LPTENS, Paris, France
 Kunsan National University, Kunsan, South Korea
 University of Massachusetts (Emeritus), Austin, United States
 Penn State University, State College, United States
 Czech Technical University in Prague, Czech Republic
 Michigan State University, Lansing, United States
 Université Paris 6, France
 University of Tokyo, Japan
 National University of Ireland, Co. Kildare, Ireland
 Science Faculty, Selcuk University, Konya, Turkey
 Tbilisi State University, Tbilisi, Georgia
 Inst. for Applied Mathematics, Odessa University, Ukraine
 Université de Versailles, France
 FernUniversität Hagen, Germany
 Okayama University, Okayama-City, Japan
 University of California - Irvine, Irvine, United States
 University of Virginia, Charlottesville, United States
 FNSPE, Czech Technical University in Prague, Czech Republic
 ETH Zurich, Switzerland
 Science Faculty, Selcuk University, Konya, Turkey
 National Academy of Sciences of Ukraine, Kiev, Ukraine
 Konkuk University, Seoul, South Korea
 Charles University, Praha, Czech Republic
 IITP RAN and Vienna University, Mytishy, Russia
 IITP, Moscow, Russia
 University of Vienna, Vienna, Austria
 IITP Russian Academy of Sciences, Vladimir, Russia
 Charles University, Prague, Czech Republic
 Institut de Physique Théorique, Saclay, Gif-sur-Yvette, France
 Ecole Polytechnique, Palaiseau, France
 CEA-Saclay, Gif-sur-Yvette, France
 Dublin Institute of Technology, Dublin, Ireland

- Kotani M.
 Kotecký R.
 Koukiou F.
 Kovařík H.
 Kozhan R.
 Kozma G.
 Krbalek M.
 Kreimer D.
 Krejčířík D.
 Kreml O.
 Kríž J.
 Krueger H.
 Krupka D.
 Krupkova O.
 Kryzhevich S.
 Kuandykov L.
 Kubiznak D.
 Kuijlaars A.
 Kuksin S.
 Kupiainen A.
 Łacki M.
 Lakshtanov E.
 Lanchon P.
 Langlois D.
 Langmann E.
 Laptev A.
 Larenas M.
 Lauridsen Ribeiro P.
 Lavička H.
 Lazar A. J.
 Lebowitz J.
 Lechner G.
 Leibler S.
 Lein M.
 Lenz D.
 Lewin M.
 Liimatainen T.
 Ling S.
 Lipert M.
 Lipovský J.
 Liu Y.
 Loebach D.
 Longo R.
 Losert-Valiente
 Kroon C. M.
 Loss M.
 Loudon A.
 Lubetzky E.
 Lukkarinen J.
 Luna-Laynez M.
 Lyberg I.
 Macek M.
 Maestre Caballero F.
 Magnitskiy N.
 Maiocchi A.
 Majewski W. A.
 Malamud M.
 Málek J.
 Maltsev A.
 Mareš J.
 Mareš J.
 Mares Gallardo R.
 Marklof J.
 Martín-Gómez J. D.
 Martinetti P.
 Martínez E.
 Masáková Z.
 Mastropietro V.
 Matsui T.
 Tohoku University, Sendai, Miyagi, Japan
 Charles University, Prague, Czech Republic
 University of Cergy-Pontoise, France
 Politecnico di Torino, Italy
 California Institute of Technology, Pasadena, United States
 Weizmann Institute of Science, Tel Aviv, Israel
 FNSPE, Czech Technical University in Prague, Czech Republic
 IHES, Bures sur Yvette, France
 Nuclear Physics Institute, ASCR, Rez, Czech Republic
 Charles University, Prague, Czech Republic
 University of Hradec Kralove, Czech Republic
 Rice University, Houston, USA
 Palacky University, Olomouc, Czech Republic
 Palacky University, Olomouc, Czech Republic
 Saint-Petersburg State University, Russia
 Corning Incorporated, Saint-Petersburg, Russia
 University of Cambridge, United Kingdom
 Katholieke Universiteit Leuven, Belgium
 CNRS-Ecole Polytechnique, Palaiseau, France
 Helsinki University, Finland
 Jagiellonian University, Krakow, Poland
 University of Aveiro, Portugal
 Snecma, Safran Group, Palaiseau, France
 Université Paris 7, France
 KTH, Stockholm, Sweden
 Imperial College London, United Kingdom
 Universidad de Chile, Santiago, Chile
 Universität Hamburg, Germany
 FNSPE Czech Technical University in Prague, Czech Republic
 Tel Aviv University, Israel
 Rutgers University, Piscataway, New Jersey, United States
 University of Vienna, Wien, Austria
 The Rockefeller University; Institute for Advanced Study, US
 Technische Universität München, Garching, Germany
 Friedrich Schiller Universitaet, Jena, Germany
 CNRS / University of Cergy-Pontoise, France
 Helsinki University of Technology, Espoo, Finland
 HKCCCU Logos Academy, Hong Kong
 Jagiellonian University, Krakow, Poland
 Nuclear Physics Institute, ASCR, Rez, Czech Republic
 Harbin Engineering University, Harbin, China
 University Bonn, Germany
 Università di Roma Tor Vergata, Roma, Italy
 UCL, London, United Kingdom
 Georgia Institute of Technology, Atlanta, United States
 American Institute of Physics, Melville, NY, United States
 Microsoft Research, Redmond, United States
 University of Helsinki, Helsingin Yliopisto, Finland
 University of Seville, Spain
 DIAS, Dublin, Ireland
 Charles University, Prague, Czech Republic
 Universidad de Sevilla, Spain
 Institute for Systems Analysis of RAS, Moscow, Russia
 Università degli Studi di Milano, Italy
 University of Gdansk, Poland
 Institute Applied Mathematics and Mechanics, Donetsk, Ukraine
 Charles University in Prague, Czech Republic
 CalTech, Pasadena, United States
 FJFI ČVUT, Prague, Czech Republic
 Nuclear Physics Institute ASCR, Rez, Czech Republic
 Instituto Politécnico Nacional, Mexico
 University of Bristol, United Kingdom
 University of Seville, Spain
 Georg-August Universität, Göttingen, Germany
 University of Zaragoza, Spain
 FNSPE, Czech Technical University in Prague, Czech Republic
 University of Rome "Tor Vergata", Rome, Italy
 Kyushu University, Fukuoka, Japan

- Matsuyama Y.
 Matveev V.
 Mei P.
 Merzon A.
 Michelangeli A.
 Michor J.
 Miermont G.
 Mine T.
 Miracle-Sole S.
 Missarov M.
 Moiseyenko R.
 Mol L.
 Monastyrsky M.
 Morfa Morales E.
 Moser D.
 Mosonyi M.
 Mostafazadeh A.
 Motl L.
 Moylan P.
 Mueger M.
 Mueller P.
 Mühlhoff R.
 Mukku C.
 Müller M.
 Muriel C.
 Musemic R.
 Myroniuk L.
 Naaijkens P.
 Nachtergaele B.
 Nagamachi S.
 Nagao T.
 Narita M.
 Naudts J.
 Navarro J. F.
 Neidhardt H.
 Neshveyev S.
 Nest R.
 Ng S.
 Nichols R.
 Nikolaev P.
 Nikolov N.
 Nordenstam E.
 Novak J.
 Novikov S.
 Novotný P.
 Núñez J.
 O'Connell N.
 Obitsu K.
 Ogata Y. O.
 Ohya S.
 Olf U.
 Oliffson Kamphorst S.
 Olkiewicz R.
 Olla S.
 Opoku A. A.
 Ordóñez-Flores B.
 Oren I.
 Oztas Z.
 Paleari S.
 Palese Winterroth M.
 Palouš J.
 Panati A.
 Pánek Š.
 Pankrashkin K.
 Pansuwan A.
 Paranjape M.
 Paszkiewicz A.
 Pauly D.
 Pelantová E.
 Chuo University, Japan, Tokyo, Japan
 University of Burgundia, Dijon, France
 University of Helsinki, Finland
 Universidad Michoacana de S.Nicolas de Hidalgo, Mexico
 LMU, Munich, Germany
 University of Vienna, Austria
 CNRS and Ecole Normale Supérieure, Paris, France
 Kyoto Institute of Technology, Kyoto, Japan
 Centre de Physique Theorique, Marseille, France
 Kazan State University, Kazan, Russia
 Donetsk national university, Donetsk, Ukraine
 Springer, Heidelberg, Germany
 Inst. for Theoretical and Experimental Physics, Moscow, Russia
 University of Vienna, Austria
 Northeastern University, Boston, USA
 Centre for Quantum Technologies, NUS, Singapore
 Koc University, Istanbul, Turkey
 Charles University, Prague, Czech Republic
 Pennsylvania State University, Abington, United States
 Radboud University Nijmegen, Netherlands
 University of Munich, Germany
 Berlin Mathematical School, Germany
 Internat. Institute of Information Technology, Hyderabad, India
 ICTP Trieste, Italy
 Universidad de Cadiz, Spain
 University of Sarajevo, Bosnia-Herzegovina
 Lesya Ukrayinka Volyn National University, Lutsk, Ukraine
 Radboud Universiteit, Nijmegen, Netherlands
 UC Davis, CA, USA,
 The University of Tokushima, Fac. Eng., Tokushima, Japan
 Nagoya University, Japan
 Okinawa National College of Technology, Nago, Japan
 Universiteit Antwerpen, Belgium
 University of Alicante, San Vicente del Raspeig, Spain
 WIAS Berlin, Germany
 University of Oslo, Norway
 Copenhagen University, Denmark
 UC Davis, United States
 University of Alabama at Birmingham, USA
 Saint-Petersburg State University, Saint-Petersburg, Russia
 Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria
 Université Catholique de Louvain-La-Neuve, Belgium
 Charles university in Prague, Czech Republic
 University of Maryland-College Park, Hyattsville, United States
 FNSPE Czech Technical University in Prague, Czech Republic
 Universidad de Sevilla, Spain
 University of Warwick, Coventry, United Kingdom
 Kagoshima University, Kagoshima, Japan
 Tokyo, Japan
 Kobe University, Kobe, Japan
 Springer, Heidelberg, Germany
 Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
 Wroclaw University, Poland
 Université Paris Dauphine, Paris, France
 University of Groningen, Netherlands
 University of Seville, Spain
 Weizmann Institute of Science, Rehovot, Israel
 Anadolu Univeristy, Eskişehir, Turkey
 Università di Milano, Italy
 University of Torino, Italy
 Rada pro zahraniční styky AV ČR, Praha, Czech Republic
 Université Toulon-CPT Luminy, Paris, France
 People in Need, Prague, Czech Republic
 University Paris Sud XI, Orsay, France
 Thammasat University, Pathumthani, Thailand
 Université de Montréal, Canada
 Łódź University, Poland
 University of Jyväskylä, Finland
 FNSPE Czech Technical University in Prague, Czech Republic

- Peña J. J.
 Pensupha L.
 Pereira E.
 Peřinová V.
 Persson M.
 Pete G.
 Petit J.
 Petz D.
 Phan T. N.
 Pinchover Y.
 Pittman-Polletta B.
 Pizzo A.
 Plenio M.
 Pliukhn O.
 Poghosyan V.
 Pogosyan G.
 Pokorný M.
 Polishchuk D.
 Popov I. Y.
 Porrmann M.
 Porta M.
 Post O.
 Pošta S.
 Postolica V.
 Prešnajder P.
 Procacci A.
 Procházka M.
 Prykarpatsky A.
 Puente M. A.
 Purice R.
 Pusztai G.
 Quastel J.
 Quella T.
 Queva J.
 Radford C.
 Raftopoulos D.
 Rasmussen M. G.
 Rastelli G.
 Raussendorf R.
 Raz H.
 Reagor M.
 Rebenko A.
 Reintjes M.
 Rejzner K.
 Renner R.
 Rezouzenko A.
 Ricca R. L.
 Roberts J.
 Rodrigues M. M.
 Rodriguez-Ceballos J. A.
 Rojas-Molina C.
 Rothlauf S.
 Rowe D. J.
 Rubio Ponce A.
 Ruder R.
 Rudolph G.
 Runkel I.
 Rutkevich S.
 Saifullah K.
 Sako A.
 Salmhofer M.
 Samtleben H.
 Samuelsson P.
 Sanchez Morgado H.
 Sanders J.
 Santos L.
 Savoie B.
 Schaposnik L.
 Schilling C.
- Universidad Autonoma Metropolitana, Mexico
 Thammasat University, Pathumthani, Thailand
 Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
 Univerzita Palackého, Olomouc, Czech Republic
 Aarhus C, Denmark
 University of Toronto, Toronto, Canada
 Bruxelles, Belgium
 Renyi Institute of Mathematics, Budapest, Hungary
 University of Copenhagen, Denmark
 Technion - Israel Institute of Technology, Haifa, Israel
 University of Arizona, Tucson, USA
 University of California Davis, United States
 Imperial College London, United Kingdom
 Kyiv, Ukraine
 Joint Institute for Nuclear Research, Dubna, Russia
 Yerevan State University, Yerevan, Armenia
 Charles University, Prague, Czech Republic
 Taras Shevchenko National University of Kyiv, Ukraine
 St.-Petersburg State University of Inform. Technologies, Russia
 University of KwaZulu-Natal, Durban, South Africa
 University of Rome "Sapienza", Roma, Italy
 HU Berlin, Germany
 FNSPE, Czech Technical University in Prague, Czech Republic
 Romanian Academy of Scientists, Bacak University, Romania
 Faculty of Mathematics, Physics and Informatics, Bratislava, Slovakia
 Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
 Charles University, Prague, Czech Republic
 The AGH University of Science and Technology, Krakow, Poland
 Technical University of Catalonia, Manresa, Spain
 "Simion Stoilow" Institute of Mathematics, Bucharest, Romania
 Bolyai Institute, University of Szeged, Hungary
 University of Toronto, ON, Canada
 University of Amsterdam, Netherlands
 Laboratoire APC, PARIS, France
 Memorial University, St John's, Canada
 NTUA, Anavissos, Greece
 University of Aarhus, Denmark
 University of Torino, Lignana (VC), Italy
 University of British Columbia, Vancouver, Canada
 University of California Davis, Del Mar, United States
 University of Notre Dame, Valrico, FL, USA
 Institute of Mathematics, Kiev, Ukraine
 University of California Davis, Dresden, Germany
 Jagiellonian University in Cracow, Poland
 Institute for Theoretical Physics, Zurich, Switzerland
 Kharkov University, Ukraine
 U. Milano-Bicocca, Milano, Italy
 Universita' di Roma Tor Vergata, Roma, Italy
 University of Aveiro, Portugal
 Instituto Tecnológico de Morelia, Morelia, Mexico
 Université de Cergy-Pontoise, France
 University Erlangen-Nürnberg, Germany
 University of Toronto, Canada
 Universidad Autónoma Metropolitana, Mexico
 FAU Erlangen-Nürnberg, Germany
 University of Leipzig, Germany
 King's College London, United Kingdom
 Minsk, Belarus
 Quaid-i-Azam University, Islamabad, Pakistan
 Kushihiro National College of Technology, Kushihiro-city, Japan
 Heidelberg University, Germany
 Ecole Normale Supérieure de Lyon, France
 Lund University, Sweden
 Universidad Nacional Autónoma de México, Mexico
 University of Goettingen, Germany
 Yeshiva University, New York, United States
 Marseille, France
 Oxford University, United Kingdom
 Zurich, Switzerland

- Schlein B.
 Schnell K.
 Schoutens K.
 Schubert C.
 Schumacher C.
 Schwaibold T.
 Šeba P.
 Seel A.
 Seiler R.
 Seiringer R.
 Selberg S.
 Sen R. N.
 Serfaty S.
 Sergyeyev A.
 Serov V.
 Sewell G.
 Sezer Y.
 Shamarova E.
 Shamis M.
 Shatashvili S.
 Shelkovich V.
 Sherratt J.
 Shimada H.
 Shtyk V.
 Siedentop H.
 Siegl P.
 Silenko A.
 Simak V.
 Simonella S.
 Sims R.
 Sinai Y. G.
 Sirko L.
 Skopina M.
 Slavík P.
 Smedbäck M.
 Smirnov S.
 Smirnova Nagnibeda T.
 Smith L.
 Smith S.
 Snobl L.
 Sodin S.
 Solovej J. P.
 Sorrell I.
 Soshnikov A.
 Sparks J.
 Spicka V.
 Spitzer W.
 Spohn H.
 Starr S.
 Štefanaák M.
 Stephens M.
 Sterbenz J.
 Sternheimer D.
 Stiller M.
 Stoll I.
 Stopa M.
 Štoviček P.
 Streater R.
 Suchanek P.
 Sukhanov A.
 Susskind L.
 Suto A.
 Svitek O.
 Szablikowski B.
 Szabo Z. I.
 Szasz D.
 Szybka S.
 Ta Ngoc T.
 Takesaki M.
- University of Cambridge, United Kingdom
 ETH Zurich, ETH Honggerberg, Zurich, Switzerland
 University of Amsterdam, Netherlands
 Chemnitz University of Technology, Germany
 Universitaet Erlangen-Nuernberg, Germany
 Springer, Heidelberg, Germany
 University of Hradec Kralove, Czech Republic
 Universitaet Hannover, Germany
 Technische Universität Berlin, Germany
 Princeton University, United States
 Norwegean Univ. of Science and Technology, Trondheim, Norway
 Ben-Gurion University, Beer-Sheva, Israel
 Universite Paris 6 and Courant Institute, Paris, France
 Silesian University in Opava, Czech Republic
 University of Oulu, Finland
 Queen Mary University of London, United Kingdom
 Institute of Science and Technology, Istanbul, Turkey
 University of Lisbon, Portugal
 Hebrew University, Bat-Yam, Israel
 Trinity College Dublin and IHES, Dublin, Ireland
 St.Petersburg State University, St.Petersburg, Russia
 Heriot-Watt University, Edinburgh, United Kingdom
 University of Tokyo, Japan
 Bogolyubov Institute for Theoretical Physics, Kiev, Ukraine
 Ludwig-Maximilians-Universität München, Germany
 Doppler Institute, Řež, Czech Republic
 Belarusian State University, Minsk, Belarus
 FJFI, ČVUT, Prague, Czech Republic
 Università degli studi di Roma "La Sapienza", Rome, Italy
 University of Arizona, Tucson, USA
 Princeton University, United States
 Institute of Physics, Polish Academy of Sciences, Warsaw, Poland
 St. Petersburg State University, St. Petersburg, Russia
 Fakulta elektrotechnická ČVUT, Praha, Czech Republic
 Princeton University, USA
 Universite de Geneve, Switzerland
 Geneva University, Switzerland
 IOP Publishing, Bristol, United Kingdom
 Tufts University, Cambridge, United States
 Czech Technical University in Prague, Czech Republic
 Tel Aviv University, Israel
 University of Copenhagen, Denmark
 Lancaster University, Lancaster, United Kingdom
 University of California at Davis, California, United States
 University of Oxford, United Kingdom
 Institute of Physics, Praha, Czech Republic
 Universitaet Erlangen-Nuernberg, Germany
 TU Muenchen, Garching, Germany
 University of Rochester, United States
 FNSPE Czech Technical University in Prague, Czech Republic
 IOP Publishing, Bristol, United Kingdom
 University of California, San Diego, La Jolla, CA, United States
 Université de Bourgogne, Dijon, France
 University Hamburg, Germany
 Czech Technical University in Prague, Czech Republic
 Jagiellonian University, Tarnów, Poland
 Czech Technical University in Prague, Czech Republic
 King's College London, United Kingdom
 Jagiellonian University, Krakow, Poland
 Joint Institute for Nuclear Research, Moscow, Russia
 Stanford University, Palo Alto CA, United States
 Hungarian Academy of Sciences, Budapest, Hungary
 Charles University in Prague, Czech Republic
 University of Glasgow, United Kingdom
 City University of New York, USA
 Budapest University of Technology, Budapest, Hungary
 Jagellonian University, Krakow, Poland
 Lancaster University, United Kingdom
 UCLA, Sendai, Japan

- Tamura H.
 Taneco-Hernández M. A.
 Tanimoto Y.
 Tater M.
 Terras V.
 Teschl G.
 Thaler H.
 Tiedra de Aldecoa R.
 Timmermann W.
 Tokareva S.
 Tolar J.
 Tomassini L.
 Toninelli F.
 Tracy C.
 Trnka J.
 Troian R.
 Truc F.
 Tůma K.
 Turek O.
 Turski J.
 Tusek M.
 Ueki N.
 Ueltschi D.
 Uhlmann G.
 Umeda T.
 Umetsu H.
 Unwin J.
 Urban D.
 Uzawa K.
 Valencia Martínez G. D.
 Valiente Kroon J. A.
 Valko B.
 Vartanov G.
 Vařata D.
 Vasilyev V.
 Verstraete F.
 Veselic I.
 Veselov A.
 Vieira N.
 Vignes-Tourneret F.
 Villalba-Vega T.
 Villani C.
 Vorotnikov D.
 Vrána L.
 Vuković M.
 Vymazal M.
 Waldmann S.
 Waldron A.
 Wang W.
 Warzel S.
 Weder R.
 Weinberg S.
 Werner R. F.
 Westrich M.
 Wilkinson A.
 Winterroth E.
 Wise D.
 Wojtkiewicz J.
 Wong L. M.
 Wood S.
 Wu S.
 Xu R.
 Xu Y.
 Yajima K.
 Yamada O.
 Yau H.
 Yeo J.
 Yngvason J.
 Yokomizo N.
- Kanazawa University, Japan
 Universidad Michoacana de San Nicolas de Hidalgo, Mexico
 Università di Roma "Tor Vergata", Roma, Italy
 Nuclear Physics Institute, Rež, Czech Republic
 CNRS & ENS Lyon, France
 University of Vienna, Austria
 University of Camerino, Italy
 Pontifical Catholic University of Chile, Santiago, Chile
 Technische Universität Dresden, Germany
 Belfort, France
 Czech Technical University in Prague, Czech Republic
 University of Tor Vergata, Rome, Italy
 ENS Lyon, France
 UC Davis, United States
 Princeton University & Charles University, United States
 Donetsk National University, Ukraine
 Institut Fourier, University Grenoble 1, France
 Charles University, Prague Czech Republic
 FNSPE Czech Technical University in Prague, Czech Republic
 University of Houston-Downtown, Houston, United States
 FNSPE Czech Technical University in Prague, Czech Republic
 Kyoto University, Japan
 University of Warwick, Coventry, United Kingdom
 University of Washington, Seattle, United States
 University of Hyogo, Himeji, Japan
 Kushi National College of Technology, Kushi, Japan
 Mathematical Institute, Oxford, United Kingdom
 Tufts University, Somerville, United States
 Kinki University, Higashi-Osaka, Japan
 INRIA - UNAM, Paris, France
 Queen Mary, University of London, United Kingdom
 University of Wisconsin - Madison, Madison, United States
 JINR, Dubna, Russia
 FNSPE Czech Technical University in Prague, Czech Republic
 Bryansk State University, Bryansk, Russia
 University of Vienna, Austria
 Technische Universität Chemnitz, Germany
 Loughborough University, United Kingdom
 University of Aveiro, Portugal
 University of Vienna, Austria
 UMSNH-UNAM, Morelia, Mexico
 Ecole Normale Supérieure de Lyon, France
 Voronezh State University, Voronezh, Russia
 FJFI ČVUT, Praha, Czech Republic
 Faculty of Science, University of Sarajevo, Bosnia-Herzegovina
 ÚJF AV ČR, Rež, Czech Republic
 Albert-Ludwigs-Universität Freiburg, Germany
 American Institute of Physics, Melville, NY, United States
 Université Paris-Sud, Orsay, France
 TU Munich, Garching bei München, Germany
 INRIA-UNAM, Le Chesnay, France
 The University of Texas at Austin, Texas, United States
 Leibniz Universität Hannover, Germany
 University of Aarhus, Mainz, Germany
 Northwestern University, Evanston, United States
 University of Torino, Italy
 UC Davis, CA, USA
 Warsaw University, Faculty of Physics, Warszawa, Poland
 Caltech, Hong Kong
 ETH Zurich, Switzerland
 University of Michigan, Ann Arbor, MI, United States
 Harbin Engineering University, Harbin, China
 University of Oxford, UK
 Gakushuin University, Tokyo, Japan
 Ritsumeikan University, Kusatsu, Shiga, Japan
 Harvard, Cambridge, United States
 Konkuk University, Seoul, South Korea
 University of Vienna, Austria
 University of Sao Paulo, Brazil

Youn H.	Korea Advanced Institute of Science and Technology, Daejeon
Zagrebnov V.	Université Aix-Marseille II and CPT, Marseille, France
Zahradnik M.	MFF UK, Praha, Czech Republic
Zajac Š	Jednota českých matematiků a fyziků, Praha, Czech Republic
Zambrini J.	Univ. Lisboa, Portugal
Zeng Y.	Tsinghua University, Beijing, China
Zinoviev Y.	Steklov Mathematical Institute, Moscow, Russia
Zirnbauer M.	Universitaet zu Koeln, Germany
Znojil M.	NPI ASCR, Rez, Czech Republic
Zuevsky A.	National University of Ireland, Galway, Ireland
Żyjewski K.	University of Warmia and Mazury in Olsztyn, Olsztyn, Poland

SPEAKER INDEX

Abdesselam, 346
 Adams, 352
 Aharonov, 656
 Alexander, 377

Ballesteros, 566
 Bazhanov, 23
 Benguria, 666
 Berry, 672
 Betz, 518
 Borodin, 486
 Bossard, 574
 Boya, 658
 Buscemi, 567

Cheng, 338
 Christodoulou, 45
 Chruscinski, 515
 Cornean, 641

Díaz, 620
 Dafermos, 421
 Dappiagi, 620
 Datta, 567
 De Lellis, 452
 De Masi, 667
 Disertori, 357
 Dittrich, 339
 Drahoš, 1
 Duplantier, 56

Eckle, 641
 Enciso, 453
 Erdős, 86
 Exner, 2

Fedorchuk, 454
 Fournais, 625
 Fröhlich, 417
 Frank, 523
 Fujimoto, 377

Giuliani, 630
 Golse, 433
 Griesemer, 642
 Gruber, 568
 Guionnet, 106

Hammond, 486
 Harrow, 536
 Henrici, 339
 Hepp, 6
 Hofer, 666
 Holmes, 123
 Hounkonnou, 675

Ionescu, 452

Jansen, 640
 Joita, 515

Kaloshin, 314
 Kalvoda, 339
 Kang, 568
 Kells, 642
 Khorunzhii, 487
 Klich, 656
 Korchemsky, 619
 Kosower, 143
 Kostenko, 568
 Kozma, 377
 Kreimer, 596
 Krejčířík, 454
 Kubizňák, 580
 Kuijlaars, 487
 Kuksin, 323
 Kupiainen, 381

Löbach, 455
 Langlois, 645
 Lebowitz, 14, 675
 Lechner, 490
 Leibler, 672

Longo, 161
 Lubetzky, 464
 Lukkarinen, 386

Müger, 496
 Müller, 602
 Málek, 456
 Magnitskii, 340
 Marklof, 162
 Mastropietro, 180
 Matsui, 514
 Miermont, 470
 Monastyrsky, 658
 Mosonyi, 569
 Mostafazadeh, 658

Nachtergaele, 391
 Narita, 456
 Neshveyev, 504
 Nest, 514

O'Connell, 486
 Olla, 397

Pánek, 675
 Paleari, 417
 Pereira, 418
 Pete, 475
 Pinchover, 456
 Pizzo, 200
 Plenio, 566
 Pokorný, 457
 Procacci, 457
 Prykarpatsky, 340

Quastel, 401

Raussendorf, 657
 Renner, 541
 Ricca, 659
 Rudolph, 620
 Runkel, 608
 Rutkevich, 621

Saifullah, 659
 Samtleben, 588
 Samuelsson, 546
 Sanchez-Morgado, 341
 Schlein, 406
 Schoutens, 635
 Schumacher, 341

Šeba, 222
 Seiringer, 231
 Selberg, 438
 Serfaty, 246
 Sergveyev, 341
 Shatashvili, 265
 Shelkovich, 457
 Sherratt, 651
 Sinai, 667
 Skopina, 458
 Smedbäck, 593
 Smirnov, 362
 Šnobl, 594
 Sodin, 342
 Solovej, 640
 Sparks, 593
 Starr, 378
 Sterbenz, 444
 Susskind, 672
 Szász, 11, 328

Terras, 614
 Teschl, 342
 Toninelli, 372
 Tracy, 417
 Turek, 570
 Turski, 659

Ueltschi, 378
 Uhlmann, 657

Valiente Kroon, 455
 Valkó, 481
 Vartanov, 594
 Verstraete, 641
 Veselić, 551
 Veselov, 333
 Villani, 452

Waldmann, 509
 Wang, 667
 Warzel, 556
 Weinberg, 290
 Werner, 668
 Wilkinson, 338
 Wu, 453

Yngvason, 8

Zagrebnov, 516
 Zirnbauer, 619